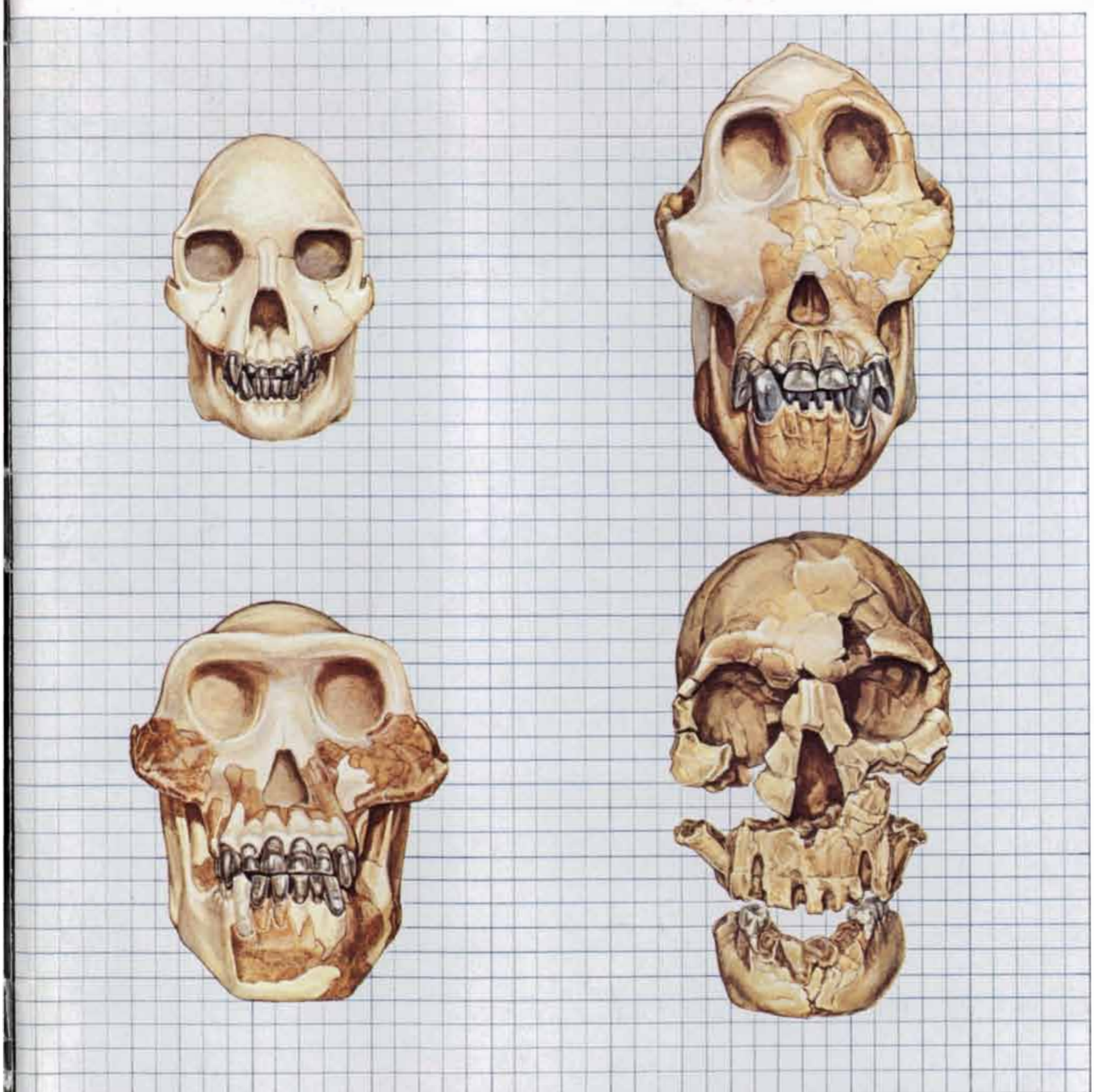
SCIENTIFIC AMERICAN



DESCENT OF HOMINOIDS AND HOMINIDS

$2.50

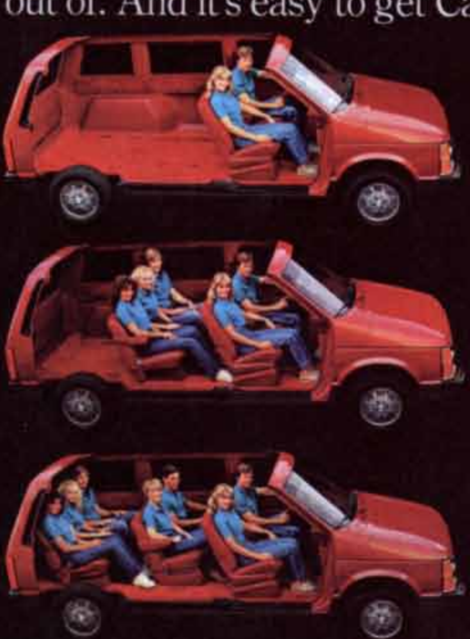
March 1984

© 1984 SCIENTIFIC AMERICAN, INC

© 1984 SCIENTIFIC AMERICAN, INC

© 1984 SCIENTIFIC AMERICAN, INC

Dodge Caravan. A truly revolutionary vehicle. It can handle two adults plus 125 cubic feet of cargo. Or five adults. Even seven with the available rear seat. Yet it's shorter than a full-size station wagon, so it's easier to maneuver and park. And since Caravan stands a mere 5'5," it's easy for you to get in and out of. And it's easy to get Caravan in and out of your garage.



Caravan's 2.2-liter engine rates an est. hwy. of 37 and EPA est. mpg of [24]*—impressive mileage for a vehicle of this sort. Caravan also has front-wheel drive to handle slippery surfaces outside and give you more room inside. Yet for all of this, Caravan is remarkably low-priced and even backed by Dodge's 5 year/50,000 mile Protection Plan on engine, powertrain and outer body rust-through.**

The totally new Dodge Caravan. You've got to see it, sit in it, and drive it to believe it. All of which can happen at your Dodge dealer—where you can buy or lease** your very own transportation revolution. Order one now.

The New Chrysler Technology. Quality backed by 5/50 Protection.

*Use EPA est. mpg for comparison. Your mileage may vary depending on speed, distance and weather. Actual hwy. mpg and CA ests. lower. **5 years or 50,000 miles, whichever comes first.

© 1984 SCIENTIFIC AMERICAN, INC

Established 1845 **SCIENTIFIC AMERICAN** March 1984 Volume 250 Number 3

ARTICLES

DEPARTMENTS

BOARD OF EDITORS Gerard Piel (Publisher), Dennis Flanagan (Editor), Brian P. Hayes (Associate Editor), Philip Morrison (Book Editor), John M. Benditt, Peter G. Brown, Michael Feirtag, Robert Kunzig, Jonathan B. Piel, John Purcell, James T. Rogers, Armand Schwab, Jr., Joseph Wisnovsky

ART DEPARTMENT Samuel L. Howard (Art Director), Steven R. Black (Assistant Art Director), Ilil Arbel, Edward Bell

PRODUCTION DEPARTMENT Richard Sasso (Production Manager), Carol Eisler and Leo J. Petruzzi (Assistants to the Production Manager), Carol Hansen (Electronic Composition Manager), Carol Albert, Karen Friedman, Lori Mogol, Karen O'Connor, Julio E. Xavier

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

GENERAL MANAGER George S. Conn

ADVERTISING DIRECTOR C. John Kirby

CIRCULATION MANAGER William H. Yokel

SECRETARY Arlene Wright

SCIENTIFIC AMERICAN (ISSN 0036-8733), PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. COPYRIGHT © 1984 BY SCIENTIFIC AMERICAN, INC. ALL RIGHTS RESERVED. PRINTED IN THE U.S.A. NO PART OF THIS ISSUE MAY BE REPRODUCED BY ANY MECHANICAL, PHOTOGRAPHIC OR ELECTRONIC PROCESS, OR IN THE FORM OF A PHONOGRAPHIC RECORDING, NOR MAY IT BE STORED IN A RETRIEVAL SYSTEM, TRANSMITTED OR OTHERWISE COPIED FOR PUBLIC OR PRIVATE USE WITHOUT WRITTEN PERMISSION OF THE PUBLISHER. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $24 PER YEAR IN THE U.S. AND ITS POSSESSIONS; $33 PER YEAR IN ALL OTHER COUNTRIES. POSTMASTER: SEND ADDRESS CHANGES TO SCIENTIFIC AMERICAN, 415 MADISON AVENUE, NEW YORK, N.Y. 10017

© 1984 SCIENTIFIC AMERICAN, INC

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, UT 84121. Offprints adopted for classroom use may be ordered direct or through a college bookstore. Sets of 10 or more Offprints are collated by the publisher and are delivered as sets to bookstores.

Photocopying rights are hereby granted by Scientific American, Inc., to libraries and others registered with the Copyright Clearance Center (CCC) to photocopy articles in this issue of SCIENTIFIC AMERICAN for the flat fee of $1.25 per copy of each article or any part thereof. Such clearance does not extend to the photocopying of articles for promotion or other commercial purposes. Correspondence and payment should be addressed to Copyright Clearance Center, Inc., 21 Congress Street, Salem, MA 01970. Specify CCC Reference Number ISSN 0036-8733/84. $1.25 + 0.00.

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

Advertising correspondence should be addressed to C. John Kirby, Advertising Director, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, NY 10017.

Subscription correspondence should be addressed to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 5969, New York, NY 10017. The date of the last issue on your subscription is shown in the upper right-hand corner of each month's mailing label. For change of address notify us at least four weeks in advance. Please send your old address (if convenient, on a mailing label of a recent issue) as well as the new one.

Name

New Address

Street

City

State and ZIP

Old Address

Street

City

State and ZIP

THE COVER

The painting on the cover shows the restored skulls of four species that have played an important part in modern understanding of human evolution (see "The Descent of Hominoids and Hominids," by David Pilbeam, page 84). At the top left is the African Miocene ape *Proconsul africanus*. At the top right is the larger and later Miocene ape *Sivapithecus*, best known from fossil deposits south of the Himalayas; this genus was probably the ancestor of the living Asian great ape, the orangutan. At the bottom left is the *Australopithecus* species uncovered in Ethiopia, *A. afarensis*. At the bottom right is the species *Homo habilis*, first found in Kenya. Both *A. afarensis* and *H. habilis* predate the better-known *Homo erectus*, a species that first appeared in Africa about 1.6 million years ago and that was the immediate predecessor of *Homo sapiens*.

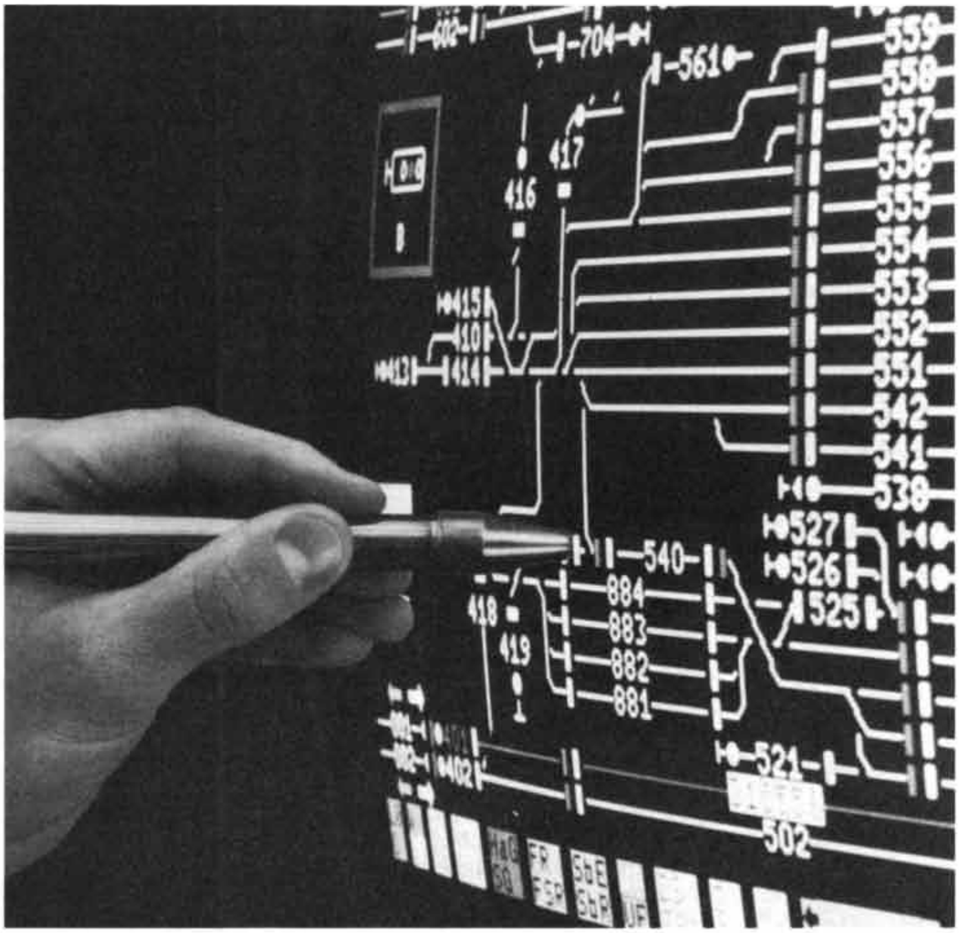
THE ILLUSTRATIONS

Cover painting by Enid Kotschnig

Page	Source
13	Thomas M. Pantages; Norman Margolus, Tommaso Toffoli and Gérard Y. Vichniac
14–16	Ilil Arbel
17	Stephen Wolfram
20	Ilil Arbel
21	Thomas M. Pantages; Norman Margolus, Tommaso Toffoli and Gérard Y. Vichniac
34–41	Ian Worpole
43	Charles D. Hollister
44–45	Walken Graphics
46	Charles D. Hollister (*top*), Walken Graphics (*bottom*)
47–53	Walken Graphics
55	Bunji Tagawa
56	Richard H. Scheller and Richard Axel
57	Richard H. Scheller and Richard Axel (*top*), Bunji Tagawa (*middle and bottom*)
59–60	Bunji Tagawa
62	Richard H. Scheller and Richard Axel
85	Enid Kotschnig
86–88	Tom Prentiss
91–95	Ilil Arbel
99	André Mysyrowicz, David P. Trauernicht and James P. Wolfe
100–101	Gabor Kiss
102	David P. Trauernicht, University of Illinois at Urbana-Champaign
103	Paul L. Gourley and James P. Wolfe, University of Illinois at Urbana-Champaign (*left*); Gabor Kiss (*right*)
104–107	Gabor Kiss
109–114	Patricia J. Wynne
117–120	Alan D. Iselin
121	F. H. Bronson, University of Texas at Austin (*top*); Alan D. Iselin (*bottom left*)
124–125	Alan D. Iselin
129	Thomas M. Pantages, courtesy of Houghton Library, Harvard University
130	Falk Library of the Health Sciences, University of Pittsburgh
131–137	Thomas M. Pantages, courtesy of Widener Library, Harvard University
138–143	Michael Goodman

© 1984 SCIENTIFIC AMERICAN, INC

How a pen can stop a train.



At a freight yard in Austria, the yard controller draws a red bar on a video screen and stops a rolling 70-ton train car on a siding.

The controller makes a green line on the screen and sends another car along a clear path.

And to make sure boxcars for Buchs don't end up in Buchenberg, the controller sets switch points with a pen point.

No ordinary pen, this light-sensing stylus is the nub of a computer railyard switching system developed by the people at ITT.

With it, the Austrian Federal Railways controls the Wolfurt yard by "writing" routing instructions on a video terminal.

Controlled by an ITT microprocessor, the video display system is connected to a fail-safe keyboard which can also direct the yard's switches.

So even if the light-pen isn't working, the trains will be.

In time, railways will use this ITT technology in freight depots around the world.

And that will help make railroad operations easier and more productive.

All at the touch of a pen.

The best ideas are the ideas that help people. ITT

© 1983 ITT Corporation, 320 Park Avenue, New York, N.Y. 10022

LETTERS

Sirs:

"The Uncertainties of a Preemptive Nuclear Attack," by Matthew Bunn and Kosta Tsipis [SCIENTIFIC AMERICAN, November, 1983], raises an interesting dilemma for the scientific community. Bunn and Tsipis maintain the U.S. need not worry because the Russians are such poor shots that they cannot kill even our present missile silos at an overpressure of 2,000 pounds per square inch and a spacing of five miles. Other respected scientists, however, sought to discredit the closely-spaced-basing concept for the MX (Peacekeeper) ICBM by contending the Russians are such unerring marksmen that they could kill "superhard" silos spaced close enough to create fratricide effects far worse than those described by Bunn and Tsipis. Perhaps the scientific community needs to get its act together!

The truth appears to lie between these two extremes. The Russians are not the technological dwarfs Bunn and Tsipis portray because the data they cite uniformly understate known Russian accomplishments. On the other hand, their marksmanship is not of the "10-feet tall" variety either.

RICHARD D. DELAUER

Under Secretary of Defense
Research and Engineering
Department of Defense
Washington

Sirs:

Bunn and Tsipis' painstaking analysis of the uncertainties of a preemptive nuclear attack shows quite convincingly that the vulnerability of ICBM's has been grossly overestimated. The article is deficient, however, it seems to me, in two important respects.

First, before spending so much time and effort to demonstrate that ICBM's are not as vulnerable as had been maintained, should not the authors have taken the trouble to tell us why we should be worried about ICBM vulnerability in the first place? Does the vulnerability of our ICBM's significantly threaten our ability to deter nuclear attack? The authors state quite flatly that it does, but they offer neither proof nor reasoning to support the assertion. With 4,656 U.S. SLBM warheads—more than 10 times the number required to lay waste the U.S.S.R.—lurking quite invulnerably under 300 million square kilometers of ocean, it is not at all obvious that it does.

Second, the whole point of the ICBM vulnerability controversy *was* the MX. The MX was to be both accurate and invulnerable, so that it could destroy, but not be destroyed by, Russian ICBM's. The Pentagon exaggerated the vulnerability of its own ICBM's in order to sell the MX.

RALPH CHERNOFF

Poughkeepsie, N.Y.

Sirs:

Dr. DeLauer further confuses the already difficult issues surrounding the vulnerability of various ICBM basing modes by implying that there are identical accuracy and fratricide issues associated with attacks on "dense pack" and Minuteman silos. As we hope he is aware, the high accuracies necessary to attack the Minuteman silos are not necessary to attack the closely spaced dense-pack silos, since an inaccurate weapon that missed one silo would hit another.

Similarly, the fratricide encountered in an attack involving a small number of large weapons each aimed several hundred meters apart from one another would not be analogous to the fratricide encountered in attempting to target two warheads against each of more than 1,000 separate aim points. Moreover, most of the debate over dense pack centered on attacks that might take place in the future rather than in the present. Thus there is no inconsistency; the scientific community can rest easy.

Dr. DeLauer believes we make the Russians out to be "technological dwarfs," yet of the various estimates publicly available we used those that imputed the highest accuracy to Russian ICBM's. If we had used the lower accuracy estimates given by the International Institute for Strategic Studies, the kill probabilities of Russian ICBM's against U.S. silos would be substantially lower than those we calculated.

Mr. Chernoff is quite correct in his judgment that the survival of the ICBM's is not central to our ability to deter nuclear attack, for the simple reason that several thousand thermonuclear weapons based on submarines and bombers would survive even the most successful attack on U.S. land-based strategic forces. The Central Intelligence Agency, the Office of Technology Assessment, the Navy and the Rand Corporation all agree that U.S. strategic submarines will be invulnerable into the foreseeable future.

There are those who believe, however, that ICBM vulnerability is important in and of itself. The argument over this point is rather arcane. To summarize it, some analysts have held that if the U.S.S.R. launched an attack limited to U.S. ICBM silos, leaving U.S. cities intact, the U.S. could not respond against Russian cities for fear that the Russians would then escalate their attack to destroy a corresponding number of U.S. cities. Since the ICBM's destroyed in the initial attack are the only weapons in the U.S. arsenal that could deliver a "mirror image" strike against the remaining Russian ICBM's, these analysts contend that the U.S. would have no alternative but to give in.

This argument would be seriously flawed *even if U.S. ICBM's were immediately vulnerable.* Since such an attack on U.S. ICBM silos alone would result in from 20 to 40 million casualties, more than 20 times the total American casualties in all previous wars combined, it is inconceivable that the U.S. would not respond, using some of the thousands of surviving warheads. Such a situation would probably escalate into an all-out strategic exchange, in which the attacker would be as totally destroyed as the attacked. Thus it is unlikely that the cautious Russian leadership would ever seriously consider such an attack.

There is a wide variety of targets for such retaliation other than cities and ICBM silos; even with only SLBM's surviving the U.S. would have many options available, any of which would cause a level of retaliatory destruction that would be completely unacceptable to any rational attacker. Indeed, a retaliatory strike against Russian ICBM's (the only military purpose of the MX) would be among the *worst* options available to the U.S. When the Russian warning systems told them that U.S. ICBM's were on their way to destroy Russian ICBM silos, it is likely the U.S.S.R. would simply launch its remaining ICBM's against the U.S., rendering the U.S. retaliation both futile and suicidal. Moreover, given the deployment of Russian ICBM's (with four major ICBM fields directly around Moscow) an attack on Russian ICBM's might seem quite similar to an attack on Soviet cities, further increasing the probability that the war would escalate to the total destruction of both nations.

As Bernard Brodie once pointed out, the statement that a scenario is conceivable is proved by the mere act of saying so; whether it is worth a second thought is another matter entirely. This scenario, although it may be worth a second or third thought, is not an adequate justification for weapon systems as destabilizing and as expensive as the MX and the Trident II. Since the arguments against it are complex, however, and we did not wish to deal with them in detail, we simply stated in our article that "concerns have arisen" over the issue, leading to Mr. Chernoff's complaint.

MATTHEW BUNN

KOSTA TSIPIS

Massachusetts Institute of Technology
Cambridge

© 1984 SCIENTIFIC AMERICAN, INC

Red Adair and his Rolex stopped at nothing to blow out The Devil's Cigarette Lighter.

Red Adair puts out oil-well fires.

The most famous of them was nicknamed The Devil's Cigarette Lighter.

It was in the Sahara, at a place called Gassi Touil.

Every day, over 15 million cubic meters of gas roared 200 meters into the desert sky from this terrifying blowout.

Working under tons of falling water to combat the heat, Red and his team were able to maneuver 250 kilos of special explosive to within half a meter of the base of the flames. They set off the blast, and the fire was out in a split second.

In the slightly less heated atmosphere of the Rolex factory in Geneva, craftsmen create a watch designed to withstand the most appalling conditions.

Every Rolex chronometer we make is officially tested night and day for two weeks. The Rolex movement must survive 15 days of extreme temperature changes, hanging unguarded by its rugged Oyster case.

But who could even guess at the sort of treatment a Rolex can expect to receive on the wrist of a man like Red Adair?

In 1977, Red and his team returned in triumph from the North Sea, after taming the Bravo blowout, a mixture of boiling oil and gas screaming out of the well-head at several hundred k.p.h.

The Daily Mail's reporter observed, "Solid gold Rolex watches were clamped to their wrists like a company badge."

Some company.
Some badge.

ROLEX

Pictured: The Rolex Day-Date Chronometer. Available in 18kt. gold, with matching bracelet.

Write for brochure. Rolex Watch, U.S.A., Inc., Dept. 452, Rolex Building, 665 Fifth Avenue, New York, NY 10022. World headquarters in Geneva. Other offices in Canada and major countries around the world.

© 1984 SCIENTIFIC AMERICAN, INC

Why? The answer lies in the difference between a luxury car and a premium automobile. The new Continental Mark VII.

A premium automobile offers qualities such as functional beauty, and the kind of performance and precise control that's both rewarding and reassuring to the driver.

Take the sleek shape of this new Mark VII, for example. Air flow over the car actually helps it move over the road more solidly and quietly.

And there's still another reason for the Mark VII's remarkable road manners: electronically controlled air suspension. This advanced suspension, offered by no other car maker, gives you a rare combination of riding comfort and control. It automatically levels the car to compensate for variations in passenger or luggage load, so that ride and handling are unusually consistent.

Inside, it's instantly apparent that the Continental Mark VII was designed for the driver, as a premium

If all you're getting out of your luxury car is luxury, we respectfully suggest you get rid of it.

© 1984 SCIENTIFIC AMERICAN, INC

© 1984 SCIENTIFIC AMERICAN, INC

© 1984 SCIENTIFIC AMERICAN, INC

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

MARCH, 1934: "Ultra-short radio waves are from 10 meters down to about one millimeter in length. Prof. G. Potapenko of the Moscow Mining Academy has succeeded in obtaining waves as short as three centimeters. Among the many technical applications of high-frequency oscillations and ultra-short waves their use in radio transmission is the best-known. Ultra-short waves are particularly well adapted for this purpose because they can be focused and transmitted in the form of narrow beams of less than one square meter cross-section. Recently, with the help of ultra-short waves, important results have been obtained with regard to what physicists call the problem of the solid state of matter. A very promising start in this direction has been made by Prof. Potapenko and Dr. R. Sänger, working together at the California Institute of Technology. These investigations have thrown new light especially on the constitution of ferromagnetic substances, such as iron, nickel and cobalt. Ultra-short waves are destined to become a universal tool in the hands of scientists and engineers."

"We wish to draw some merited attention to that extraordinary and little-appreciated sky boat, the humble 'blimp.' We put the word in quotation marks because it is really a nick-name, being a contraction of the official British Air Service designation, 'B-Limp,' which applied to small dirigibles that had no interior framework to help them hold their shape. Since 1925 the Goodyear blimp fleet has carried, in round numbers, 100,000 passengers and has covered a million and a half miles round and about the United States. Blimps have been used (1) to rescue persons from terrain on which no kind of airplane could possibly land, (2) to land passengers on the roof of skyscrapers, in small city parks and on the deck of ships, (3) to set down scientific workers and their equipment on mountain ranges, (4) to remain for hours aloft watching for forest fires, with little expenditure of fuel, (5) to operate on regular passenger runs without the need for landing fields at every stop and finally (6) to serve as observation patrol in wartime."

"Edison Pettit and Seth B. Nicholson of the Mount Wilson Observatory report the long series of measures of heat radiation with the thermocouple. Stars at lower temperatures radiate the bulk of their energy in the infra-red, leaving but a beggarly fraction in the visible region. Could we see heat, therefore, the coolest stars would come up so much in brightness as to change the familiar aspect of the constellations. Betelgeuse, for example, would look as bright as Sirius, and Alpha Herculis, to actual eyes a not very conspicuous star of the third magnitude, would rival Canopus. The long-period variables would show still greater changes. Mira Ceti, which to the eye rises to the fourth magnitude at its maximum, to the thermocouple is as bright as Vega."

"Correct exposure is one of the vital problems of good photography. So many factors enter into proper exposure that even those with long experience, who pride themselves on their picture results, freely admit their inability consistently to judge photographic light under all conditions. Now seasoned photographers need no longer rely on their personal judgment of light values but are able to devote all of their attention to composition. They determine exposures with the modern and positive way—with a Weston exposure meter. The untiring 'electric eye' in the exposure meter never guesses. It properly gauges the photographic light regardless of whether exposures are made in sunlight or in deep shade. Thus film wastage is eliminated and disappointment over the loss of prized pictures is a thing of the past."

SCIENTIFIC AMERICAN

MARCH, 1884: "Malaria has been in time past reckoned a marsh fever, and it was commonly believed to be limited to regions and to seasons where vegetable matters were undergoing decomposition from moisture and heat. New England was formerly free of it. It is possible that a single case may have occurred occasionally in the extreme southwestern part of Connecticut, but ague was practically unknown there until 1860. In that year an onward movement to the east and the north commenced, and it has continued unchecked till the present time, and in its progress it has overturned theories at a rate as remarkable as its own advance. Commencing on Long Island Sound, at Southport, in 1860, it spread with an irregular front, reaching New Haven in 1864, Hartford in 1872 and the northeastern part of the state not until 1882. It has maintained a steadily epidemic condition ever since, having swept the state so fully that not a town is now exempt from its sway. It has already invaded the sacred precincts of Boston and is in high march for Nova Scotia and Newfoundland, for all that we can tell."

"Prof. Charles A. Young of Princeton in a recent lecture asks: 'Is there any connection between the maximums of sun spots and terrestrial disturbances of any kind?' The average of the occurrence of sun spots is, from maximum to maximum, either 13.11 years or 9.11 years. It was said that in 1870 there was a maximum of sun spots and an extraordinarily hot summer, but it was forgotten that observations to be conclusive must cover the whole of the earth's surface and not a part only. It so happened that in 1870 at our antipodes in New Zealand there was at the time of our hot summer an intensely cold winter, something phenomenal for that region. But an unmistakable connection has been discovered between the sun-spot maximums and aberrations of the magnetic needle. It is well established that the period of a maximum of spots is coincident with a period of the greatest declination of the needle."

"Within recent years it has been established beyond doubt that there exists in the blood of mammals a third type of corpuscle, differing morphologically from both the red and the white corpuscle and possessing certain distinctive properties of the greatest importance in coagulation. These elements were called hematoblastes by Georges Hayem, upon the supposition that they are eventually transformed into red corpuscles. As this view is by no means established, it is better to speak of them as blood plates, the name given to them by Giulio Bizzozero."

"A Mexican correspondent writes: 'There is a considerable variety of tongues among the messages going over the telephone wires in Mexico. The peremptory method of making telephone calls—"Hello! Hello? Give me 1299!" etc.—would never do in the Castilian tongue. Courtesy of intercourse must be preserved even between invisible communicants, and the unseemly vexatiousness and petulance that the telephone seems to provoke in Anglo-Saxon moods is never allowed to obtain utterance here. The regular responses from the central office to a telephone call is "Mande usted!" which is equivalent to "At your command!" Then preliminaries are gone through something as follows: "Very well, I thank you. What service may I render you?" "Will you kindly do me the favor of enabling me to speak with Don So-and-So, No. 777?" "With much pleasure!" etc., etc., and when the connection is made, the usual polite introductories are gone through before proceeding to the business in hand.'"

© 1984 SCIENTIFIC AMERICAN, INC

THE AUTHORS

KURT GOTTFRIED, HENRY W. KENDALL and JOHN M. LEE ("'No First Use' of Nuclear Weapons") are respectively professor of physics at Cornell University, professor of physics at the Massachusetts Institute of Technology and vice admiral (retired) in the U.S. Navy. They collaborated on a study of the doctrine of "no first use" done by the Union of Concerned Scientists. Gottfried was born in Vienna in 1929 and emigrated to Canada in 1939. He received his undergraduate education at McGill University and his doctorate at M.I.T. Thereafter he was a Junior Fellow at Harvard University and a Ford Foundation Fellow at the Niels Bohr Institute in Copenhagen. From 1960 to 1964 he was assistant professor of physics at Harvard. He has been a member of the Cornell faculty since 1964 and professor of physics since 1968. Kendall received his undergraduate education at Amherst College and his doctorate from M.I.T. Thereafter he was a National Science Foundation Fellow at M.I.T. and a member of the faculty of Stanford University; he has been professor at M.I.T since 1967. Lee had 38 years of active duty in the navy, including posts at the North Atlantic Treaty Organization, the United Nations and the Pentagon. He was head of the Planning Staff of the Office of the Assistant Secretary of Defense for International Security Affairs and assistant director of the Arms Control and Disarmament Agency. He directed the Union of Concerned Scientists' study of no first use.

CHARLES D. HOLLISTER, ARTHUR R. M. NOWELL and PETER A. JUMARS ("The Dynamic Abyss") are oceanographers with a common interest in deep-ocean processes. Hollister is Dean of Graduate Studies at the Woods Hole Oceanographic Institution. He got his B.S. at Oregon State University and did his graduate work at the Lamont-Doherty Geological Observatory of Columbia University, which awarded him his Ph.D. in 1967. His dissertation was the first to deal with the geologic effects of bottom currents in the deep sea, a region thought up to that time to be tranquil. In addition to being an oceanographer Hollister is a mountaineer; he has climbed in both the Arctic and the Antarctic, and he is past president of the American Alpine Club. Nowell is associate professor in the School of Oceanography of the University of Washington and Visiting Investigator at the Woods Hole Oceanographic Institution. He was an undergraduate at Trinity College, Cambridge, and received his master's and doctor's degrees from the University of British Columbia. Jumars is professor of oceanography at the University of Washington. Born in Germany, he studied as an undergraduate at the University of Delaware and was granted his Ph.D. by the Scripps Institution of Oceanography. His oceanographic work has a strong biological component; he is particularly interested in the interaction of bottom organisms and sedimentary processes.

RICHARD H. SCHELLER and RICHARD AXEL ("How Genes Control an Innate Behavior") are assistant professor of biological sciences at Stanford University and professor of pathology and biochemistry at the Columbia University College of Physicians and Surgeons. After Scheller's graduation from the University of Wisconsin at Madison, he went to the California Institute of Technology, where he got his Ph.D. in 1980. In 1980 and 1981 he was a postdoctoral fellow in the laboratories of Axel and Eric Kandel at the College of Physicians and Surgeons. Axel received his A.B. from Columbia College and his M.D. from the Johns Hopkins School of Medicine. He joined the Columbia faculty in 1972. The main theme of his work has been the development of universal techniques for transferring genes from virtually any cell to any other cell.

DAVID PILBEAM ("The Descent of Hominoids and Hominids") is professor of anthropology at Harvard University. He did his undergraduate work at the University of Cambridge and continued there as a demonstrator in anthropology. He was awarded his Ph.D. by Yale University in 1967 and joined the Yale faculty the followng year. He moved to Harvard in 1981. He remarks: "I am drawn to human evolutionary studies because they require one to be something of a historian and philosopher too. I am very interested in the interplay between 'fact' and 'theory,' in understanding more about where many of our basic theories come from."

JAMES P. WOLFE and ANDRÉ MYSYROWICZ ("Excitonic Matter") are professor of physics at the University of Illinois at Urbana-Champaign and a physicist with the French National Center for Scientific Research (CNRS). Wolfe's degrees (B.A. 1965 and Ph.D. 1971) are from the University of California at Berkeley. From 1971 to 1981 he was on the Berkeley faculty; he went to Illinois in 1981. There he has worked mainly on techniques of imaging the transport of phonons and excitons in crystals. Mysyrowicz has studied optical effects in semiconductors since his doctoral thesis at the University of Strasbourg in 1968.

GERALD OSTER ("Muscle Sounds") is professor of biophysics at the Mount Sinai School of Medicine of the City University of New York. He received his Sc.B. at Brown University in 1940 and his Ph.D. in physical chemistry from Cornell University in 1943. He did postdoctoral research at the Massachusetts Institute of Technology and Princeton University before he joined the Rockefeller Institute for Medical Research in 1945. In 1949 and 1950 he was a Rockefeller Foundation Fellow in London at Birkbeck College and at the Royal Institution. He did research at the University of Strasbourg and at the University of Paris before going to the Polytechnic Institute of Brooklyn (now the Polytechnic Institute of New York) in 1952. There he was professor of polymer chemistry until he moved to Mount Sinai in 1969. Oster is also working at his experimental plantation in the Caribbean toward scientific solutions to problems of agriculture in Third World countries.

F. H. BRONSON ("The Adaptability of the House Mouse") is professor of zoology at the University of Texas at Austin. He was educated at Kansas State University (B.S. 1957 and M.S. 1958) and Pennsylvania State University (Ph.D. 1961). From 1961 to 1968 he was a staff scientist at the Jackson Laboratory in Bar Harbor, Me., where he first became interested in the house mouse; he has been at Texas since 1968. He remarks: "I would describe myself as a physiologist who is interested broadly in the environmental control of mammalian reproduction."

I. BERNARD COHEN ("Florence Nightingale") is Victor S. Thomas Professor of the History of Science at Harvard University. With the exception of three years as a fellow at the Carnegie Institution of Washington, he has been at Harvard since he began his undergraduate education in 1933. He received his B.S. in 1937 and his Ph.D. (the first ever awarded in the history of science) in 1947; he has been professor of the history of science since 1959. He has also been the editor of *Isis,* the quarterly journal of the History of Science Society, and president of that society. Over the years he has mainly been interested in the science of Isaac Newton and Benjamin Franklin; the first of his 17 articles in SCIENTIFIC AMERICAN was "In Defense of Benjamin Franklin" (August, 1948). He has recently finished a historical and analytical study of revolutions in science. Currently Cohen is doing research on the historical interactions of the natural and exact sciences with the social sciences.

© 1984 SCIENTIFIC AMERICAN, INC

The "brilliant and completely convincing"* portrait of this year's Nobel Laureate in Medicine

A FEELING FOR THE ORGANISM

THE LIFE AND WORK OF BARBARA McCLINTOCK

Evelyn Fox Keller
Northeastern University

"Excellent...a welcome and useful addition to the growing literature on the recent history of biology and women's achievements in science. It *is* interesting to read about women geniuses, thorns, bristles, and all..."
—*The New York Times*

A Feeling for the Organism is the revealing and compelling story of Barbara McClintock, the 1983 Nobel Laureate for her discovery of genetic transposition and rearrangement. Now accorded worldwide recognition for her research, McClintock followed a path that diverged significantly from the mainstream. This major biography is a salute to the product of that independence.

"One of the most interesting of contemporary scientific biographies. I suspect that the portrait drawn here is the most we shall ever learn about this remarkable woman."
—*Science 83**

235 pages, illustrated. $14.95 cloth.
Available in fine bookstores or use the accompanying coupon.

Please send me____copy(ies) of A FEELING FOR THE ORGANISM (ISBN 1433) at $14.95 each.

☐ I enclose a check made payable to W. H. Freeman (please include $1.50 for postage and handling; New York, California, and Utah residents add appropriate sales tax)

Charge ☐ MasterCard ☐ Visa

Expiration Date ________/________

ACCOUNT NUMBER________________________

SIGNATURE________________________
(All credit card orders must be signed.)

NAME________________________

ADDRESS________________________

CITY________________________

STATE/ZIP________________________

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

0127

© 1984 SCIENTIFIC AMERICAN, INC

COMPUTER RECREATIONS

The cellular automaton offers a model of the world and a world unto itself

by Brian Hayes

It is due cause for wonder that molecules of water "know" how to frame the elaborate symmetries of a snowflake. There is no architect directing the assembly, and the molecules themselves carry within them no template for the crystalline form. Pattern on a large scale emerges entirely from the short-range interactions of many identical units. Each molecule responds only to the influence of its nearest neighbors, but a consistent arrangement is maintained throughout a structure made up of perhaps 10^{20} molecules.

One way to approach an understanding of this process is to imagine that each site where a molecule might be emplaced is governed by a rudimentary computer. As the crystal grows, each computer surveys the surrounding sites and, depending on its findings, determines by some fixed rule whether its own site should be occupied or vacant. The same calculation is made at all the sites according to the same rule.

The computational model of snowflake growth is a cellular automaton: a uniform array of many identical cells, or sites, in which each cell has only a few possible states and interacts only with a few neighboring cells. The components of the system—the cells and the rule for calculating the next state of a cell—can be simple indeed and nonetheless give rise to a remarkably complex evolution.

The idea of the cellular automaton is roughly as old as the electronic digital computer. The first investigations were carried out by John von Neumann (with an important contribution from Stanislaw Ulam) in the early 1950's. Von Neumann's primary aim was to devise a simple system capable of reproducing itself in the manner of a living organism. The best-known cellular automaton, the "game of life" invented in 1970 by John Horton Conway, also has a biological aspect, as the name suggests; cells are born, live or die depending on the local population density.

In more recent work on cellular automata the emphasis has shifted somewhat. Arrays of locally interacting cells are seen as potentially useful models of physical systems, ranging from snowflakes to ferromagnets to galaxies. They may also have applications to questions in computer science, both practical (How should one organize a network of many interacting computers?) and theoretical (What is the ultimate limit to the power of a computing machine?). Perhaps most intriguing, the cellular automaton can be viewed as a "digital universe" worth exploring for its own sake, quite apart from its utility as a model of the real world.

The resurgence of interest in cellular automata was marked by a workshop on the subject held a year ago at the Los Alamos National Laboratory. The proceedings (some 20 papers) have since been published in *Physica D* and in book form by the North-Holland Publishing Company. Almost all of what is reported here is based on work discussed at the Los Alamos meeting.

Four properties characterize a cellular automaton. The first property is the geometry of the array of cells. For a model of snowflake growth a two-dimensional hexagonal array would be appropriate, but in most contexts a rectilinear lattice is chosen, one made up of identical squares. Arrays with three or more dimensions are readily constructed but are not readily visualized. Lately surprising discoveries have been made with the still simpler one-dimensional array: a mere line of cells.

Within a given array it is necessary to specify the neighborhood that each cell examines in calculating its own next state. In the two-dimensional rectilinear array two neighborhoods have been given much attention. Von Neumann confined each cell's attention to its four nearest neighbors, those to the north, south, east and west; this set of cells is now called the von Neumann neighborhood. The neighborhood that includes these four cells and the four diagonally adjacent ones is called the Moore neighborhood, after Edward F. Moore. Obviously neighborhoods overlap, and a given cell is simultaneously included in the neighborhoods of several adjacent cells. In some cases the center cell—the cell making a calculation—is considered a member of its own neighborhood.

The third factor to be considered in describing a cellular automaton is the number of states per cell. Von Neumann found a self-replicating pattern made up of cells with 29 possible states, but most automata are far simpler. Indeed, there is ample scope for variation even among the binary automata, those with only two states per cell; the states may be represented as 1 or 0, true or false, on or off, living or dead.

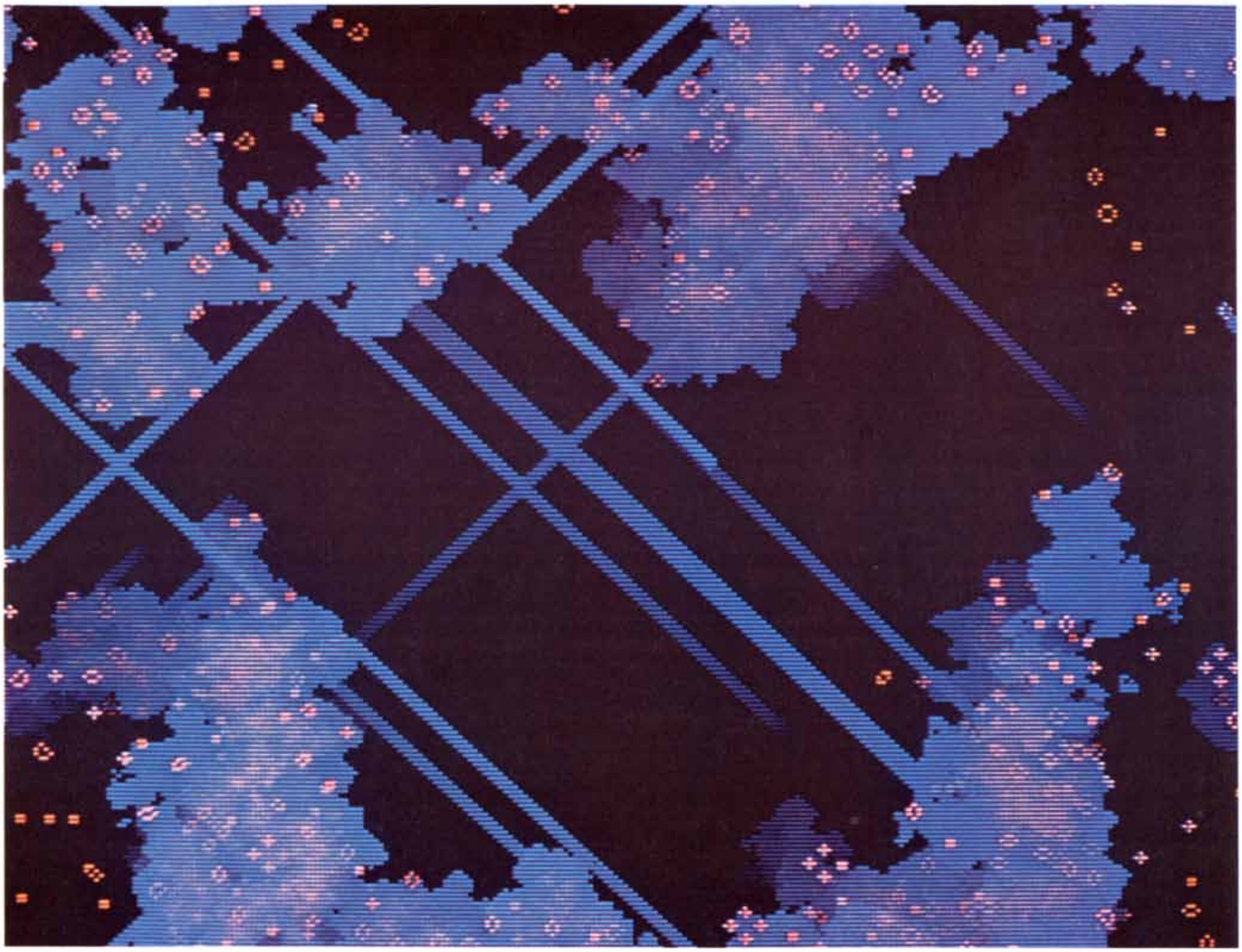
The primary source of variety in the universe of cellular automata is the enormous number of possible rules for determining the future state of a cell based on the present configuration of its neighborhood. If k is the number of states per cell and n is the number of cells included in the neighborhood, there are k^{k^n} possible rules. Thus for a binary automaton in the von Neumann neighborhood (where n is 4) there are more than 65,000 possible rules; in the Moore neighborhood (where n is 8) there are 10^{77}. Only a trifling fraction of them have been examined at all.

The game of life is played with two-state cells on a rectilinear lattice in the Moore neighborhood, with the additional complication that the center cell is significant. In other words, at each step in the evolution of the system every cell checks the state of the eight surrounding cells as well as its own state. According to the rule defined by Conway, if the center cell is living, it will continue to live in the next generation if either two or three of the eight cells in the neighborhood are also living. If there are three live cells in the neighborhood, the center cell is alive in the next generation regardless of its present state. Under all other circumstances the center cell either dies or remains dead.

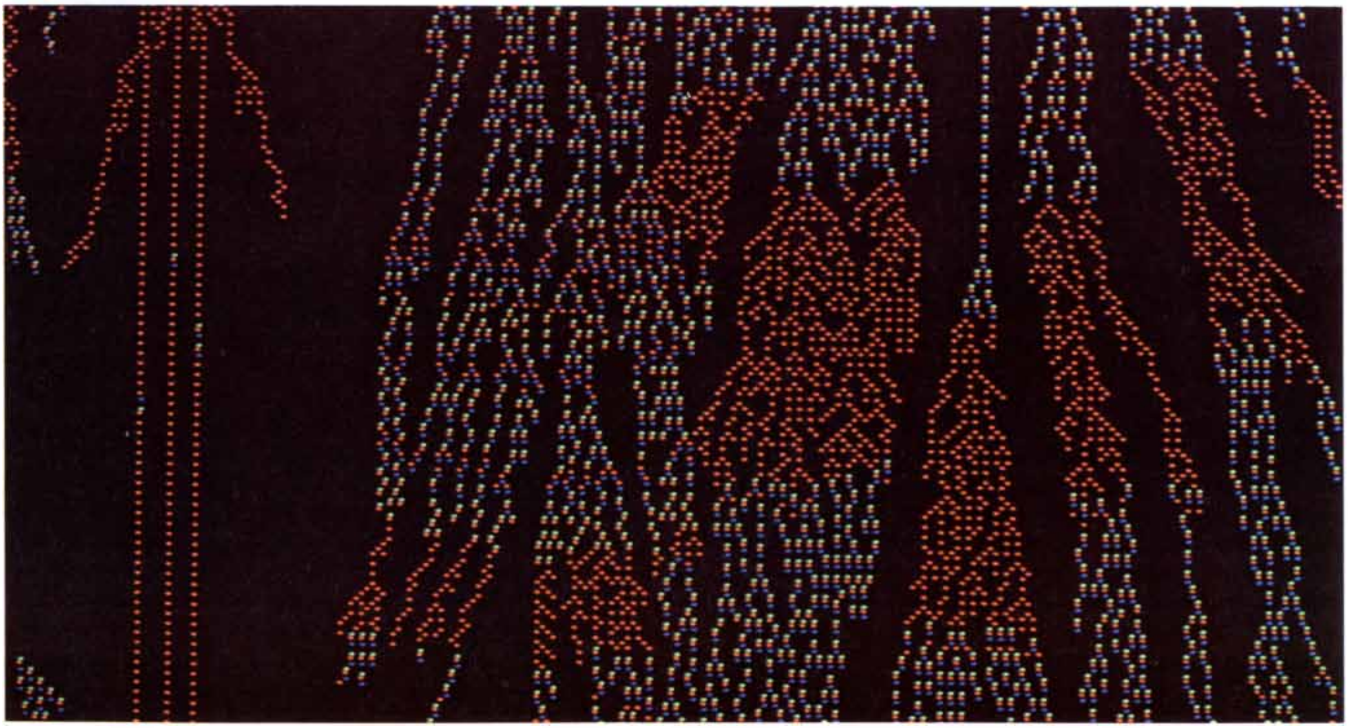
The fascination of the game of life is its unpredictability. Some patterns die out entirely; many more lapse into a stable configuration or a cyclical one with a period of a few generations. Over the years, however, a number of more interesting initial states have been discovered, such as the "glider gun" that launches an unending stream of projectiles. The exploration of life's byways continues. Recent developments are described by Martin Gardner in *Wheels, Life, and Other Mathematical Amusements.* Here I should like to turn to other cellular automata whose properties are just beginning to be elucidated.

Among the multitude of possible transition rules, many hold little intrinsic interest. For example, a rule stating that a cell will be on if and only if the cell to its left is on specifies an evolution that is quite easy to predict: any initial pat-

© 1984 SCIENTIFIC AMERICAN, INC

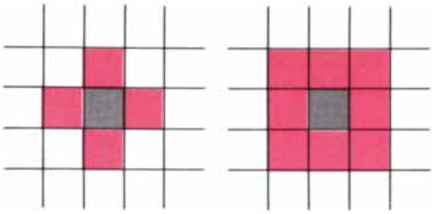


The "game of life" evolves on the screen of Tommaso Toffoli's cellular-automaton machine



A pattern of dendrites is created by a cellular automaton with an asymmetric transition rule

© 1984 SCIENTIFIC AMERICAN, INC



The von Neumann and Moore neighborhoods

tern preserves its shape but shifts to the right by one cell with each time step. A subclass of rules called counting rules or totalistic rules seems to include specimens of almost all the observed varieties of cellular automata. With rules of this kind the new state of a cell depends only on the number of neighbors in a given state, not on their position. Many automata based on such rules have been investigated by members of the Information Mechanics Group of the Laboratory for Computer Science at the Massachusetts Institute of Technology. The group consists of Edward Fredkin, Norman Margolus, Tommaso Toffoli and Gérard Y. Vichniac.

One of the simplest counting rules is the parity rule, which assigns a cell a value of 1 if an odd number of the neighboring cells are 1's and otherwise assigns it a value of 0. The evolution of this system, when the rule is applied in the von Neumann neighborhood, was described in this space last October. Any starting pattern is replicated four times; the four copies are then replicated in turn, and so on.

Another class of counting rules are the "voting" rules, which give the center cell a value of 1 whenever the number of 1's in the neighborhood exceeds some threshold. Vichniac, in a paper presented at the Los Alamos meeting, points out that rules of this type yield models of percolation and nucleation, phenomena of importance in solid-state physics and other fields. Percolation is the term applied to the formation of an unbroken path across some space; for example, when a metal is dispersed in an insulating matrix, the conductivity of the composite depends on the probability of forming a continuous chain of metal atoms. Likewise the transmission of an infectious disease is possible only through an unbroken sequence of susceptible individuals. Nucleation is the process that initiates the growth of a crystal, the boiling of a liquid and similar events.

One transition rule that gives rise to percolation makes the center cell a 1 only if there are 1's in at least three out of the five cells that constitute the von Neumann neighborhood plus the center cell. The onset of percolation is extremely sensitive to the initial concentration of 1's. If the concentration is less than one-half, continuous chains of 1's spanning the array are not likely to form in the course of the evolution. At a concentration of one-half or greater the chains do appear, but the entire lattice still does not fill with 1's; islands of 0's remain in the final stable state. Nucleation, in which the array does fill solidly with 1's, is observed when the rule is changed to require only two out of five 1's. The critical concentration is .0822.

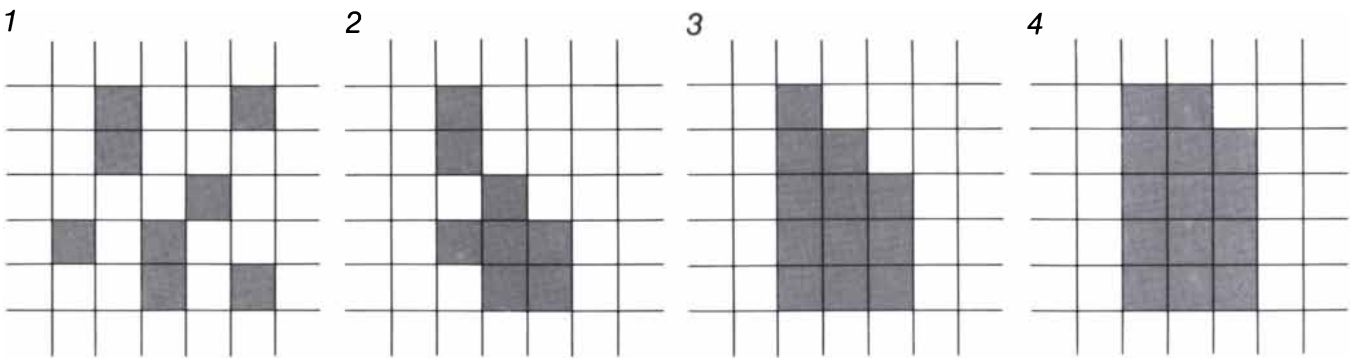
The Ising model is a conceptual tool of physics that seems superficially to be much like a cellular automaton. The model is a rectilinear lattice where each site has two possible values and interacts only with its four nearest neighbors. The model is often employed to describe ferromagnetic materials; each site represents an atomic spin that must point either up or down. Below a critical temperature (the Curie temperature) the spins tend to be aligned, so that the material is magnetized, but at higher temperatures they are more or less randomly distributed.

In October I discussed a version of the Ising model created with a spreadsheet program, whose lattice of cells lends itself naturally to cellular-automata studies, albeit a lattice with probabilistic rules to emulate temperature. I observed a curious phenomenon: at low temperature the spins did not assume a uniform alignment in one direction; instead they adopted a checkerboard configuration of alternating up and down spins. With each time step all the spins flipped. In a ferromagnet the checkerboard pattern is the configuration of highest energy and should therefore be unstable; it is the pattern characteristic of an antiferromagnet.

Vichniac had already discovered the problem and explained it. In the standard implementation of the Ising model only one spin is allowed to change in each iteration. It follows that when a particular site surveys its neighborhood, some of the spins it inspects are "old" ones and some are "new." Under these conditions the oscillating antiferromagnet cannot arise. It is only when all the spins are recalculated simultaneously that the high-energy antiferromagnet is favored. There are strategies for avoiding this "feedback catastrophe," but the lesson of larger significance is that the simplest intuitive correspondence between the Ising model and cellular automata is misleading.

Vichniac and others in the M.I.T. group point out that cellular automata have a status fundamentally different from that of other physical models. The commonest device for building a mathematical model of the natural world has long been the differential equation, which can describe the change in some quantity as a function of position and time. For example, Maxwell's equations give the variation in the value of an electromagnetic field from point to point and from moment to moment. All the quantities in such equations are continuous: they vary smoothly. A cellular automaton, on the other hand, is a fully discrete system. Space is not a continuum but an array of cells; time too is broken down into discrete steps, and whereas the magnitude of a field can vary over a continuous range, the cells of a cellular automaton can have only a finite number of states.

Of course, real space and time and many physical variables are thought to be continuous rather than discrete (at least at the scale commonly considered). It does not follow, however, that differential equations yield inherently superior models of nature. Often it is not the precise numerical value of a variable that is significant but only its overall size, as in whether a particular point in



Evolution of a cellular automaton under the two-out-of-five voting rule

© 1984 SCIENTIFIC AMERICAN, INC

RENAULT FUEGO WATCHWORD: RACY

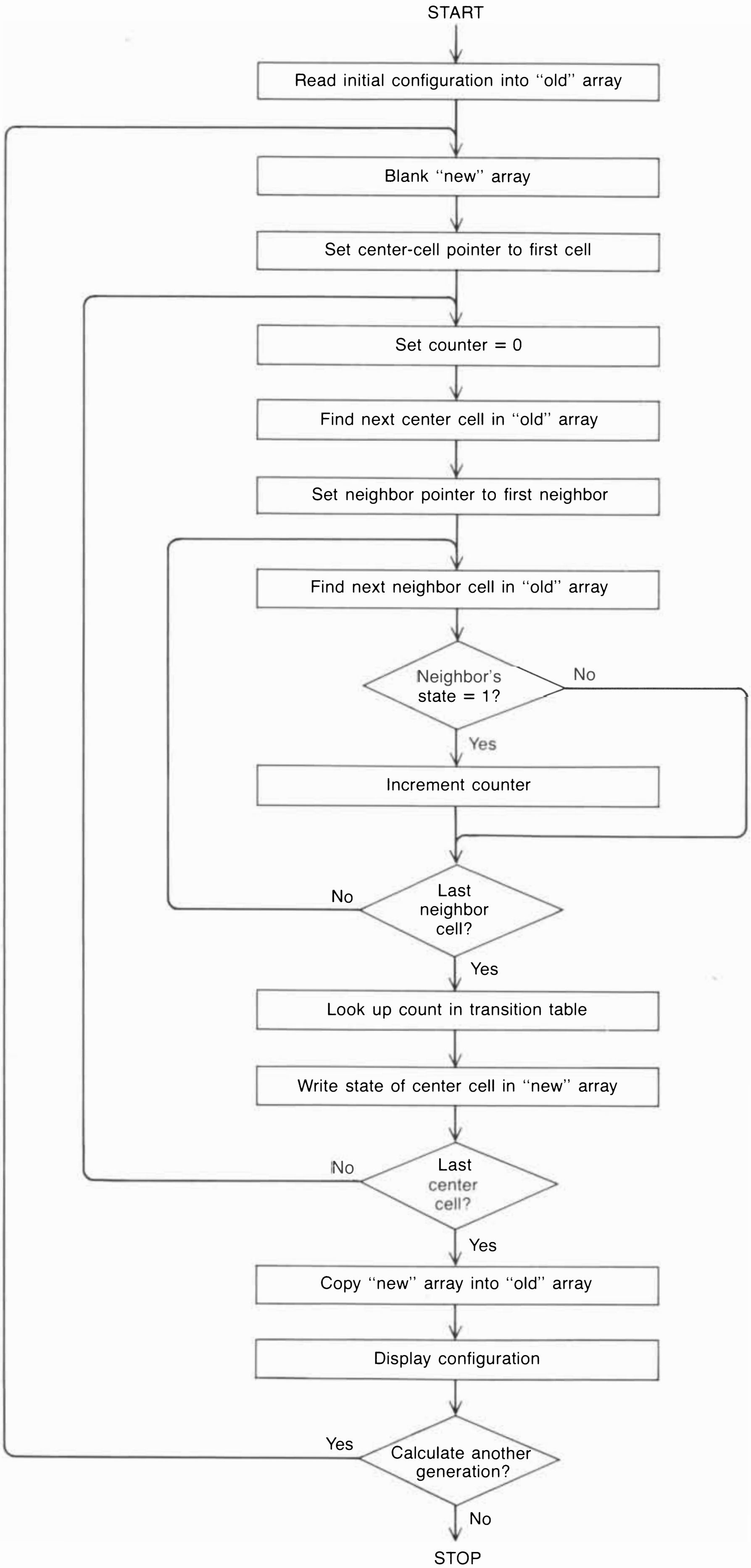
The racy Renault Fuego—descended from the winningest turbos in Formula One history. Too racy for average tastes? Definitely. But when you consider the 31% boost in power, the laser-calibrated cockpit with tach-to-turbo gauge telemetry and a seductive shape that literally cleaves the air, Fuego Turbo just might be right for you. Fuego's competition breeding includes a 5-speed transmission, Bosch L-Jetronic fuel injection and Michelin TRX radials. Plus creature comforts like AM-FM stereo, fully adjustable reclining bucket seats and air conditioning are standard. Renault Fuego Turbo. The raciest **$11,095*** on the street. And Renault Fuego, the one to watch from **$8,695***.

*Manufacturer's suggested retail price. Tax, license, destination charges and other optional and regionally required equipment extra. **Sold by American Motors**

RENAULT

THE ONE TO WATCH

© 1984 SCIENTIFIC AMERICAN, INC



Algorithm for cellular automata based on "counting" or "totalistic" transition rules

a growing snowflake is ice or water vapor. Cellular automata make this discreteness explicit on a digital computer. In addition their temporal evolution can be computed exactly; there is no need for approximations. Furthermore, they can make far more efficient use of the digital computer's resources.

A program for simulating a cellular automaton can be written for even the smallest computer. Indeed, Per Bak of the Brookhaven National Laboratory has recently argued in *Physics Today* that many simulations in physics can be done more effectively and more cheaply with a small personal computer than they can be with more powerful shared facilities. The example he chose to illustrate was a simulation of the three-dimensional Ising model, done with a Commodore VIC-20 computer at an estimated cost of $4.

The most straightforward cellular-automaton program simply embodies the method one would be likely to adopt in carrying out the procedure by hand with graph paper. First an array of cells is established, with each cell being represented by a memory element in the computer. For each time step the program must attend to every cell in turn, examine its neighbors and calculate the appropriate value for the cell's next state. The calculation itself is conveniently done by looking up the value in a table. If only counting rules are considered, the table needs only one entry for each possible number of "on" cells. When other kinds of rules are allowed, the table can become quite elaborate.

A few subtleties must be kept in mind when one writes such a program. Most important is the need to avoid altering the content of a cell before its value has been checked by all the other cells to which it is a neighbor. The easiest way of meeting this requirement is to maintain two copies of the array; the program examines one copy to determine the current state of the neighborhood and enters the result of its calculation in the other copy. Boundary conditions must also be defined. Ideally the array would be infinite, but that is clearly impractical. A common technique is to effectively join the edges of an array, so that cells on opposite edges become neighbors. In one dimension an array of this kind is topologically a circle and in two dimensions it is a torus; although it is finite, it has no boundaries.

A program of the kind described above, running on a general-purpose digital computer, is a sequential procedure that simulates the actions of an array of many computers running simultaneously. Far better would be an actual network of multiple computers with the structure of the cellular array. Building such a machine is by no means out of the question: the individual computers

© 1984 SCIENTIFIC AMERICAN, INC

C O L O R T R A K

T W O T H O U S A N D

© 1984 SCIENTIFIC AMERICAN, INC

ANNOUNCING
A PROUD ADDITION
TO YOUR FAMILY.

© 1984 SCIENTIFIC AMERICAN, INC

The next addition to your family could be the bright little newcomer in the growing family of IBM® personal computers.

Name: PC*jr.* Weight: 12 pounds. Heritage: more than 30 years of computer experience.

"Junior" is a powerful tool for modern times. Yet it's simple enough for a child to enjoy.

BRINGING HOME BABY

It's a big day when PC*jr* comes home.

The surprises begin the moment you open the carton.

Surprise #1 is the IBM "Freeboard"—a keyboard that doesn't need a connecting cord. The Freeboard frees you to move around and relax.

Then there's the Keyboard Adventure—an instructional exercise for first-time users. It's built into the computer and explained step-by-step in the Guide to Operations. It will help anyone begin learning as soon as PC*jr* is hooked up to a TV set.

In systems equipped with a diskette drive, there's a program that lets you explore computer fundamentals at your own pace, with PC*jr* as your teacher.

And to get you off and running from the very first day, a sample diskette with eleven useful mini-programs (ranging from a spreadsheet for monthly expenses to a word game and a recipe file) is also included.

But there are still more surprises.

FAMILY COMPUTING MADE EASY

Many IBM software programs written for other IBM personal computers will run on PC*jr.* And inexpensive new ones written especially for PC*jr* are being released.

An easy-to-use diskette word processing program, for example, uses pictures as well as words to guide you along. A comprehensive IBM home budget program makes keeping track of money easier. There's also a selection of educational programs for children at home and at school.

And when the work is finished (or perhaps before), the fun can begin. Just slip in a game cartridge and stand back.

GROWING UP WITH JUNIOR

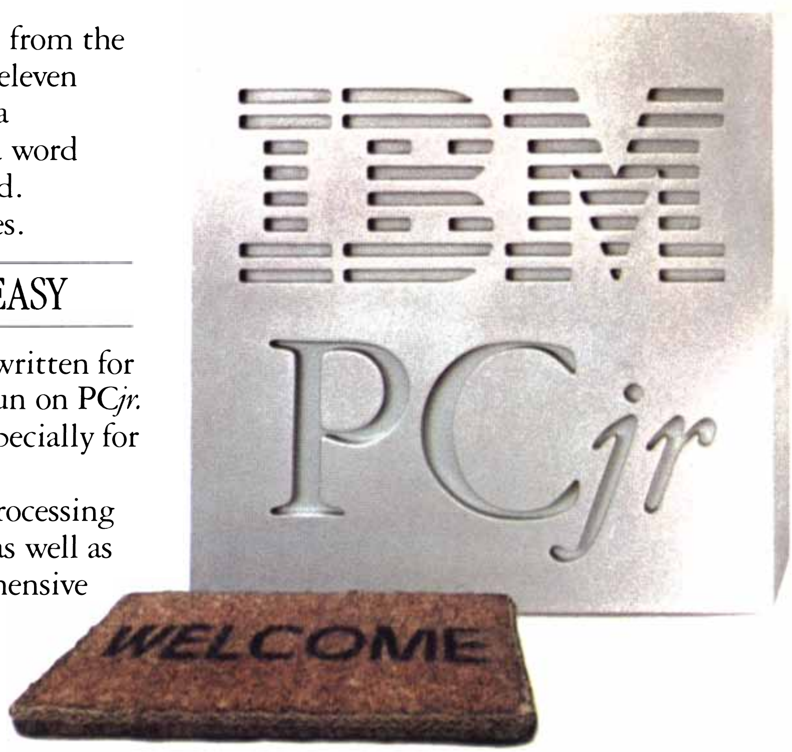
Add a printer. A diskette drive. An internal modem for telecommunications. Increase user memory from 64KB to 128KB. With these and other add-it-yourself options, even the lowest-priced PC*jr* can grow up *real* fast.

PC*jr* is a powerful tool for home, school or college. With its optional carrying case, it's a powerful tool anywhere you care to take it.

SEE JUNIOR RUN

Junior's starting model includes a 64KB cassette/cartridge unit and Freeboard for about $700. A 128KB model with diskette drive is about $1300. (Prices apply at IBM Product Centers. Prices may vary at other stores.)

Your local authorized IBM PC*jr* dealer proudly invites you to see this bright little addition to the family. For the store nearest you, just call 1-800-IBM-PCJR. In Alaska and Hawaii, 1-800-447-0890.



© 1984 SCIENTIFIC AMERICAN, INC

Digital read-outs. Light-emitting diodes. On-board diagnostics. They're elements of the advanced electronics designed by GM scientists and engineers for 747's.

Now, after years of work, many of the same experts are building the same technology into new GM cars and trucks.

In some cases, aircraft solutions were appropriate and we adapted them. When they weren't, we went back to the drawing board to create new technology.

Not just to give you "gee whiz" gadgetry. But to help your GM car run more efficiently and intelligently. And to help you keep it running.

We believe in taking the extra time, giving the extra effort and paying attention to every detail. That's what it takes to provide the quality that leads more people to buy GM cars and trucks than any other kind. And why GM owners are the most loyal on the road.

That's the GM commitment to excellence.

GM

Nobody sweats the details like GM.

© 1984 SCIENTIFIC AMERICAN, INC

would be so simple that many of them might be fitted onto a single semiconductor chip. The fact that only nearby computers need to communicate with one another would also reduce the complexity of the device. Toffoli has estimated that such a processor might operate faster than a general-purpose computer by a factor of a million or even a billion. Preliminary work on computers of this kind is under way at the Massachusetts Institute of Technology and the Thinking Machines Corporation of Waltham, Mass.

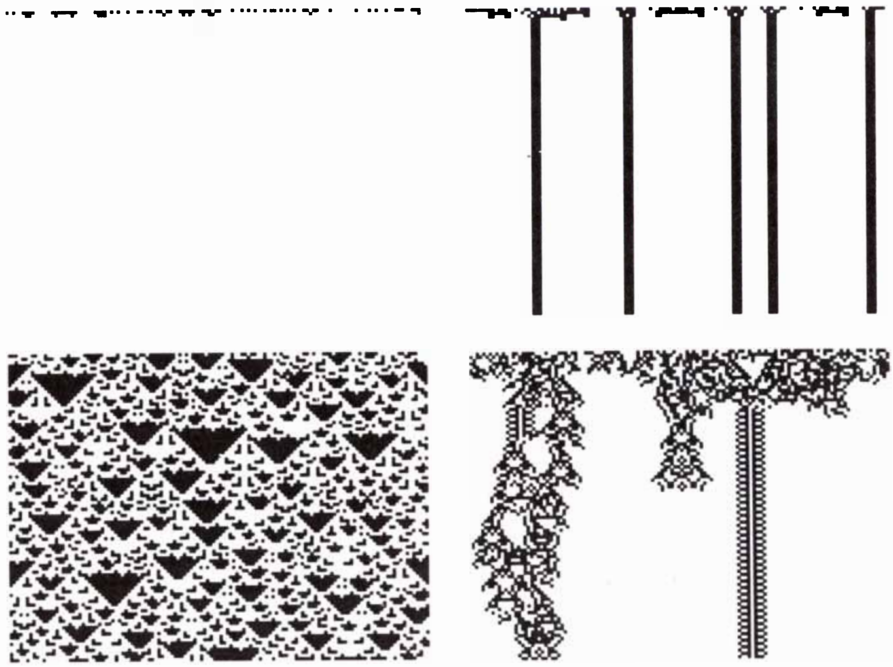
In lieu of a special-purpose chip, Toffoli has constructed a dedicated cellular-automaton machine out of standard microelectronic components. Calculations are done serially rather than for all the cells at once, but because the device is finely tuned to a single kind of calculation, it is roughly 1,000 times faster than a general-purpose computer. The machine itself consists of a few printed-circuit boards mounted in a frame; it is connected to a color display and is controlled by another small computer, an Atari 800.

Toffoli's cellular-automaton machine provides an array of 256 by 256 cells, each of which can have up to 256 states. The state of every cell is recalculated 60 times per second. Watching a system evolve at this rate is quite different from watching a slower device. Instead of a sequence of still photographs one sees a motion picture. The game of life no longer appears as a stately progression of abstract patterns; it is more like a view through the microscope of bacteria and protozoa swimming, spinning, breeding, eating and being eaten.

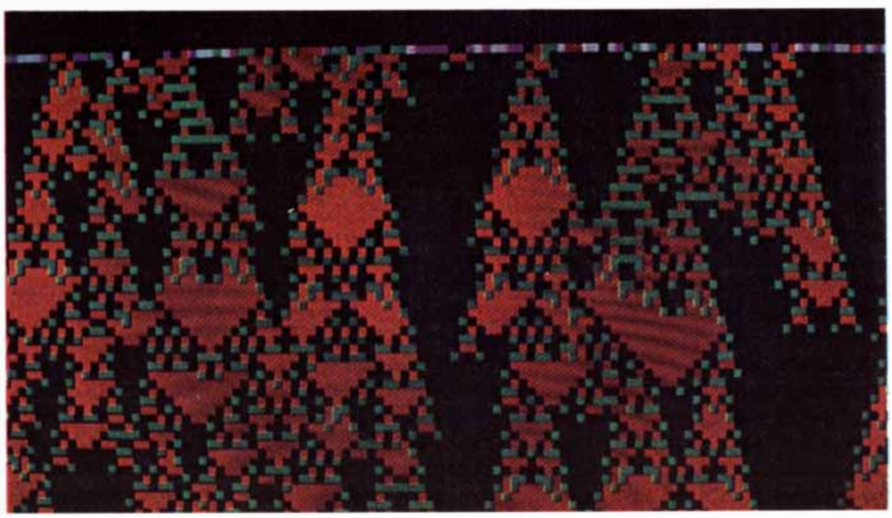
A one-dimensional cellular automaton is much less demanding of computer resources, both spatial and temporal, than a two-dimensional system. Writing a program for the one-dimensional system is also easier. The linear array has still another advantage over the planar one: because of the simpler geometric structure, there is more hope of gaining an analytic understanding of the automaton's evolution. In the past two years Stephen Wolfram of the Institute for Advanced Study has undertaken to do just that.

A single generation of a one-dimensional array is merely a line of cells, but successive generations can be plotted next to one another. In this way a two-dimensional pattern is formed that has one spatial axis and one time axis, and the entire evolution of the system can be taken in at a glance.

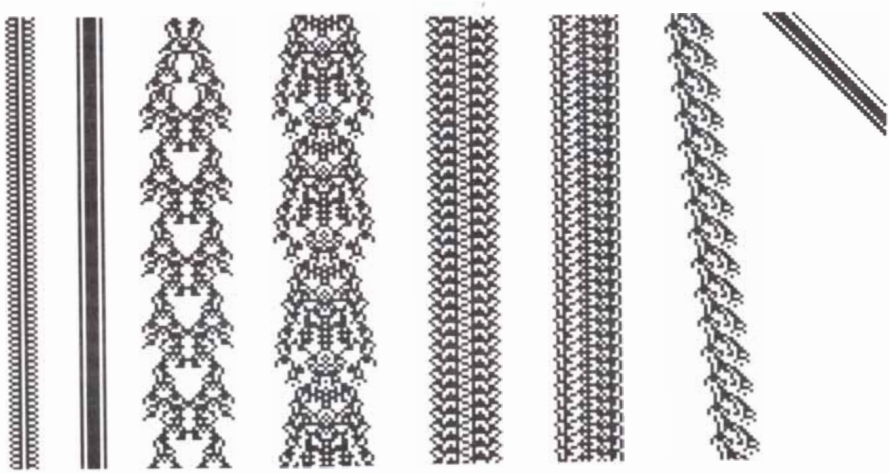
Wolfram has found that all the transition rules he has investigated so far can be put into just four classes. Class 1 consists of those rules whose evolution leads to a stable and homogeneous state; for example, all cells might take on a value of 0 or of 1. Class 2 rules give rise



The four classes of totalistic rules in one dimension



Successive states of a Class 4 one-dimensional automaton



Some components of a possible universal computer

© 1984 SCIENTIFIC AMERICAN, INC

How Exxon's advanced the refining of

FLEXICOKING converts 99% of even the poorest feedstocks into high-quality fuels.

A key challenge facing the oil industry today is to ensure the most efficient conversion of available crude oil supplies into the highest value products. This is increasingly difficult because, while demand is shifting toward light, cleaner-burning products, available crudes are becoming heavier.

In addition, some heavy crudes contain high concentrations of impurities, such as sulfur, nickel and vanadium, so that key catalytic refining processes cannot readily handle them. These contaminants "poison" process catalysts just as lead deactivates an automobile's catalytic convertor. Future unconventional feedstocks, such as bitumen from tar sands and shale oil, are likely to be even harder to process.

A Powerful New Refining Tool

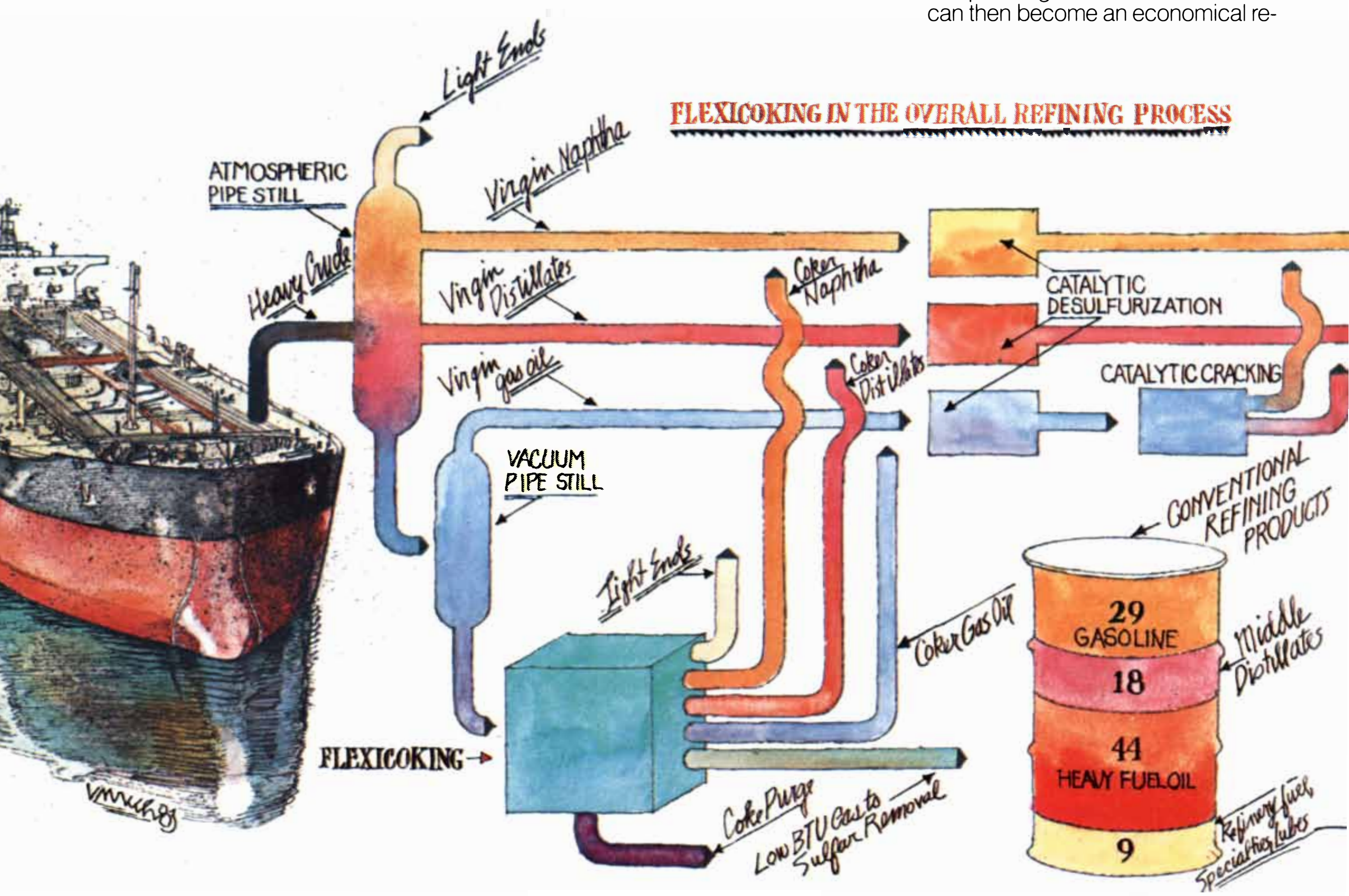
But scientists and engineers at Exxon Research and Engineering Company have a unique solution—an extremely rugged process called FLEXICOKING that converts in excess of 99 percent of the worst possible feedstocks into high-quality liquid and gaseous fuels.

FLEXICOKING relies on thermal cracking rather than catalytic cracking to convert heavy feedstocks to lighter products. Thermal conversion is not new. In the 50's, Exxon developed FLUID COKING to upgrade heavy feedstocks. FLEXICOKING further extends this pioneer thermal process with one major improvement.

In FLUID COKING, heavy feed is sprayed into a large reactor containing very hot, granular coke. The upward flow of hydrocarbon vapors and injected steam suspends the finely divided particles, permitting the hydrocarbons to contact the coke on all sides. The heavy hydrocarbon molecules fracture as they contact the hot coke. This results in additional gasoline and middle distillate liquids. However, about 30 percent of the heavy feedstock is rejected as solid coke, which represents almost 20 percent of the energy in the original feed.

Replacing Expensive Refinery Fuel

FLEXICOKING represents a key advance by converting the coke into a clean-burning low-BTU fuel gas (LBG) in a separate gasification reactor. This LBG can then become an economical re-



© 1984 SCIENTIFIC AMERICAN, INC

technology is revolutionizing difficult crudes.

placement for refinery fuels costing three to four times as much, or it can be sold to local industries. By successfully integrating coking and coke gasification into a single process, Exxon researchers have reduced the volume of purge coke to about one percent of the heavy feedstock. Metals are concentrated in this coke purge.

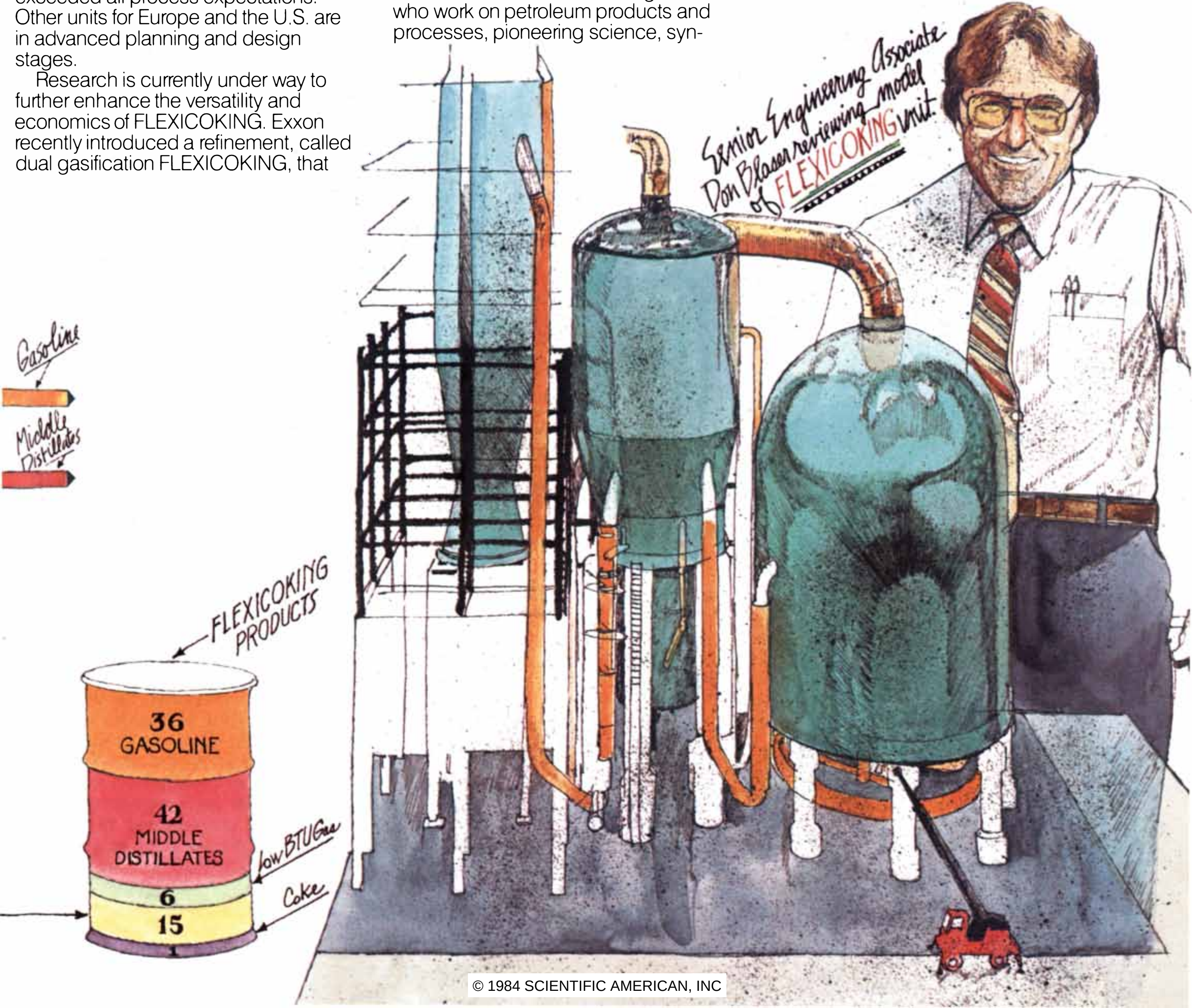
The first commercial FLEXICOKING unit went into operation in Japan in 1976. A second unit was started in Venezuela at the end of 1982 and a third in California in early 1983. Each unit has met or exceeded all process expectations. Other units for Europe and the U.S. are in advanced planning and design stages.

Research is currently under way to further enhance the versatility and economics of FLEXICOKING. Exxon recently introduced a refinement, called dual gasification FLEXICOKING, that enables refiners to convert part of the LBG into a nitrogen-free synthesis gas. This gas can be used as a source of hydrogen or as a feedstock for making chemicals.

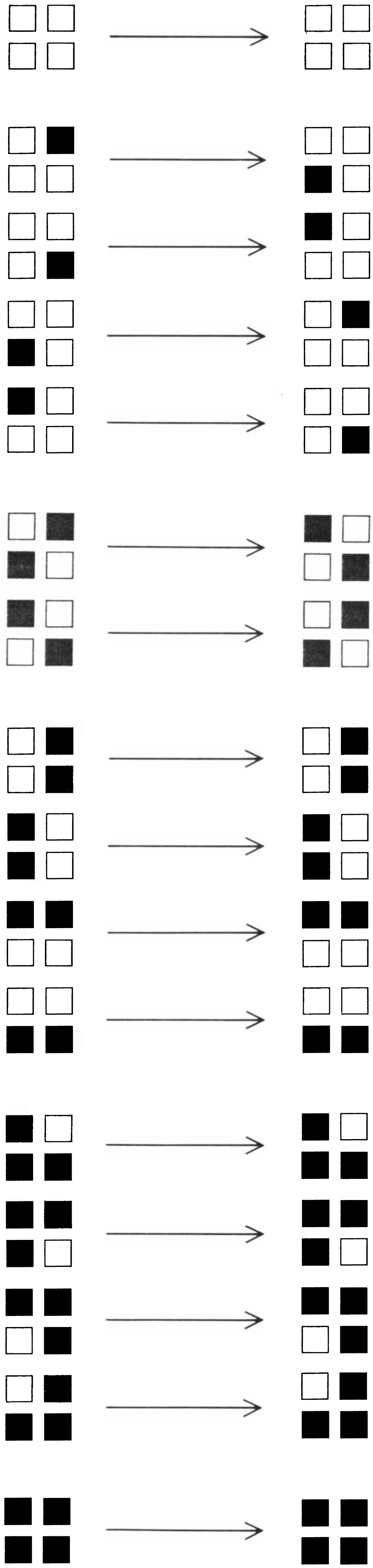
Exxon Research and Engineering Company

FLEXICOKING is but one product of the research and engineering programs at Exxon Research and Engineering Company. A wholly owned subsidiary of Exxon Corporation, ER&E employs about 2,000 scientists and engineers, who work on petroleum products and processes, pioneering science, synthetic fuels, and the engineering required to develop and apply new technology in the manufacture of fuels and other products. For more information on FLEXICOKING or ER&E, write to E.E. David, Jr., President, Exxon Research & Engineering Company, Room 100, P.O. Box 101, Florham Park, NJ 07932.

EXXON



© 1984 SCIENTIFIC AMERICAN, INC



Transition rule for a billiard-ball computer

to simple structures that are either stable or periodic but in either case remain isolated from one another. The rules in Class 3 create chaotic patterns, although not random ones. In Class 4 are the few transition rules that generate structures of substantial spatial and temporal complexity.

Wolfram conjectures that the one-dimensional cellular automata may be the simplest well-defined systems capable of complex self-organizing behavior. In nature many continuous dynamical systems have such a capability: beginning in a random initial state, they evolve a highly ordered structure. (The snowflake is an example.) The evolution can be explained in terms of attractors, which seem to draw the system toward a subset of all the possible configurations.

A parallel has been established between the classes of cellular automata and the kinds of attractors observed in physical systems. A Class 1 automaton is analogous to a continuous system with the simplest attractor: a limit point, which invariably brings the system to the same final state. The evolution of a Class 2 automaton is rather like that of a system with a limit cycle, a set of configurations that is repeated indefinitely.

Class 3 automata, with their chaotic patterns, can be associated with the more interesting entities called strange attractors, which are characteristic of physical phenomena such as the onset of turbulent flow. In a system governed by a strange attractor evolution proceeds toward a subset of all the possible configurations, but the subset can have an exceedingly intricate structure. When the set is visualized as an array of points in space, it is in many cases a fractal, a geometric figure with a fractional number of dimensions.

The distinctions between the classes of automata can be made clearer by considering a simple experiment. Suppose a cellular automaton is started in some randomly chosen initial configuration and allowed to evolve for many time steps; the final state is then noted. Now return to the starting configuration, change the value of a single cell and allow the system to evolve for the same number of steps. What effect will the small change have on the final state? In a Class 1 automaton there is no effect at all: a Class 1 system reaches the same final state no matter what the initial state is. A Class 2 automaton may show some effect, but it is confined to a small area near the site of the change. In a Class 3 system, however, altering a single cell can set up a disturbance that propagates throughout the array.

The Class 4 rules are the rarest and the most intriguing. Some quite simple transition functions fall into this class; for example, in the neighborhood defined to include the center cell and the two cells on each side of it, the rule stating that the center cell is a 1 if either two or four cells in the neighborhood are 1's leads to Class 4 patterns. Sensitivity to small variations in the initial conditions is even greater in Class 4 than it is in Class 3. It is conjectured that in predicting the future state of a Class 4 automaton there can be no general procedure more efficient than allowing the automaton itself to compute the state.

A related conjecture has even grander scope: it suggests that Class 4 automata may qualify as universal computers. The Turing machine is the most familiar device of this kind; if a function can be computed at all, a Turing machine can presumably do it. Other computers can be proved universal by showing that they are equivalent to a Turing machine. Several two-dimensional cellular automata (including the game of life) have been shown to be universal computers, and a proof has also been given for a complicated one-dimensional system with 18 states per cell. The Class 4 automata would be the simplest universal computers known. Most of the essential components have been identified. One important missing element is a clock: a structure that issues a train of pulses at regular intervals, like the glider gun in the game of life.

The view of cellular automata as computers suggests that their self-organizing behavior can be characterized in terms of their computational capabilities. Thus, for example, sets of configurations generated by the evolution of a cellular automaton can be thought of as a formal language. Each configuration is considered as a word in the language, formed from a sequence of symbols representing the cellular-automaton site values according to a set of grammatical rules. Wolfram has shown that the configuration generated by any cellular automaton after a finite time can be described by a simple class of formal languages known as regular languages. For any of these regular languages it is possible to find a simplest grammar. That grammar gives a minimal description of the cellular-automaton configurations, and its size can be taken to measure the complexity of the configurations. For cellular automata of Class 1 and Class 2 the complexity tends to a finite limit at large times, so that the structures generated by these systems are described by regular languages. For cellular automata of Class 3 and Class 4, however, the complexity usually increases rapidly with time, and it appears that more complicated formal languages are required to describe the large-time behavior of such systems.

There is a special class of cellular automata that are said to be reversible, or invertible. From any starting configu-

© 1984 SCIENTIFIC AMERICAN, INC

ration a reversible automaton can be allowed to evolve for any number of time steps, then stopped and run in reverse, and it will return to its exact initial state. The patterns formed by a typical reversible automaton have a qualitatively different appearance from those characteristic of a nonreversible automaton. In particular, if the pattern is initially random, it tends to remain random; no self-organizing structures appear.

A necessary condition for reversibility is that the transition rule be deterministic in both the forward and the backward directions, that is, every possible state of a neighborhood must have both a unique successor and a unique predecessor. The game of life is nonreversible because the predecessor of a state cannot be identified unambiguously: if a cell is currently "dead," for example, in the preceding generation it could have had any number of living neighbors other than three. A systematic way of creating reversible transition rules was invented by Fredkin and has been further investigated by Margolus. The essence of the method is to let the next state of a cell depend on the two previous states of the neighborhood. The state at time $t+1$ is given by any function of the neighborhood at time t minus the state at time $t-1$. The reversal is then straightforward: the state at time $t-1$ must be given by the state at time t minus the state at time $t+1$.

Because of the requirement of bidirectional determinism, there can be no attractors in the evolution of a reversible automaton. The presence of an attractor implies that many initial states evolve along paths that merge with one another; in the reversed evolution the merger points would become branch points, where determinism would fail. Similarly, a reversible cellular automaton can never enter or leave a loop, or cycle of states, because again a branch point arises in one direction or the other. Because attractors and the associated self-organizing patterns are excluded, it may seem that reversible transition rules would give rise to quite dull cellular automata, but other features of the systems offer compensating points of interest. Most notably, the information content of a pattern of cells in a reversible automaton turns out to be a conserved quantity (one that cannot increase or decrease in the course of the automaton's evolution). This property makes the reversible systems valuable models of computation.

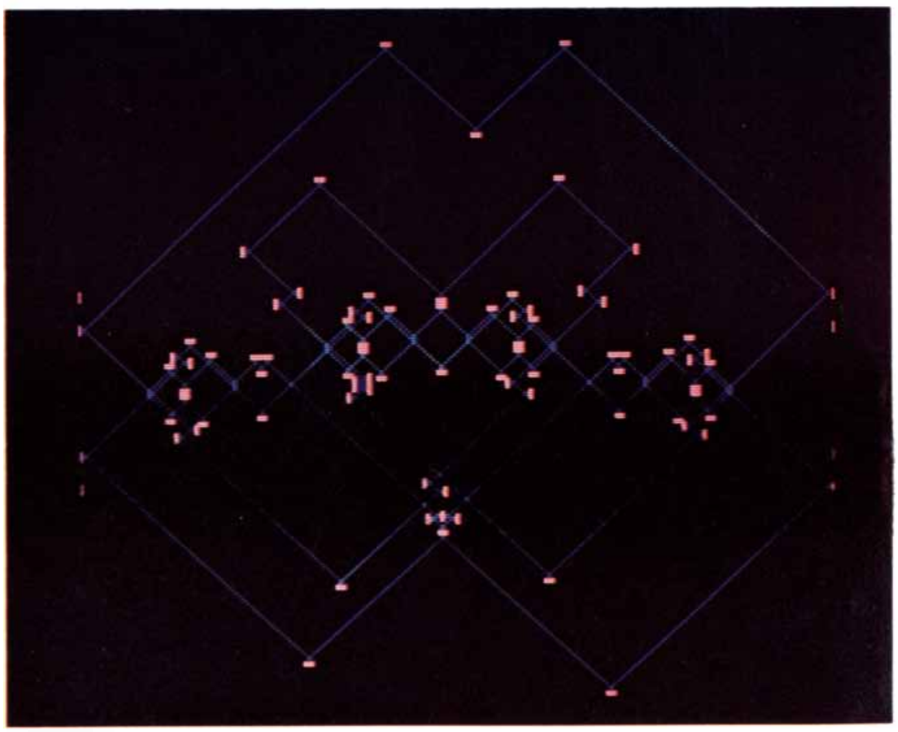
Margolus has constructed a cellular-automaton computer based on an imaginary mechanical system first discussed by Fredkin: the billiard-ball model of computation. In the model bits of information (1's and 0's) are carried by idealized billiard balls that move without friction and rebound from one another and from other obstacles with perfect elasticity. The presence of a ball at a designated position represents a binary 1 and the absence of a ball represents a binary 0. Through a clever arrangement of bumpers it is possible to create various logic gates analogous to those of an electronic computer. In an AND gate, for example, one billiard ball passes through the output region (and thereby registers a binary 1) only if two balls approach the gate simultaneously along specific trajectories.

Margolus' cellular-automaton version of the billiard-ball model is an example of a simple but somewhat unusual reversible transition rule. Cells are considered not individually but in blocks of four; every possible pattern within a block is transformed into a unique product pattern. The rule is designed so that a single 1 in a background of 0's propagates along one of the four diagonal directions of the lattice at a speed of one cell per time step; the isolated 1 is the embodiment of a billiard ball. A solid block of four 1's remains unchanged and acts as a perfect reflector. When the model is set going on Toffoli's cellular-automaton machine, the "billiard balls" streak across the display screen in intricate interwoven patterns. Watching this orderly (if frenetic) motion, it is hard to keep in mind that the program has no representation of the balls' paths but merely applies a single rule to all the cells.

The billiard-ball model and its cellular-automaton implementation have an important bearing on the theory of computation. It has been conjectured that any computer must have components that dissipate both energy and information; according to this argument, there is a thermodynamic limit to the efficiency of a computer just as there is to the efficiency of a heat engine. The supposedly inevitable losses of information and energy result directly from the irreversibility of the computational process. (When a computer adds the numbers 5 and 3 to get 8, the procedure cannot be reversed because there are infinitely many numbers that could have been added to get the same result.)

Fredkin, Toffoli and Margolus point out that the billiard-ball model offers a counterargument to the notion of inevitable dissipation. In the billiard-ball computer no information is lost. Indeed, the billiard balls themselves cannot be created or destroyed, and all the information that defines their initial pattern is preserved as the system evolves. The inputs to an addition operation can be recovered simply by reversing the trajectories. In principle the billiard-ball computer could operate with no internal power consumption.

The connection between physics and computing has been made with particular clarity by Toffoli in a statement that could be read as a description of the largest of all cellular automata. "In a sense," he writes, "nature has been continually computing the 'next state' of the universe for billions of years; all we have to do—and, actually, all we *can* do—is 'hitch a ride' on this huge ongoing computation, and try to discover which parts of it happen to go near to where we want."



The billiard-ball computer in action

© 1984 SCIENTIFIC AMERICAN, INC

What should from a working Hewlett-

"At American Cyanamid's Medical Group, HP computer systems will cut annual production costs by $4.4 million.

"Because the HP systems deliver more accurate analyses of test results, we're saving $250,000 in raw materials."

Dr. Bill Zeitz, Director of Medical Group Information Services, says: "American Cyanamid's Medical Group is replacing mainframe use in its manufacturing plants with six HP 3000 computers. The HP systems automate materials and production management, factory order entry, and quality assurance programs.

"Because the HP 3000s can communicate with each other and with our corporate IBM mainframe, we've eliminated the need for additional communications links, thus cutting annual computing costs by $1 million. And by 1985, with all HP systems up and running, we expect to realize an annual savings of $4.4 million.

"We needed a lot-control program tailored to the pharmaceutical industry—one that would track and process information by product lot number. In just four months, an HP task force developed a software product that met our requirements, and provided a lot-control solution for the entire process industry."

© 1984 SCIENTIFIC AMERICAN, INC

you expect partnership with Packard?

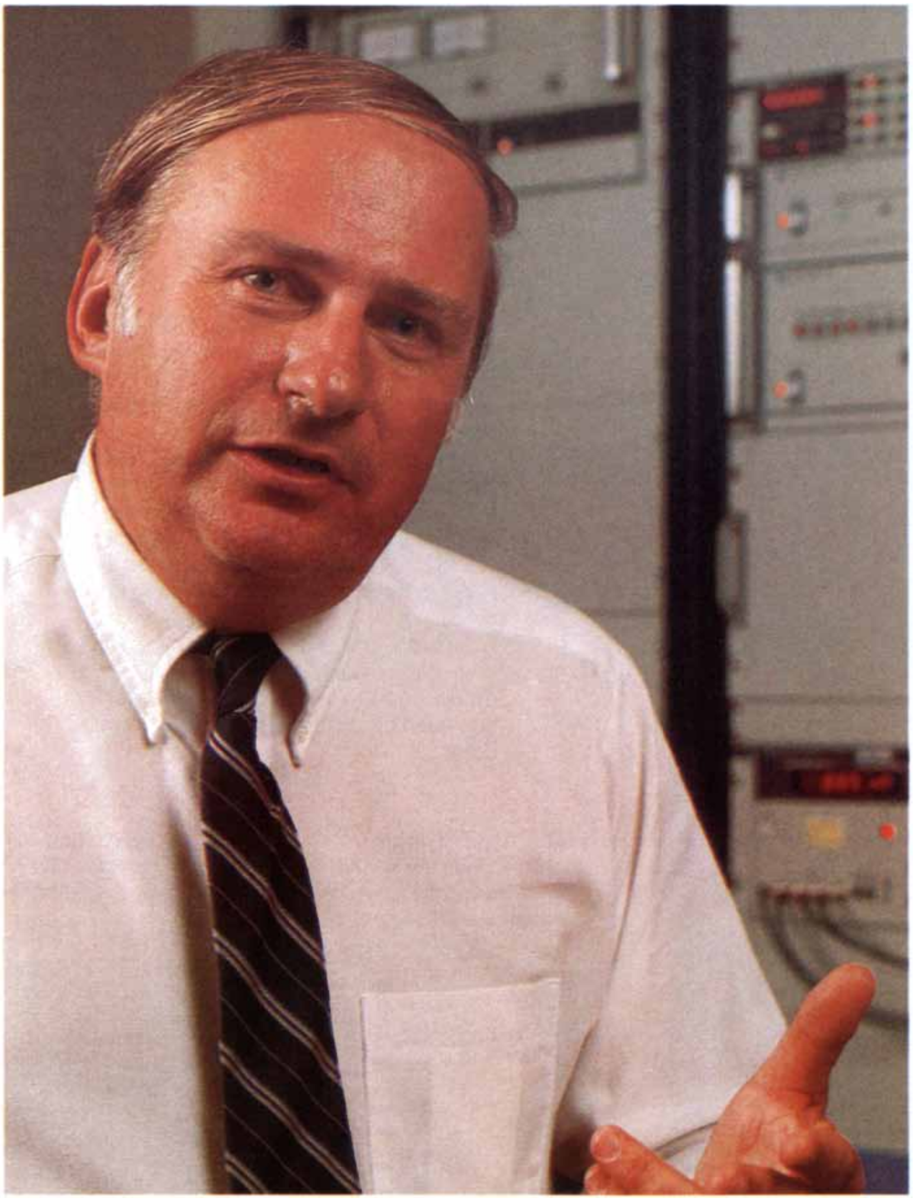
'At M/A-COM, HP computer-based instrument systems speed final test 12-fold.

"By saving test time and increasing throughput of our microwave components, each of our 45 HP computer-based systems paid for itself in less than a year."

M/A-COM, Inc., is a major manufacturer of electronic components for commercial telecommunications and defense markets. Dr. Alan Carlson, Manager of the Automatic Test Equipment Group in Burlington, Mass., says: "We assemble special-purpose systems to test hundreds of M/A-COM products. To automate, we needed precise instruments plus computer power.

"HP was the only vendor to offer us a complete solution. With 45 HP computers driving more than 400 HP instruments, we've totally automated product tests.

"HP application engineers work with us from the start to design, develop, and implement new test systems. Once a new system is installed, we rely on HP's support teams to keep our testing networks up and running. HP's technical expertise continues to play a major role in our success."



Productivity. Not promises.

HEWLETT PACKARD

0002403

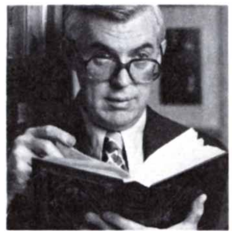
© 1984 SCIENTIFIC AMERICAN, INC

We asked these three experts to put a price on this edition of Dickens' DAVID COPPERFIELD.

$12.50.

"The raised spine is a giveaway. That's quality binding. **$30.**"
KEN McCORMICK,
Sr. Consulting Editor,
Doubleday.

"The illustrations are superb. Excellent typography. **$35.**"
CLIVE BARNES,
Cultural critic, New York Post.



"The paper is fine mill. The end pages an unusual design. A first-class job. **$35.**"
PHILIP LYMAN,
Gotham Book Mart.

WRONG. WRONG. WRONG.
The price, of course, is only $12.50 for *David Copperfield* and every other volume of the new Collector's Library of the World's Great Books.

Again and again, Franklin Library, the world's leading publisher of fine books, has seen its remarkable new Collector's Library confound the experts.

How can these authorities be so right in their praise and so wrong in the price?

Let's take a close look and see why.

Each of the fifty volumes is individually designed; each cover is a work of art, each binding distinct in size, grain, color.

(The effect in your room is uncanny. Each volume different but harmonious. A luxurious library that is a tribute to your taste.)

Each spine is hubbed in the great tradition of book binding.

Each endsheet (inside covers) is an original design.

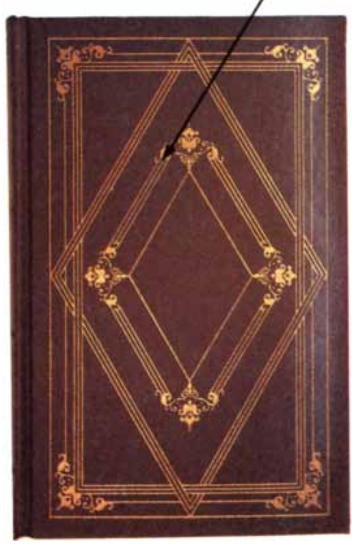
The page edges are gilded; the paper, itself, is crisp, opaque, specially milled and acid free to last for generations.

Many artists were commissioned for the project. And the illustrations are magnificent. (The color paintings by Thomas Rowlandson in TOM JONES make us tingle with pleasure.)

Distinctive endpapers, specially designed for the collection, and varied among the volumes.

Superb illustrations, many specially commissioned, including a number in full color.

Each binding is individually designed: the copper cover dies are hand finished.



Hubbing—a valued characteristic of the traditional bookbinder's art.

Page edges gilded with a tarnish-free finish for both protection and beauty.

Specially milled, acid-free paper that will endure for generations.

© 1984 SCIENTIFIC AMERICAN, INC

How did we do it?

The typography is classic, the printing first rate, the contents immortal and priceless.

Yet the price for each volume is only $12.50, less, as you know, than ordinary hardcover books.

It is a publishing conundrum with a simple, albeit spectacular answer. Franklin Library has made a publishing breakthrough of which it is very proud.

Here is the genius of mankind.

Eternal classics, from Homer's ODYSSEY to Melville's MOBY DICK.

The chilling tales of Edgar Allan Poe. The romance of Emily Bronte's WUTHERING HEIGHTS. The earthy humor of CANTERBURY TALES by Chaucer. The transcendent dramas of Shakespeare (illustrated by Henry Fuseli), Ibsen, Chekhov. Great English and American poetry.

Fifty volumes in all. A library of genius that belongs in every home.

Here is adventure for the mind, for yourself and for your children. Inspiration and pleasure more abiding than television, more memorable than movies. At a price that makes it almost mandatory. (And *that* is Franklin Library's intention.)

Subscribe without risk.

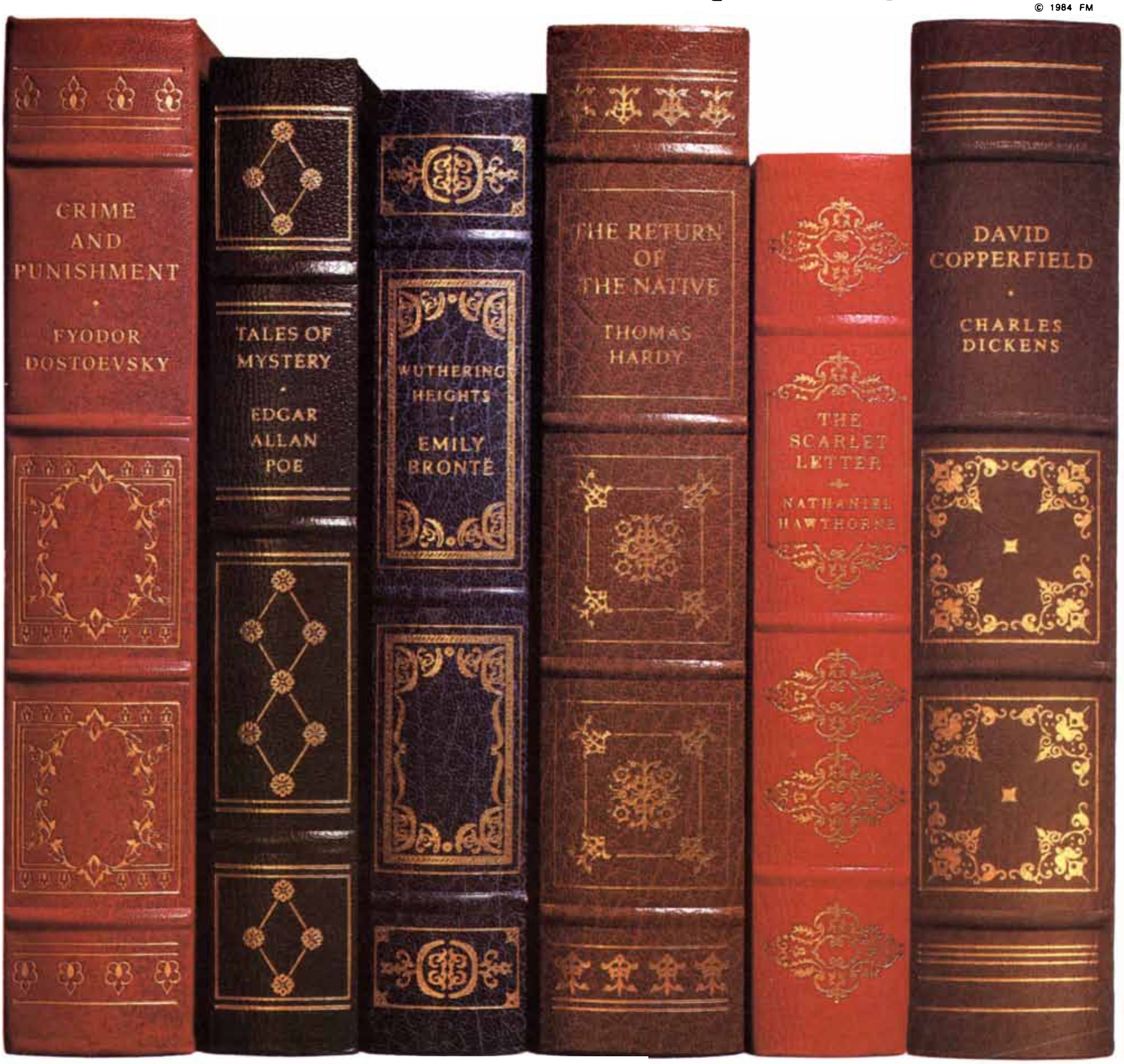
As a subscriber, you build your collection one book each month, and payment is on the same convenient basis. The subscription price is a remarkable $12.50 per volume.

A *guaranteed* price, if your subscription is mailed by May 31, 1984. You need send no payment now.

You may cancel your subscription at any time, upon 30 days written notice, and return any book for any reason, within 30 days.

The Collector's Library of the world's great books. A revolution in publishing.

© 1984 FM



© 1984 SCIENTIFIC AMERICAN, INC

BOOKS

Volcano skies, art of the chart, nuclear munitions, the editing of DNA, apples

by Philip Morrison

KRAKATAU 1883: THE VOLCANIC ERUPTION AND ITS EFFECTS, by Tom Simkin and Richard S. Fiske. Smithsonian Institution Press ($25). SUNSETS, TWILIGHTS, AND EVENING SKIES, by Aden and Marjorie Meinel. Cambridge University Press ($29.95). When the hush-hush technical library was deviously assembled for wartime Los Alamos, two titles were included by the witty librarians that might have brushed the limits of discretion. One was the science fiction novel *Krakatit,* by the Czech novelist Karel Capek, who imagined a chemist hero capable of coaxing explosive reactions out of almost any raw material. The title celebrates the chemist's most alarming compound, for which he became the cynosure and pawn of governments. The other title was a matter-of-fact, although sumptuous, 1888 volume of the Royal Society, which carried its expert committee's compilation of material on the Krakatau eruption itself. There is no doubt that the eruption, with its catastrophic loss of life, its physical effects noticed worldwide and its full accounting by unprecedentedly prompt stories in the press, borne everywhere by the new cable links, caught the imagination of a generation.

The centennial of Krakatau was last year. Two scholarly young volcanologists have here compiled a modern version of that 1888 tome out of London. They offer us an exemplary treatment: documented, comprehensive, fascinating, beautiful, even cheap. The bargain volume consists of a useful introduction and concise chronology, then a hundred pages of eyewitness accounts, half in terror and half in wonder, a first translation into English (done by E. M. and A. F. Koster van Groos and J. A. Nelen) of much of the monograph prepared by the prescient leader of the official on-the-spot investigation sent in seven weeks afterward by the government of the Dutch East Indies, and an anthology of scientific papers about Krakatau.

The cited papers cover the geology, the prodigious waves radiated the world around through air and sea, the visual atmospheric phenomena seen for once in a blue moon, the plausible effect on climate and the slow, steady return of life to the wrecked and denuded island. The papers span a century of interpretation, and they are eloquent witness to the blend of truth and error in science at any one epoch. The young mining engineer Rogier D. M. Verbeek, who led the official investigation, had sought the services of a photographer. That specialist objected to a long field trip, and in his place was sent "the sergeant-major draughtsman P. B. Schreuders." Schreuders' paintings, done with the aid of the camera lucida, were reproduced for publication by the new process of chromolithography; they supply vivid evocations of the scenes. (The most important single view, that of the cliffed ruin of the half cone as seen from the sea, was reproduced in the November 1983 issue of this magazine; here there are many more views as well.)

Since the geology has recently been described in *Scientific American* and the strange atmospheric effects are covered in the second book reviewed here, it is the waves we shall cite to sample the science reported in the first book. Death's chief minister was the fearful tsunamis. On the sloping coasts in the area the sea rose as high as 40 meters and more, carrying one warship inland for miles and sweeping prosperous market towns clean of their green groves and paddy, leaving them melancholy featureless tidal flats. For weeks the inland people toiled to burn or bury the corpses cast up here and there in heaps. As far away as San Francisco and even in the English Channel the tide gauges showed Krakatau-timed trains of waves inches high. The dynamics, however, did not fit well: those little waves had come too fast from the Sunda Straits, and in some instances any sea path seemed implausibly circuitous.

A fine paper of 20 years ago (by Frank Press and David G. Harkrider, who were then at the California Institute of Technology) goes far toward a solution of the century-old riddle. That small distant effect was not the sea wave direct but a secondary sea disturbance driven by the shocked atmosphere. (The sound of "distant roars of heavy guns" was reliably reported from an island 3,000 miles away, after four hours of sound transit time.) There are modes of vibration of the entire atmosphere that nearly match the phase velocity of the long waves in the deepest seas. It is these air modes that jumped over land barriers to reexcite local sea waves one way or another, with unexpected amplitude and speed. And there is no anxiety about the fearful substance krakatite quite stilled: Simkin and Fiske estimate that a surface explosion "of about 100 to 150 megatons would produce pressure pulses equivalent to those observed from Krakatau."

The Meinels are an astronomical couple from Tucson, well known at many a mountain dome around the world. To their professional work, now centered on the exploitation of solar energy, they have for two decades added an open-eyed study of sunset phenomena. They report it all here delightfully "in the old fashion, when science was open to every reader," the fascinating facts told in personal stories and as tales from friends. Catastrophe apart, their book, aglow with dozens of color plates of fiery sunsets, stilled by the purple vault of desert nights, echoes and explains the puzzled witness of those Krakatau-painted sunsets a century ago.

Their subject is vista. Narrow valleys and lowering rainclouds are not the places to look; long, low horizons and clear desert air are needed. The scale is set by the layered spheres of the earth and its atmosphere; the quantum clarity of the air molecules allows the effects of dust and droplets to shine out. The book begins with the geometry of sunset and its physics, made plain in diagram and graph. The green flash is given a chapter of its own; the most agreeable novelty is a simple sketch and account of how a triangular shadow converges to the horizon at sunset and sunrise to an observer watching from a mountaintop. That triangle does not depend on the shape of the mountain but is the necessary consequence of the set of shadow forms converging by perspective to a single highest point: the shadow point of the observer. Just as we stand always at the top of our world—the highest surface point along our own vertical radius—so do we stand at the tip of our mountain's shadow pyramid.

The most original part of this delightful book is volcano watching, albeit at a big distance. The dusts and aerosols from the eruptions did not stop, of course, with old Krakatau but continue still. Here are sunsets of recent decades colored by Agung and El Chichón and El Fuego and the rest, with the explanation of what you see and why. Some of the interpretation calls for rather more gear than a keen eye and a clock, but a good start can be made with what every viewer sees. The brilliant crimsons of post-Agung sunsets disappeared in half

© 1984 SCIENTIFIC AMERICAN, INC

a year, as had those of Krakatau. No one was surprised; the dust had simply settled out. But a century ago, as they did in the late 1960's, the enhanced sunsets unexpectedly recurred in the fall and winter for some years. The reappearances bore the mark of high aerosol layers: a sharp upper edge to the glow, together with spectral details. More is going on than a simple settling of dust.

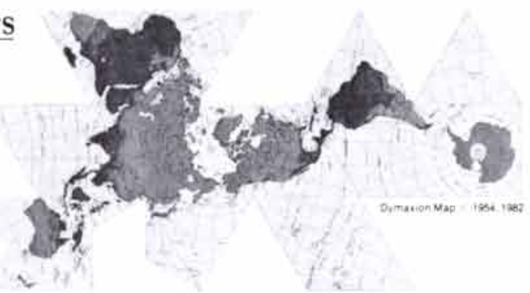
It seems probable that the stratospheric aerosol is in fact droplets of sulfuric acid (they have been found on the windshields of the high-flying U-2's) generated not by dust but by sulfur dioxide ejected by the volcano. That reagent slowly diffuses upward, reacting finally with the ozone layer to form the acid droplets. The abundance of the droplets is under the control of the seasonal response of ozone formation to the insolation of the high atmosphere. Acid droplets re-form season after season until most of the gas is used up. Blue suns and emerald green moons are not just centennial memories either: in northern China the effect is not infrequent, as dust borne by winds from the distant loess of the Gobi fills the skies to provide a complex example of angle- and color-dependent light scattering.

The night sky draws the Meinels' attention as well. In the dark are subtler glows: the aurora, high clouds of artificial and natural origin, effects of interplanetary dust in the plane of the ecliptic, odd visitors (such as asteroids and comets) and city lights. All of these are briefly introduced for the benefit of would-be watchers of the skies. Other worlds are not neglected; for twilight glory, though, the earth remains the "best of all the planets," at least while it enjoys a climate that allows clear air. It seems probable that cloudy skies were general during epochs as warm and rainy as the Mesozoic; the swampy earth of the coal beds was no planet for sky watchers. The Meinels aptly chose the psalmist to close in eloquence their engaging and enthusiastic book: "The dawn and the sunset shout for joy!"

THE VISUAL DISPLAY OF QUANTITATIVE INFORMATION, by Edward R. Tufte. Graphics Press, Box 430, Cheshire, Conn. 06410 ($34). The old page reproduced shows an outline map of a part of eastern Europe, with a few towns and river courses marked. Across the map there is plotted an unfamiliar current, certainly no river we know, wide in the west, dwindling as it makes its way eastward to Moscow. A smaller return current, graphed in black, flows back to reach the west again, but now wasted to a trickle. What is plotted is the tragic ruin of the Grande Armée, the triumph of General Winter and his hardy Russian allies over Napoleon I in the year 1812. The width of the band is proportional to the number who marched be-

BUCKMINSTER FULLER'S

MAP FOR SPACESHIP EARTH



A ONE-WORLD ISLAND IN A ONE-WORLD OCEAN

Forty-one years ago, Buckminster Fuller invented this cartographic projection system as an alternative to the Mercator world map which shows Greenland three times larger than Australia (actually, the opposite is true) and ignores Antartica.

Fuller's Dymaxion Sky-Ocean Map is the only flat map of the world with no visible distortion in relative shapes and sizes of its geographic features, and with no breaks in any of the continental contours. The Dymaxion Map received the first U.S. patent ever granted on a method of cartographic projection.

You may order a full color print of the Dymaxion Map, suitable for framing (34" x 22") $10 postpaid or five maps for $35 postpaid. The map comes with an illustrated chart describing the map projection philosophy and mechanics. Profits from map sales will support the research and development work of the Buckminster Fuller Institute. A newsletter from the Institute is also available.

Mail to: Buckminster Fuller Institute
1743 South La Cienega
Los Angeles, CA 90035
or call: (213) 837-7710

Please send me:
☐ ___ Dymaxion Maps at $10 each postpaid (5 Maps for $35 postpaid)
☐ One year's subscription (8 issues) to the Institute's newsletter for $8
☐ A free catalog of other educational materials

☐ Enclosed is a check or money order
☐ Please charge to my Visa or MasterCard
Make checks payable to Buckminster Fuller Institute.

Name ______
Address ______
City ______
State ______ Zip ______
Card number ______
Expiration date ______

© 1984 SCIENTIFIC AMERICAN, INC

hind their emperor at each place. The invaders crossed into Russia 422,000 strong; only 10,000 were left to cross the Niemen a second time on the way back. Below the map lies a graph of winter's tactics: the drop of the temperature during the retreat, the curve carefully aligned in place and time. In all its "brutal eloquence" this multidimensional classic "may well be the best statistical graphic ever drawn." It is an 1861 product of a gifted French engineer, Charles Joseph Minard.

Original, beautifully presented, sharp and learned, this book is a work of art history and reasoned criticism. The art here is a cognitive art, the graphic display of relations and empirical data, now an indispensable tool of science and engineering. Maps were made long before graphs; the representation of spatial relations may be older than writing itself. But young Edmund Halley's world map of winds plotted as "little stroaks," published in a 1686 *Philosophical Transactions of the Royal Society,* is about as early an example of thematic or data mapping as we know.

That degree of abstraction was difficult. There are rare examples of time-series plots even before then; one is shown here (it plots the inclinations of planetary orbits) from a manuscript of perhaps the 11th century. Graphs of time series become familiar late in the 18th century; the versatile physicist and mathematician J. H. Lambert was a pioneer in their thoughtful use by 1770, and a decade later the political economist William Playfair presented an entire atlas of economic data such as wages and trade balances convincingly graphed over time. He also invented the familiar bar and pie charts.

Among all data graphics it is only maps, time-series plots and bar and pie charts (sometimes jazzed up with more representational symbols) that to this day reach the wide public. Professor Tufte has made some sample surveys. Consider relational graphics, any plot that is neither a map nor a time series yet links two or more variables. They are the everyday stuff of scientific publication, amounting to 40 percent of the graphics in a typical scientific weekly. But of 15 widely known news publications around the world, about half (including *Time, Pravda* and *The Wall Street Journal*) carried no such plots at all. *The New York Times* presented one plot for each 200 graphical items. It is probably no surprise that *Asahi, Der Spiegel* and *The Economist* did better; the data-loving Japanese even observe a national statistics holiday, complete with graph-drawing contests.

Graphs can lie handsomely, of course, but so can words. As Tufte sees it, undue concern with the detection of deception and the related idea that graphics were meant mainly for the naive ("showing the obvious to the ignorant") stultified the art from 1930 to about 1970. In the second half of the book he takes the stance of philosopher-critic, more normative now than descriptive. The basis of his theory commands assent: above all else show the data. On this rock he builds firmly, with examples in rich variety, from electroencephalograms with eight aligned tracks from which nothing can be erased to preposterous examples of false perspective and multicolor fields presenting little information. He dubs the latter chartjunk, along with various forms of undisciplined crosshatch dazzle, emphatic but information-free grids and the crude busy labels so easily entered on the keyboard. All these elements are now close to high fashion in the world of computer graphics, infesting advertisements, corporation reports and television.

The book ends with what a keen critic often neglects: real help. Statistician and economist, Tufte has become seized with the visual art. His proposals for putting to useful statistical work the redundant axes of a scatter plot are practical. (What use is the origin, anyway? The ends of the framing axes can mark the data range, and an offset in the line can give quartiles.) Clever little ideas such as the white grid for scaling bar charts and histograms should become popular; they fit Tufte's theory. Multiple small graphs, evoking the frames of a motion picture, are a successful recent scheme for data-rich designs. Typically a fixed geographic background, say a thumb-size map of Los Angeles County, is presented repeatedly on the page, each time with shifting data on the map plotted as contours or in perspective peaks and valleys.

It is not the density of information that bores and confuses; more often it is the density of noninformation. One vapid bar chart displays about a sixth of one data number per square inch; better news comes with the *Times*'s annual New York weather-summary plot, with nearly 200 numbers per square inch, and a 1973 outline map of the communes of France, nearly 9,000 numbers per square inch. The current record is held by the well-known map of the galaxy density on the sky; its two million encoded gray rectangles yield a data density of about 110,000 numbers per square inch.

A final chapter adds technical hints, on line weights and graph shapes, on the right combination of words, numbers and pictures. An epigraph from Henri Matisse is apt: "I want to reach that state of condensation of sensations which constitutes a picture." The designer of graphics has the task of giving easy visual access to the subtle and the difficult. This volume brings together a gallery of models, a workable theory of approaches and a cautionary display of failures. One hopes it will find not only general readers who will take delight in it but also heavy service among those who elaborate the long codes that rule a million screens.

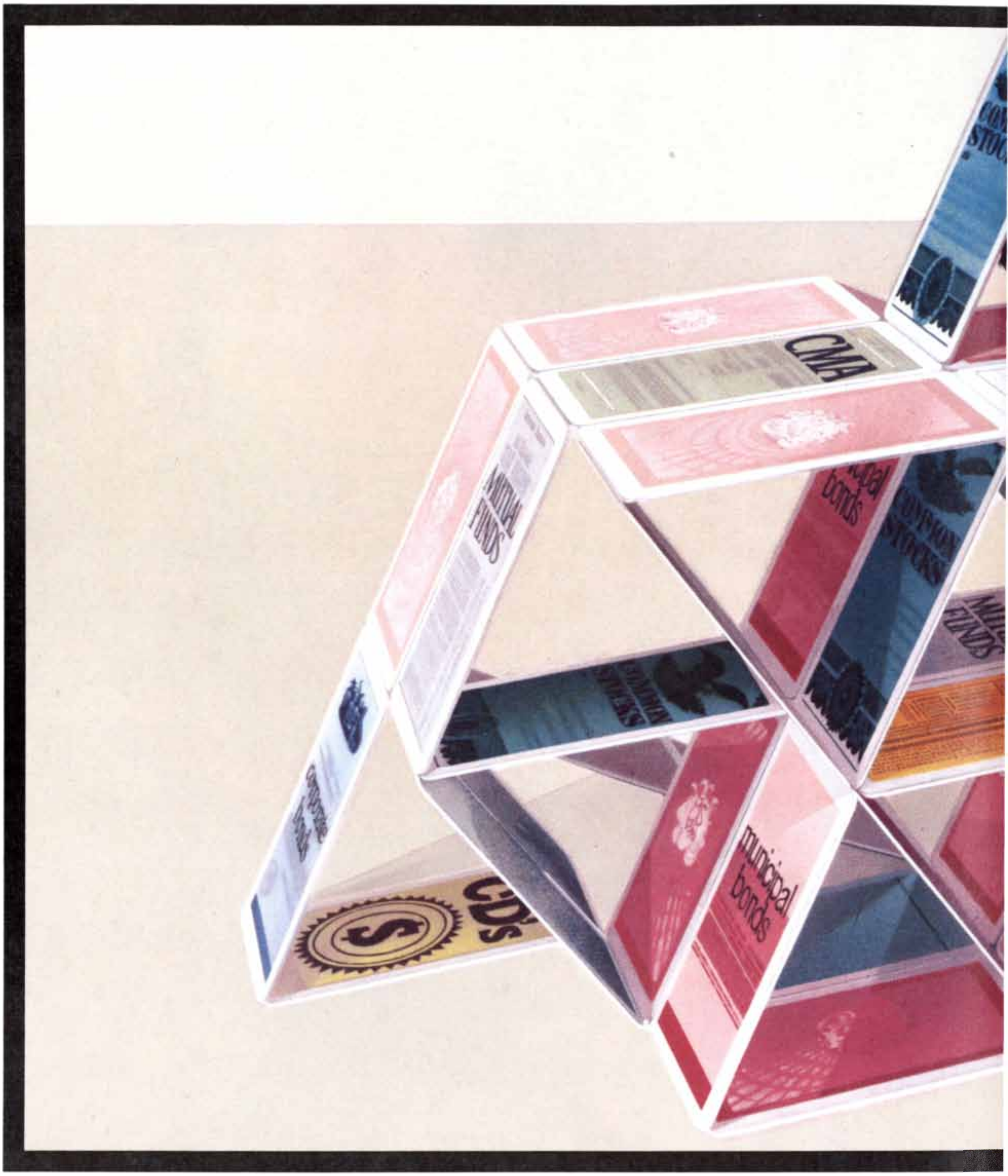
NUCLEAR WEAPONS DATABOOK, VOLUME 1: U.S. NUCLEAR FORCES AND CAPABILITIES, by Thomas B. Cochran, William M. Arkin and Milton M. Hoenig. Ballinger Publishing Company ($38). Enough survivors at Hiroshima saw the slowly falling instrument-carrying parachutes released by a telemetry plane that accompanied the strike bomber to support the widespread belief the nuclear bomb itself had come down by parachute. That unlikely and unused tactic of 1945 is now a matter of design. All development complete, engineered for first production this year (the total production run is expected to be 2,500), enters the Modern Strategic Bomb, Type B83. Neat external drawings of five other nuclear bombs are here with their dimensions, although not yet the B83. A photograph of the prototype dropped on its parachute from an F-111 aircraft is given. Some of these new bombs will be placed on the B-52's and FB-111's of the U.S. Strategic Air Command. B83 is intended as the major gravity weapon for the forthcoming B-1B bomber, the first squadron of which will be operational out of Dyess Air Force Base in Texas late in 1985. That supersonic plane, whose production will cost the taxpayers some $30 billion for the 100 aircraft on order, can carry 24 of the 12-foot bombs internally and perhaps another 14 as external load (at some cost in speed from their inevitable drag). Each B83 weighs some 2,400 pounds. The parachute, no novelty in itself, is newly designed to retard the fall of the bomb from the supersonic aircraft roaring evasively past its target at altitudes as low as 150 feet and allows a laydown accuracy of some 600 feet. The surface explosion is delayed long enough by parachute and fuzing to let the bomber escape.

The yield of the weapon is a nominal megaton-plus, probably from a mixed fission primary of plutonium and uranium, plus the thermonuclear secondary. It is therefore suited to the hard targets, such as missile silos and command bunkers, for which it has been specifically designed. Not many witnesses are likely to be close enough to see the parachuting of B83 and survive to report it.

B83 is a clever bargain design out of the Lawrence Livermore Laboratory. It offers the possibility of high-speed low-altitude delivery, together with a number of improved safety features, at a cost reduced from the excessive price tag of a canceled design of the late 1970's. It will replace three or four earlier bombs, one of which, also used as the warhead of the Titan missile, has an explosive yield of nine megatons. That

© 1984 SCIENTIFIC AMERICAN, INC

© 1984 SCIENTIFIC AMERICAN, INC



© Copyright 1984 Merrill Lynch, Pierce, Fenner & Smith Incorporated. Published in U.S.A. All rights reserved. Member, SIPC.

It's how we make each piece support all the others that makes us what we are:

Merrill Lynch. A breed apart.

© 1984 SCIENTIFIC AMERICAN, INC

Arranging your finances can be a fragile affair. That's why you need to build a lasting relationship with a firm that offers a full range of integrated financial services. A firm that has the commitment to deliver each one with the same uncompromising quality.

For an introduction to the many ways in which we make each part of your finances work in concert with the others to create a single, solid structure, please look at our new brochure, *A Commitment to Excellence.* For your copy, just call 1-800-MERRILL (ext. 9080) Monday through Friday from 8:30 a.m. to midnight Eastern time.

Merrill Lynch
Merrill Lynch, Pierce, Fenner & Smith Inc.
A breed apart.

© 1984 SCIENTIFIC AMERICAN, INC

model is the highest-yield weapon ever in our arsenal and is now out of date. Four tons in weight, 20 years old, it is shown in a photograph from the National Atomic Museum; its New Mexico origin is easily read from the ornamental clasp worn instead of a necktie by one of the engineers standing beside it.

No sharp issue arises around B83; it simply represents the technical change normal for nuclear weapons in the U.S. forces for nearly 30 years, since the pioneering early 1950's. Those details of specifications, purpose, evolution, production, deployment and the like, however, stand as a sample of this unusual reference work, factual and comprehensive, although only modestly technical. It has been compiled over the past few years out of the open literature, now voluminous, by two physicists and one military analyst working in Washington.

Their numerous sources are documented by meticulous footnotes; chief among them are the declassified hearings of the main congressional committees that act on the appropriations for the armed services and the Department of Energy. The first three chapters are the most interesting for the general reader. They present an overview, a primer on the nature and functional features of nuclear weapons (which wisely does not attempt any conjecture about their structure except in general terms) and a remarkable warhead-by-warhead description of the current U.S. stockpile, 24 types of nuclear warhead in 40 pages of data and photographs. These proceed from W25, the warhead of the Genie air-to-air interceptor rocket, with a yield of one or two kilotons, first flown in 1957, up to W80, the common warhead to be carried by the air- and sea-launched cruise missiles whose deployment is only now under way. W80 manages its yield of 200 kilotons in a weight of about 270 pounds. Of Los Alamos origin, with a 20-year evolution, it is probably a deuterium-tritium-boosted fission device, as compact as they come.

Warheads projected for the future are also well covered, albeit with more gaps in the data. Some 15 planned new weapons can be identified over the eight or 10 years ahead. There are many replacements of older warheads (a 30-year life cycle is about average) and a net increase of about 4,000 weapons by 1990, for a total of 30,000.

The rest of the book centers on the delivery systems. The material on new systems—such as Peacekeeper (MX), Trident, Pershing, the new cruise missiles and the new reentry vehicles—is timely and important for the discussions of the day. The comprehensive approach brings with it quite a few pages on aircraft, ships, missiles and artillery capable of delivering either conventional or nuclear weapons.

The most remarkable photograph here is one of a wildly painted tanker aircraft refueling two Tornado fighter-bombers at once. The strangest device is surely W54, the special atomic demolition munition, 60 pounds in its backpack for a low subkiloton yield, to be emplaced by an engineer at some bridge or tunnel not too far. The most disturbing Department of Defense projection is that of the Trident II CDD5 missile: "hard target kill across full-spectrum, higher-yield warheads," a program under acceleration for 1990. It promises some 5,000 accurate high-yield warheads, with terminal-guidance option and instant retargeting. Such powerful new offensive strength at sea is said to be in part a hedge against the eventual cancellation of the MX. (CDD5's targets would seem to include the congressional end of Pennsylvania Avenue.)

In 1980 we had 722 military units officially certified by the Defense Department for some direct role in nuclear warfare. (In the Air Force, for example, such a unit would usually be a squadron.) Those forces number more than 100,000 specifically trained and cleared persons in all the services, their weapons "properly wired and inspected." That is the force assessed overall by this careful book. Its deployment maps, stockpile inventories and detailed lists of features, dates and models make visible the forms that hide in the shadow of potential destruction. Here is the real world, if a cockeyed one.

The ambitious authors propose seven companion volumes over the next few years. The second volume will cover the production complex for nuclear weapons in the U.S., the third (now in preparation) Russian weapons, the fourth other national nuclear forces, the fifth health and safety (sic), the sixth command and strategy, the seventh arms control. The eighth will close the set less technically with a review of arms control and a history of nuclear weapons.

RECOMBINANT DNA: A SHORT COURSE, by James D. Watson, John Tooze and David T. Kurtz. W. H. Freeman and Company ($27.95). "There is no substance as important as DNA." An astronomer might contest the point, but let it stand. In a swift, crisp, often aphoristic way these three molecular biologists and their talented illustrator George Kelvin have set out the state of their marvelous art in mid-1983: how the living cell analyzes, edits and publishes the genetic message. The level is that of *Scientific American* at its most austere, perhaps even more demanding because the topic is sustained over 18 chapters. Attention is centered on the laboratory and the experiments and concepts developed there (although a final chapter enters the spectacular new industry, and the hopes of therapy are not forgotten).

A sense of history and impact is conveyed throughout. There is a time line of main events and a few excellent photographs of major actors in the drama. The material recalls a more extensive historical publication, *The DNA Story,* published by Watson and Tooze in 1981. Their new "course" has grown out of the scientific summary in that book. Indeed, the organization remains broadly chronological: the essential foundation of the discipline was laid down over a period of 25 years, from the recognition of the genetic transforming power of DNA through the discovery of its double-helical form and replication chemistry to the establishment of the full genetic code in the late 1960's. The background account is thus historical in feel. The subject has grown apace since the 1970 discovery of the first restriction enzyme—the editor's scissors, so to say. As in the rise of the large-scale silicon chip, the tale is mainly one of a decade.

The first four chapters set out the background, up to the studies of the control of gene expression in cells higher on the evolutionary ladder than bacteria. Almost everything depends on replication. For any molecule in the living cell a sample large enough to characterize or to reuse can be obtained once the molecule is synthesized by the cell in obedience to some part of the DNA instruction tape. Ten chapters, filled with ingenuity and insight based on experiment after experiment, treat the central issue: getting cells to make what DNA you will. The problem is soluble because the evolution of life, particularly at the bacterial level, has been marked by a struggle between intruding DNA capable of exploiting a host and a host cell determined to resist the intrusion.

Thus did bacteria evolve enzymes able to locate foreign DNA by noting unusual words in the text, and there to fragment it. Those enzymes are the scissors, the restriction enzymes reproducibly cleaving the DNA text at given phrases. Other bacteria found it useful to maintain a set of circular minichromosomes, DNA loops of no great length but in large numbers of copies. These could multiply rapidly, to allow when it was necessary the quick manufacture of enzymes to attack various antibiotic molecules. The clever human editors simply put their own little texts into the loops by finding sticky ends that could enter the specifically cleaved loops frequently. The cells copied the new recombinant texts just as well as the old. The art had begun.

There are other natural tactics for introducing DNA texts where it can prosper even though it is unwanted. There are bacterial viruses, miniloops in higher cells such as yeast, tumors in plants and even in animals—all with mechanisms leading to the acceptance of foreign text by the cell's printshop. The newly printed foreign texts are the DNA

© 1984 SCIENTIFIC AMERICAN, INC

fragments cloned in bacterial cultures. (It is unnecessary to continue any longer with the text-composition analogy, which in fact is not found at all in this concise book.) The largest fragments that can now be cloned reliably run to a length of some 50,000 pairs of nucleotide bases. A single human chromosome is a couple of hundred times longer.

By optical-mechanical sorting out of a cell culture it was possible to obtain a sample of human *X* chromosomes. That sample was fragmented with restriction enzymes and introduced into the bacterial host by means of a virus. The result is a large number of distinct virus colonies. In each colony we can expect a yield of many copies of one random fragment, 1 percent or less of the human chromosome. Most of the human DNA input is to be found in one or another distinct virus patch. Any means of analysis of that DNA—mapping its genes by hybridization, even identifying its sequences, meaningful or merely repetitive—can now be carried out.

This process, far from complete, is cited here as only one example typical of DNA work in the 1980's. The DNA fragments might be studied at some points, at least, down to the level of single base pairs, with the right restriction enzymes. Classical genetics was able to recognize under the microscope the bands in single stained chromosomes or to map chromosomes cytogenetically, resolving a length of a million or so base pairs. Write the authors: "We can confidently look forward to medium-resolution maps of entire human chromosomes, and high-resolution maps of interesting and medically important segments."

This is only a coarse account of a fine text whose power is in its explicit detail. It is as successful a summary of a discipline in flower as one can find, a worthy follower to the senior author's earlier comprehensive text on the gene. There is an air of good fortune about this subject. The dangers have dwindled and success has multiplied like the bacteria. It was not expected. Several shrewd investigators "left molecular genetics to start up new careers in neurobiology" just at the dawn of recombinant success. Perhaps the time of neural and developmental understanding will yet come, but this marvelously dexterous science is pretty surely its prerequisite.

APPLES: A GUIDE TO THE IDENTIFICATION OF INTERNATIONAL VARIETIES, by John Bultitude. University of Washington Press ($50). In 1905 the Agricultural Experiment Station of the Empire State published a two-volume celebration of its orchards: *The Apples of New York,* by S. A. Beach. It was a survey of apples in all their variety, with many photographs in color and in black and white. This book, a cosmopolitan modern version of Beach's proud old work, stems from England. Bultitude is a world authority on apple varieties with a lifetime of experience in the propagation of apples. His field guide to apples rests on the remarkable collection of 2,000 varieties at the National Fruit Trials in Brogdale, where over some 10 years a colleague has made beautiful photographs in color of 250 of the best-known varieties, all as fruit growing on the bough, a study in the riches of red, green and gold. For each variety Bultitude has prepared a page of description and brief history, with photographs of two cut sections, one polar and one equatorial, to help in identification.

The marketplace centers on very few varieties, through the shifting choices of the grower and the consumer alike, seeking size, yield, keeping quality, flavor and eye appeal. In 1905 the most popular American variety was the Baldwin. Now Delicious and Golden Delicious make up more than half of all apples grown in the U.S.; in the U.K. the weather is not warm enough for the Delicious to reach a popular size. There Cox's Orange Pippin holds more than half of the total market. On the Continent those three leading U.S. and U.K. varieties together add up to a market majority. In Australia the green Granny Smith supplies 56 percent; in New Zealand the Delicious and the Granny Smith share more than half of the market. Variety remains, but somewhat off in the corners.

The breeding of apples by crossing known varieties was vigorous in the 19th century. Most commercially successful varieties, however, are sports: mutants that arose by chance in some orchard. The Delicious was found in 1880 as a shoot growing from a rootstock in Iowa; the Granny Smith was "raised from chance seed thrown out by Mrs. Thomas Smith, Ryde, New South Wales" before 1868. Mutants are under steady selection from among all these major varieties, often chosen for color. Desired stocks are propagated by grafts or are cloned: reared vegetatively from a single parent plant, since apples do not usually breed true from seed. Genetics is beginning to make its mark, although chance still rules.

The scientific reader will not want to overlook one "distinctly ugly cooking apple, [a] very shy cropper." The dully reddened green fruit traces back to a tree growing in Isaac Newton's mother's garden at Woolsthorpe, and it is held to have been repropagated these three centuries and more. Cortlands, McIntoshes, Mutsus (called Crispins in the U.K.), Northern Spys, Red Astrachans, Jonathans, Gravensteins—they are all here, and a couple of hundred more. What the supermarket does not provide may well be cherished in some local garden or old orchard.

NordicTrack

Jarless Total Body Cardiovascular Exerciser
Duplicates X-C Skiing For The
Best Way To Fitness

NordicTrack duplicates the smooth rhythmic total body motion of XC Skiing. Recognized by health authorities as the most effective fitness building exercise available. Uniformly exercises more muscles than jogging, swimming, cycling and rowing.

Does Not cause joint or back problems as in jogging. Highly effective for weight control and muscle toning.

Easily Adjustable for arm resistance, leg resistance and body height. Smooth, quiet action. Folds compactly to require only 15 by 17 inches of storage area. Lifetime quality.

Used In thousands of homes and many major health clubs, universities, and corporate fitness centers.

Call or Write for FREE BROCHURE
Toll Free 1-800-328-5888 MN 612-448-6987
PSI 124 F Columbia Crt., Chaska, Minn. 55318

Play the finest
HARPSICHORDS & FORTEPIANOS

Build it yourself from one of our kits or let our experienced craftsmen build it for you.

Three Centuries of Harpsichord Making
by Frank Hubbard (Harvard Univ. Press.
397 pgs. $25.00 postpaid)

For brochure send $1.00 to:
HUBBARD HARPSICHORDS
INCORPORATED
144-S Moody St., Waltham, MA 02154
(617) 894-3238

© 1984 SCIENTIFIC AMERICAN, INC

Epson.
For those who need it, simplicity.

One computer.
Two points of view.
The Epson QX-10 personal computer.

To many, the Epson represents the ultimate in simplicity.

Just press a single key for the function you require: word processing, scheduling, business graphics, address book or file management. One keystroke produces your program. There are no rigamaroles to remember. No disks to change.

The result: you start to work immediately. And you start being productive, immediately. With step-by-step prompts. In plain English, not computerese.

Simplicity itself.
Or is it?

The plain fact is that the ease of operation the Epson offers today is accomplished with a degree of technological sophistication most other computers can only promise for tomorrow—specifically, fully integrated software, operating in an interactive environment.

The few other computers offering such "simplicity" cost $5,000 to $15,000 more. And most other computers can't offer it at any price. Which makes one wonder exactly what they do offer, in terms of either simplicity, or performance.

HOW MUCH CAN YOU DO ON THE EPSON? HOW MUCH ARE YOU READY TO DO?

The Epson's ease of operation may spoil you, but it certainly won't limit you.

Case in point: every Epson comes complete with an integrated software system—*Valdocs*®—to effortlessly provide the basic functions for which most people buy computers. The Epson also comes with *CP/M*®-80 2.2, so you can choose from the hundreds of programs in the CP/M library. And only Epson offers an exciting new collection of seven best-selling programs now specially enhanced to give you every powerful feature, plus Epson one-button simplicity. Included are *dBase II*,® *Friday!*,™ *Microplan*,® *Graphplan*,™ *WordStar*,® *SpellStar*,® and *MailMerge*.® And the Epson also allows you to add *MS*™*-DOS* compatibility, so you have access to best-sellers like *Lotus*® *1-2-3*.®

Best of all, you will run the software of your choice on the computer of choice. The high-performance Epson. With 256K RAM. 128K dedicated video memory. The breathtakingly sensible *HASCI*™ keyboard. Dual 380K double density disk drives.

Graphics capabilities unequalled in its price range. A high resolution monitor, 640 by 400 pixels, for clarity few computers in any price range can offer. Plus, an RS-232C interface, a parallel printer interface, and internal space for up to five peripheral cards so you can expand your Epson as your needs require.

One further point: all these features, and quite a few more, are included in the Epson's $2,995 price. Some com-

© 1984 SCIENTIFIC AMERICAN, INC

Epson. For those who can handle it, performance.

puter companies ask you to pay extra for features like these. Most can not offer them at any price.

That, too, is performance. The kind of performance that can make choosing a personal computer very simple, indeed.

EPSON QUALITY. OR, WHY WONDER WHAT TERRIBLE KLUDGES LURK IN WHICH SLEEK BOXES.

If you know computers, you know Epson. Epson printers set the industry standard for quality, reliability and value. Rest assured, the same can be said for the Epson personal computer.

The satisfying silence of the slim, Epson-designed disk-drives is one way for you to judge or, for an inside-out perspective, here is an excerpt from a review by Jim Hanson in the April, 1983 issue of *Microcomputing*.*

"The Epson QX-10 is soundly designed and executed. I looked hard and found no evidence of kludging or shorting out anything in the name of economy. All the connectors have gold on them and are of quality manufacture. The printed circuit boards are heavy, with soldermarks on both sides of double-sided boards. The circuit boards are completely silk-screened with component labels, and the layout is as professional and clean as you will find anywhere."

Isn't this what you expect? After all, it's an Epson.

A WORD TO THE WISE: GET YOUR HANDS ON THE EPSON.

Is the Epson a simple, easy-to-use computer for beginners? Or a sophisticated high-performance computer for the experienced? The answer is "yes." And when you think about it, aren't those two computers the one you need now.

For technical specifications, and the complete, 3-part *Microcomputing* review, along with the name of your nearby Epson dealer, call toll-free (800) 421-5426. California residents, call (213) 539-9140.

*Excerpt reprinted by permission of Microcomputing Magazine. All rights reserved.
CP/M, dBase II, Microplan, WordStar, SpellStar, MailMerge, Lotus, 1-2-3 are registered trademarks of Digital Research, Ashton-Tate, Chang Labs, MicroPro (3), and Lotus Development (2). Valdocs, Graphplan, MS, and HASCI are trademarks of Rising Star, Chang Labs, Microsoft, and Rising Star, respectively.

EPSON

STATE-OF-THE-ART...SIMPLICITY.

© 1984 SCIENTIFIC AMERICAN, INC

Ford Escort Diesel: Better mileage than this leading import.

We didn't believe it at first, either.

But EPA testing figures established it. Our new Escort Diesel is rated approximately four miles per gallon higher than a Honda 750.*

Just take a look at our numbers:

46 EPA EST. MPG.

68 EST. HWY.

And because this diesel is an Escort, there's a lot more to talk about than great economy.

Like the fact that Escort's the best-selling car in the world.**

Or that it comes with more total passenger room and more total cargo room than a Honda Accord.†

More standard features than a Toyota Tercel.††

There's even a fully independent suspension system for a smoother ride than a Nissan Sentra.

All of which means Ford Escort not only gives you a big advantage over that motorcycle pictured above.

It also beats more than its share of cars.

THE BEST-BUILT AMERICAN CARS.

When we say "Quality is Job 1," we are talking about more than a commitment. We are talking about results. A recent survey concluded Ford makes the best-built American cars. The survey measured owner-reported problems during the first three months of ownership of 1983 cars designed and built in the U.S.

And that commitment continues in 1984.

* For comparison. Honda 750 mileage is obtained from EPA emissions testing and is *not* an official rating. FS Diesel mileage applicable to sedans without power steering and A/C. Available for order. Your mileage may vary depending on speed, trip length, weather. Actual highway mileage lower. All Escort Diesel models except the FS available in California.

** Sales estimates based on worldwide production figures.

† Based on EPA Interior Volume Index.

††Escort GL (shown) compared to Toyota Tercel 3-door deluxe liftback.

Get it together — Buckle up.

Have You Driven A Ford... Lately?

Ford

© 1984 SCIENTIFIC AMERICAN, INC

Established 1845 SCIENTIFIC AMERICAN March 1984 Volume 250 Number 3

"No First Use" of Nuclear Weapons

The security of all nations would be enhanced if the U.S. and its allies were to adopt a military strategy that did not rely on nuclear arms to counter a non-nuclear attack

by Kurt Gottfried, Henry W. Kendall and John M. Lee

Since the late 1940's the U.S. and its allies in the North Atlantic Treaty Organization have based their common defense on the declared strategy of initiating nuclear warfare if their conventional, or non-nuclear, military forces should be threatened with defeat in Europe or elsewhere. The doctrine of "first use" of nuclear weapons was formulated at a time when NATO was faced with two options for countering the large standing army maintained by the U.S.S.R.: either conscripting and outfitting large numbers of men (an obviously unpopular move after six years of combat) or relying on a comparatively cheap and enormously destructive weapon their adversary did not possess. Little sleep was lost in making the choice or in reconsidering it as long as the U.S. held a significant lead in nuclear arms over the U.S.S.R.

The era of unquestioned U.S. nuclear superiority has long since passed, and with its passing confidence in NATO's first-use policy has eroded. Doubts have arisen partly because it has proved impossible to devise plans that promise to gain a military advantage from any limited use of nuclear weapons. In addition no method has been proposed for stopping a nuclear war once it has started. Accordingly it has become clear to many people that even the most limited use of nuclear weapons could well lead to the ultimate catastrophe of a global nuclear war. Nevertheless, under the present first-use policy nuclear weapons are integral to the training, planning and equipping phases of all NATO military operations. In the event of reverses in a major conflict there would be almost irresistible pressure to use them.

These concerns have led some prominent observers to call for a new NATO policy on the initiation of nuclear warfare. In particular, in the spring of 1982 four former high officials of the U.S. Government (McGeorge Bundy, George F. Kennan, Robert S. McNamara and Gerard C. Smith) published an article in *Foreign Affairs* advocating that serious consideration be given by the U.S. and its NATO allies to a policy of "no first use" of nuclear weapons. At the same time the Union of Concerned Scientists publicly explored a set of measures to enhance international security and reduce the risk of nuclear war. One of the recommendations that emerged from this review was that NATO should move toward a no-first-use policy in Europe and that the U.S., independently of its European allies, should do the same everywhere else in the world.

The scientists' group, of which we are members, then sponsored a more detailed study of the no-first-use option, with particular reference to the concomitant need to strengthen NATO's conventional forces. The study was directed by one of us (Lee) and involved (in addition to the other two of us) Gerald Steinberg and Peter Trubowitz. A number of retired senior military officers and former civilian defense officials on both sides of the Atlantic took an active part in the project and endorsed the final report. This article is based largely on the results of that joint effort.

Any analysis of the no-first-use proposal must deal primarily with the military situation on the central European front, where the countries of NATO and the Warsaw Pact confront each other with the two most powerful military forces ever assembled in peacetime. It is generally agreed that NATO would resort to nuclear weapons only if an offensive by the Warsaw Pact appeared to be on the verge of success. Accordingly military casualties would already number in the thousands, and the rapidly shifting front would run somewhere through densely populated West Germany. In the circumstances the use of tactical, or short-range, nuclear weapons could lend effective support to troops on the battlefield only if the commanders were able to make rapid decisions on the basis of accurate intelligence and if the weapons could be promptly released for use. Even if the initial NATO nuclear operations were for purposes other than direct battlefield support (for targeting Warsaw Pact command-and-control centers or support forces far in the rear, say), there would be the same need for coordinating them with a rapidly changing situation on the basis of the best available information.

Formidable obstacles would stand in the way of such NATO operations. The intricate system NATO relies on for command, control, communications and intelligence (C^3I), which is shared by its conventional and nuclear forces, is highly vulnerable to attack and would presumably be a prime target for the Warsaw Pact forces from the very beginning of hostilities. At present the "nuclear threshold" cannot be crossed by NATO field commanders acting alone. The highest political leaders of the NATO countries are supposed to agree on the timing, magnitude and location of any nuclear attack, and for this reason NATO has set up an elaborate procedure for reaching such decisions.

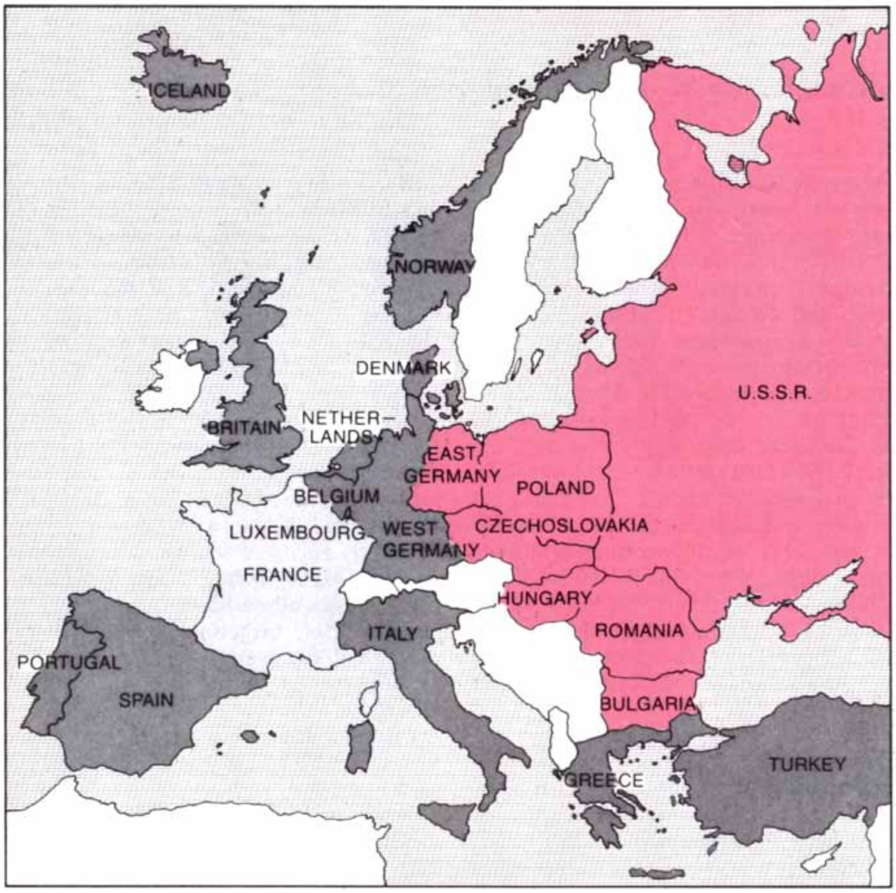
Thus NATO's military and political requirements are in direct conflict with each other. From the military stand-

© 1984 SCIENTIFIC AMERICAN, INC

point it is essential to take prompt action while the C^3I system is still able to provide reliable data and before the NATO armies, which have been trained to depend on nuclear weapons, begin to disintegrate. From the political standpoint the decision would require a consensus capable of withstanding the unprecedented strains inherent in initiating and sustaining a nuclear war. Such a consensus could be bought only with time, whereas the military situation could deteriorate to the point of collapse if the political leadership delayed—or refused to allow—the use of nuclear weapons.

War games and other military exercises by the American and British defense establishments throw further doubt on the military utility of the present first-use policy. They indicate that if NATO were to introduce tactical nuclear weapons into the conflict, the nuclear exchange that would almost certainly ensue would inflict more damage on NATO than on the Warsaw Pact. The retreat of the conventional NATO forces would create an ideal target for nuclear attack: a concentrated army immersed in its own population. Hence the Warsaw Pact forces could afford to be much less discriminating in their use of nuclear weapons than the NATO forces would have to be. The highly centralized political structure of the Warsaw Pact, dominated by the U.S.S.R., would enable it to quickly exploit opportunities in any tactical nuclear engagement.

The destruction caused on both sides by any such nuclear exchange could of course be immense. Of the roughly 9,000 tactical nuclear weapons deployed in the European theater of operations thousands would presumably be detonated in East Germany and West Germany, directly exposing many urban areas to blast and fire. The resulting radioactive fallout could contaminate tens of thousands of square miles on both sides of the central front [*see illustration on page 37*].



MEMBERS OF TWO MILITARY ALLIANCES opposed to each other in Europe are identified on this map. The nations belonging to the North Atlantic Treaty Organization are shown in gray. The Warsaw Pact nations are in color. France is a member of NATO but does not participate in its military-command structure. Transatlantic NATO members are U.S. and Canada.

	NATO	WARSAW PACT
POPULATION	575 MILLION	375 MILLION
ARMED FORCES	4.90 MILLION	4.76 MILLION
GROSS NATIONAL PRODUCT	$5.77 TRILLION	$2.02 TRILLION
ANNUAL MILITARY EXPENDITURES	$260 BILLION	$127 BILLION

AGGREGATE STATISTICS for the nations of NATO and the Warsaw Pact are compared. The NATO totals include figures for France but not for Spain, which joined NATO in 1982.

For a nuclear war to remain limited the two sides would have to negotiate quite specific, verifiable limits on weapons and targets. This unprecedented feat of diplomacy would have to be achieved in a state of crisis and uncertainty unknown to history, with each side holding in reserve a nuclear arsenal that could carry the conflict to virtually any level of destruction. It would be reckless to assume that escalation far beyond the battlefield could be averted in such circumstances. It is hardly surprising that many senior military officers and civilian defense officials on both sides of the Atlantic have long held that nuclear explosives have no utility whatever as battlefield weapons.

The recognition that the decision to use nuclear weapons is too grave to be tied to the tactical imperatives of the moment has led some strategists to espouse another nuclear option for averting a conventional defeat. In this case first use would take the form of a nuclear attack on one or more military targets far from the front that have a value sufficiently high to "demonstrate will" but not so high as to provoke a much greater counterresponse. Whether or not such targets exist is open to question. In any event whatever effectiveness this strategy would have would depend on the threat of escalation inherent in it, and so it offers no escape from the dilemma presented by the strategy of first use of nuclear weapons on the battlefield.

In short, there is no plausible scenario for the use of nuclear weapons in a conflict between the two superpowers that does not carry with it the danger of catastrophic escalation. To some this conclusion merely emphasizes the value of the first-use doctrine as a deterrent. Others have come to believe NATO cannot stake the future on a strategy that is "either a bluff or a suicide pact," to quote Field Marshal Lord Carver, former chief of the British Defence Staff. They argue that a no-first-use policy would provide a sounder foundation for defense. The rest of this article is devoted to an examination of that assertion.

A no-first-use policy would relegate nuclear weapons to a single task: retaliation for a nuclear attack. Such a policy

© 1984 SCIENTIFIC AMERICAN, INC

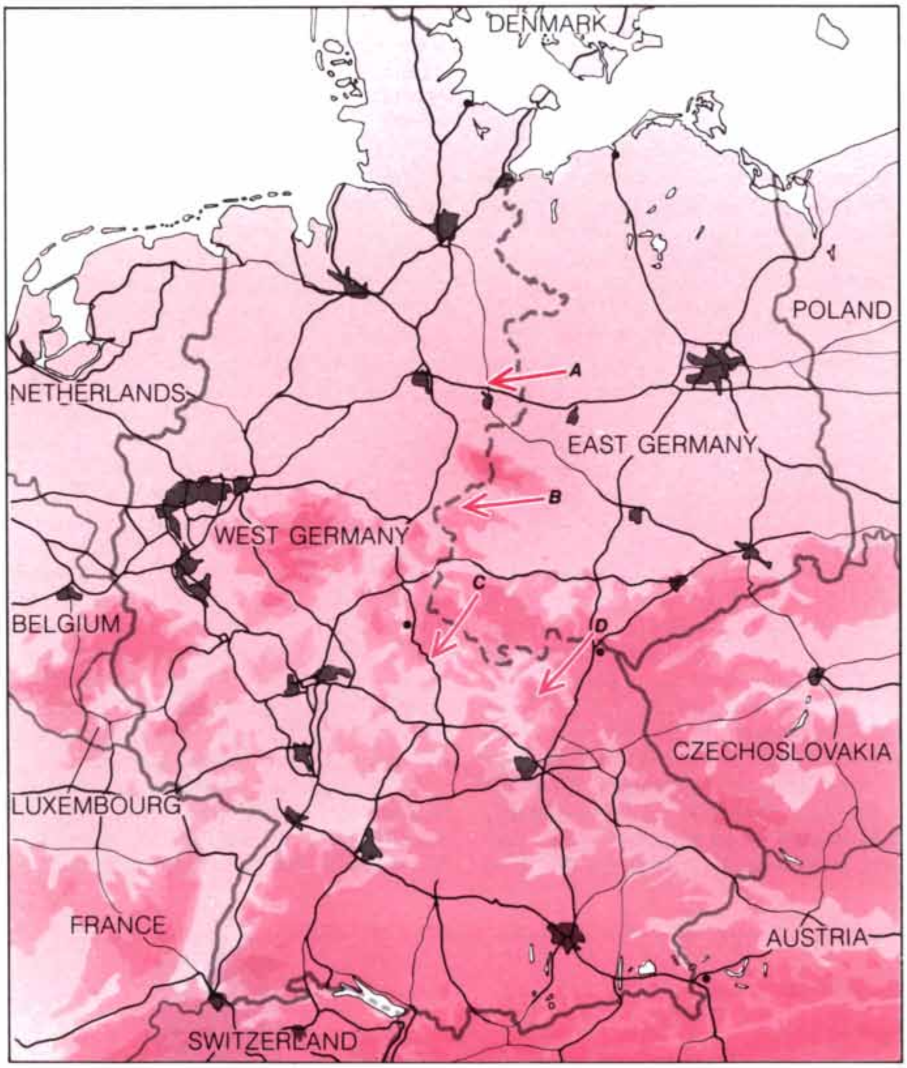
would be much more than a declaration of intent; it would require substantial changes in the training, planning and equipping of NATO's conventional forces, so that the ultimate reliance on nuclear weapons for military operations would be eliminated. Once the policy was adopted no combat commander could expect nuclear weapons to be used for support, just as today there is no expectation of a first use of nerve gas or biological weapons.

The widespread misperception that the U.S.S.R. now has an overwhelming superiority in conventional forces has led some critics to dismiss the no-first-use option as utopian. As the conventional balance of forces is usually represented, it ignores the fact that NATO's task is to defend, not to attack. Although it is true that the Warsaw Pact is stronger in conventional arms than NATO, it must be remembered that in a conflict between armies of comparable competence an attacker must hold a substantial advantage to be confident of success. The disparity between the Warsaw Pact and NATO armies is not of such proportions. Furthermore, we believe NATO could mount an even more formidable conventional deterrent at or near current manpower levels with the adoption of an improvement program at a cost that is small compared with current expenditures. Such a program could be funded largely by reduced spending on nuclear weapons.

Under a no-first-use policy NATO's conventional forces would have to deter—and if necessary stop—a conventional Warsaw Pact attack without resorting to nuclear weapons. To that end the composition and structure of NATO's present forces would have to be altered, operational plans would have to be modified and much larger stocks of supplies would have to be on hand.

The intelligence services of the Warsaw Pact would be fully aware of these developments. Although the Warsaw Pact leaders may never have absolute confidence in a NATO no-first-use declaration, they would probably be won over to a cautious belief that the new doctrine is actually in force. Similarly, from surveillance of Warsaw Pact military exercises and maneuvers, among other indicators, the NATO intelligence services would know whether the Warsaw Pact's recent no-first-use declaration was reflected in military training and planning.

Under the prevailing first-use policy many people in Western Europe are more frightened by their own defense than they are by the potential attack. This fear threatens the entire alliance with political paralysis and has created concern in military circles that NATO may already have a de facto no-first-use policy without any of the requisite preparation. A no-first-use policy that has been deliberately prepared and agreed on in peacetime would lead to more confident and coherent decisions in a crisis, a step that would in itself enhance deterrence.



CENTRAL EUROPEAN FRONT is considered by NATO planners to be the most likely site of an attack by the Warsaw Pact. Four possible major invasion routes are shown: the North German plain (*A*), the Göttingen corridor (*B*), the Fulda gap (*C*) and the Hof corridor (*D*).

Major changes in the tactical nuclear arsenal should also accompany the declaration of a no-first-use policy. Indeed, there is a consensus, encompassing many adherents of the current first-use policy, that the vast array of NATO tactical nuclear weapons based near the front presents a greater threat to NATO than it does to the Warsaw Pact. The problem is that in the face of a Warsaw Pact offensive there would be enormous pressure to use such weapons rather than to let them be overrun and captured. Under a no-first-use policy most of these systems would lose whatever reason for existence they now have.

It has been said that a no-first-use declaration would be tantamount to abandoning Europe and that the commitment to the present first-use policy is essential to linking the fate of the U.S. to that of its NATO allies. This assertion assumes that the immediate exposure to battle of some 300,000 American troops is a weaker link than the rather implausible promise to expose the U.S. to total devastation in the event of a Warsaw Pact incursion into Western Europe. A no-first-use policy would call for an enhanced U.S. commitment to a realistic and credible conventional defense of its European allies and should thereby engender increased confidence in the transatlantic "linkage." The U.S. would still continue to offer its allies a nuclear guarantee of the most profound nature: its ability to respond to a nuclear attack, which protects them from the threat of nuclear destruction.

A more concrete appraisal of the no-first-use option can be achieved by examining several possible scenarios.

© 1984 SCIENTIFIC AMERICAN, INC

First, consider an all-out preemptive invasion of Western Europe by the Warsaw Pact during a crisis. The very existence of vast nuclear arsenals provides a measure of "extended deterrence" against the first use of nuclear weapons. Neither side, however, could be confident that large-scale nuclear war could be averted whatever the two sides' true intentions were before the outbreak of hostilities—even if nuclear escalation makes no military or political sense. Such an attack would therefore imply that the Warsaw Pact accepted some risk of the conflict's ultimately becoming nuclear.

Second, consider a limited thrust by the U.S.S.R. reaching for one or more lucrative prizes made more attractive by the withdrawal of the nuclear deterrent against conventional attack. Such a prize might be Hamburg; seizing it could give the Russians a strategically useful Atlantic port. Because it is within artillery range of the border, Hamburg might be vulnerable to sudden capture even with an improved NATO defense and with comparatively little risk of escalation, assuming NATO were to abide by its no-first-use declaration. An attack of this kind that left NATO essentially intact would be an extremely risky venture. Such a great gamble by the Warsaw Pact leaders would make sense only if the U.S.S.R. could count on the subsequent disintegration of a demoralized NATO. At least as likely an outcome would be an era of dangerous tension, the cessation of all East-West trade and resumption of the conflict after the West had rearmed itself to the teeth.

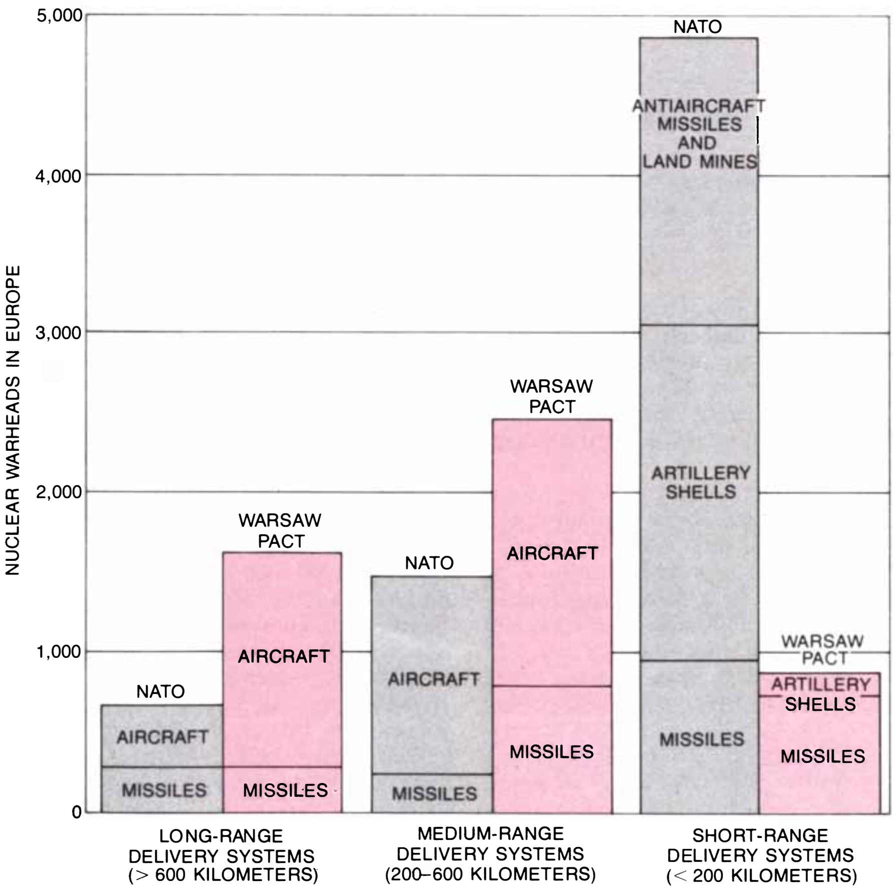
A third scenario is a 1914-style sequence of miscalculation and misperception that leads to war against the will and interests of the participants. This event could come about if NATO felt compelled to mobilize because of a massive influx of Russian troops to quell widespread unrest in the eastern European countries. Many senior officers and political leaders consider this sequence of events to be one of the most likely routes to conflict in Europe. A variant of the scenario involves the spillover of other hostilities between the U.S. and the U.S.S.R. (starting, for example, in the Middle East) into the European theater. In either case a no-first-use declaration would provide the time and relative calm to allow negotiations to begin before the conflict reached proportions comparable to an all-out invasion.

To summarize, a no-first-use strategy would be of some value in the event of a large-scale preemptive invasion by the Warsaw Pact; it might increase the likelihood of reckless adventures such as the surprise capture of a vulnerable NATO asset, and it would be distinctly beneficial in case of an inadvertent or spillover war. As long as the superpowers hold vast arsenals of long-range missiles a massive conventional attack—out of the blue, so to speak—appears to be quite improbable, and it will become progressively more improbable because space-based and airborne surveillance techniques will eventually provide all-weather, night-and-day, real-time intelligence over all areas relevant to the front. Although a limited attack against a target such as Hamburg cannot be ruled out, it would be militarily indecisive and a high-risk gamble from any point of view. Such an action does not conform to the pattern of great caution that has been characteristic of the foreign policy of the U.S.S.R. Blunders and misperceptions are by far the most probable causes of war. Given the ever-present danger of disastrous escalation, a no-first-use policy significantly increases military stability.

Although it is beyond the scope of this article, a similarly strong case can be made for adopting such measures elsewhere in the world. By reducing the risk of local nuclear wars, such measures could make escalation to global nuclear war less likely and thereby enhance the security of all nations.

At first it might appear that all the advantages of a declared no-first-use policy could be achieved by keeping the policy secret or, alternatively, by declaring a policy of "no early first use," according to which one would plan not to use nuclear weapons in the early stages of a conflict. The latter policy is indistinguishable from the present NATO policy, and, as we have argued, it shows little promise of producing the concrete military measures that would free NATO troops of their dependence on nuclear weapons. If the political and military benefits of a no-first-use policy are to be garnered, the policy must be publicly accepted by the leadership and citizenry of the NATO countries, otherwise the necessary improvements in military preparedness and the deterrent value of the new policy will never be realized.

In central Europe the U.S.S.R. and its allies are numerically superior to NATO in both troops and weapons. The Warsaw Pact forces are capable of launching a formidable offensive from their peacetime positions or progressively more powerful attacks at later stages of mobilization and deployment. A major Warsaw Pact offensive that is meant to be decisive would probably move along one or more major east-west axes: the North German Plain, the Göttingen corridor south of the Harz Mountains, the Fulda gap or (less probably) the Hof corridor [*see illustration on*



NUCLEAR WARHEADS IN EUROPE are categorized according to their means of delivery and approximate range. The chart does not include the large numbers of strategic, or intercontinental-range, nuclear weapons deployed by both sides on land, submarines and aircraft.

© 1984 SCIENTIFIC AMERICAN, INC

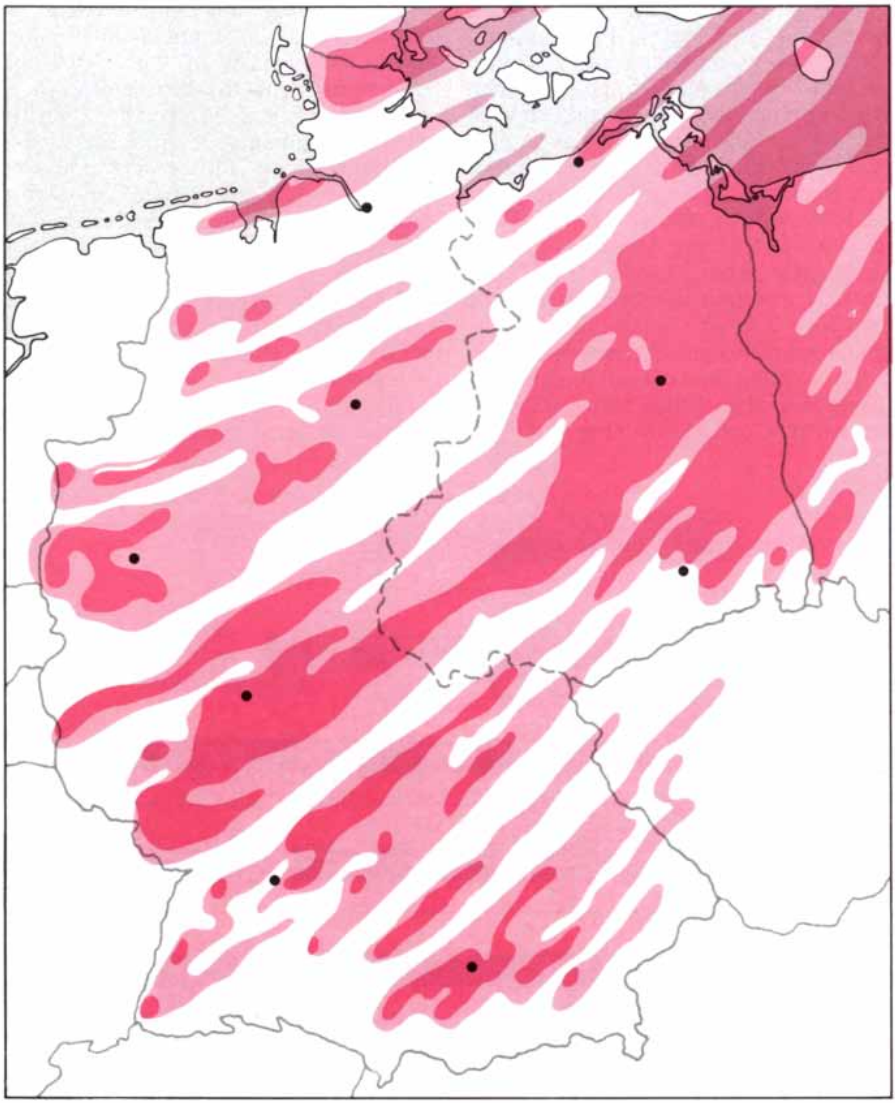
page 35]. These expected attack corridors are defined by the terrain and road system of West Germany and by the most likely campaign objectives of the Warsaw Pact forces.

The offensive would be designed to rapidly achieve its initial objectives, such as the line of the Rhine. The attack could take the form of either a blitzkrieg or a broad frontal offensive. In a blitzkrieg the attacker concentrates armored forces at one or more points, perhaps initially as narrow as three to five miles, and suddenly drives forward to create a breakthrough. Thereafter the attacking forces pour through the gap and race deep to the opponent's rear, avoiding set battles but destroying the communications system and taking key points vital to the defense.

A broad frontal attack exploits firepower (for example massed artillery) to wear down and defeat the opposing forces. In either case the Warsaw Pact forces would presumably apply their system of "attack by echelons," meaning that as one major unit is worn down it goes to the rear for regrouping and is replaced in its entirety by a fresh one.

NATO would meet these attacks by forcing the enemy to traverse a belt of terrain occupied by small, mutually supporting antitank units capable of holding their ground in difficult terrain and of impeding the enemy forces in the major attack corridors. The attack would be met there by NATO's main armored and mechanized forces, taking advantage of both prepared positions and intrinsic mobility.

A number of military and geopolitical factors have led NATO to adopt a policy of forward defense. For one thing, the eastern side of West Germany is the best defensive terrain short of the Rhine. For another, the lateral lines of communication are excellent close to the western side but limited on the east. Finally, there is little ground to give without incurring losses of critical importance.



PATTERN OF RADIOACTIVE FALLOUT resulting from a hypothetical "limited" nuclear war confined to targets in East Germany and West Germany is represented here for the wind and weather conditions prevailing on a typical June day in central Europe. The fallout pattern is based on a scenario in which it is assumed that each side launches a preemptive nuclear attack on military targets (missile sites, air bases, nuclear-weapons storage depots and the like) on the other side. The attacks produce a total of 171 ground bursts of 200-kiloton nuclear explosives. The colored areas correspond to doses of radiation measured in rads (short for "radiation absorbed dose"): the light-color areas are between 200 and 600 rads, the dark-color ones more than 600 rads. Deaths from radiation sickness would begin at doses of less than 200 rads. The illustration is based on a study by William Arkin, Frank von Hippel and Barbara G. Levi that appeared originally in *Ambio,* a publication of the Royal Swedish Academy of Sciences.

The forces that NATO and the Warsaw Pact are potentially able to bring to bear in these engagements are actually quite well matched numerically, given that NATO would be on the defensive. The U.S.S.R. and its allies are superior in number of men in combat units, in number of divisions (although the Warsaw Pact's divisions are somewhat smaller than NATO's) and in "armored divisional equivalents," a measure of the strength of a fighting unit based on the number and performance of the weapons it deploys. This margin not only holds for peacetime deployments but also would hold during and after mobilization. Nevertheless, defending forces traditionally have an advantage over the offense. Given prompt mobilization by NATO, the ratio of forces in the combat theater would at all stages remain below the two-to-one ratio generally accepted by military planners as being necessary for a successful attack against a competent defender in prepared positions.

In terms of the ratio of combat-ready troops NATO has a slight edge in peacetime, but the Warsaw Pact, with the advantages of the initiative and shorter deployments, would pull ahead and peak below the critical two-to-one ratio briefly about two weeks after the start of mobilization. Thereafter NATO reinforcements would predominate for several months until the possible arrival of some low-readiness reserve divisions from the U.S.S.R.

The Warsaw Pact would concentrate its forces along the main axes of attack, aiming at overwhelming superiority at the breakthrough points. The strength of the Warsaw Pact forces is adequate to reach ratios approaching 10 to one at such points about two weeks after mobilization, if NATO fails to reinforce the divisions under attack. By reinforcing from available reserves or adjacent divisions, however, NATO can hold the ratio to three or four to one at two weeks and at lower levels thereafter; thus the attacker's advantage would remain below the four- or five-to-one ratio needed locally for a breakthrough.

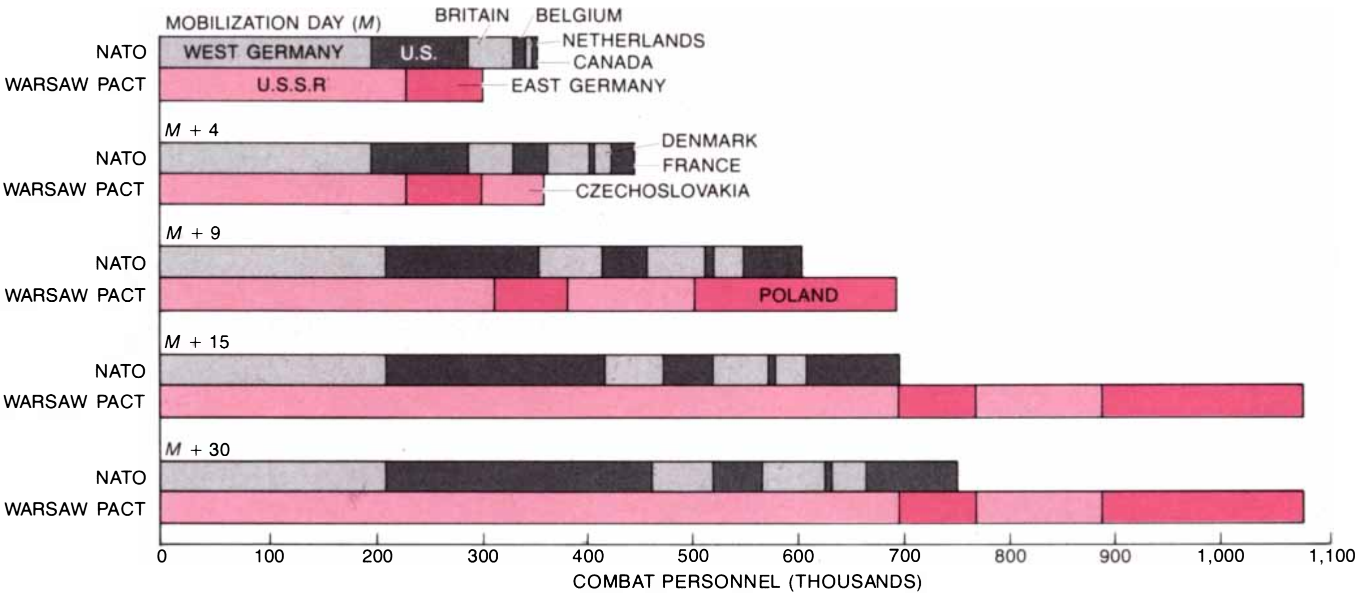
Similar conclusions follow from an

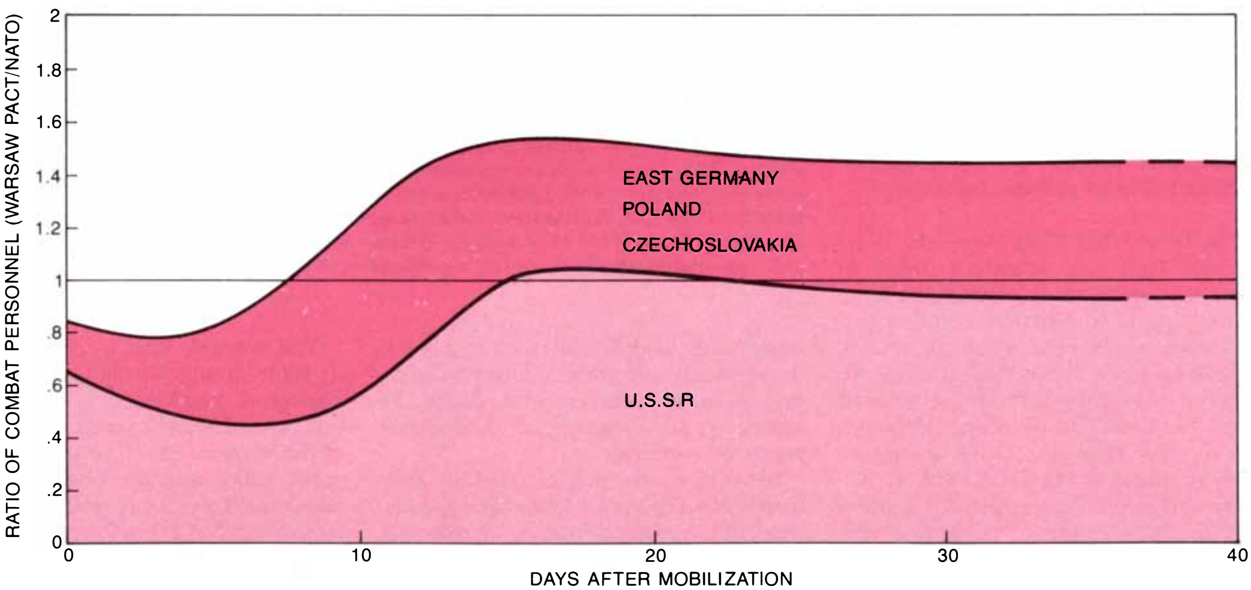
© 1984 SCIENTIFIC AMERICAN, INC

assessment of the balance of air forces. At present the Warsaw Pact outnumbers NATO in combat aircraft by a margin that would remain roughly constant after mobilization. NATO aircraft typically have a longer range and larger payload, however, and there are indications American and Western European pilots are better trained. Moreover, most Warsaw Pact aircraft are designed for a single purpose, whereas many NATO aircraft are multipurpose ones that can be used for air-to-air combat, bombing or launching air-to-ground missiles.

In assessing the overall balance of forces several uncertain political factors must also be considered. One of them is whether or not the NATO countries would be able to act quickly and in unison in the event of an attack. Another uncertainty concerns the role of France. Although France is formally a member of the alliance, it does not participate directly in the NATO military-command structure. France maintains three active divisions in West Germany, but it is not known how soon these forces could be integrated into the NATO command. Nevertheless, it is considered most unlikely that France would remain neutral in the event of an attack by the Warsaw Pact.

The questionable reliability of non-Russian Warsaw Pact forces is a serious problem for the U.S.S.R. that is not measured by any assessment of numerical strength. Would Polish and Czechoslovak troops participate fully in combat not directly linked to their own defense? Could East German troops be used in an invasion of West Germany? Indeed, it is possible that the U.S.S.R. would have to commit its troops to guarantee the loyalty of its allies in a war against NATO. Furthermore, the U.S.S.R. would have to plan for the possibility that China would take advantage of the hostilities in Europe, which would limit the ability of the U.S.S.R. to call on its strategic reserve for the European theater.





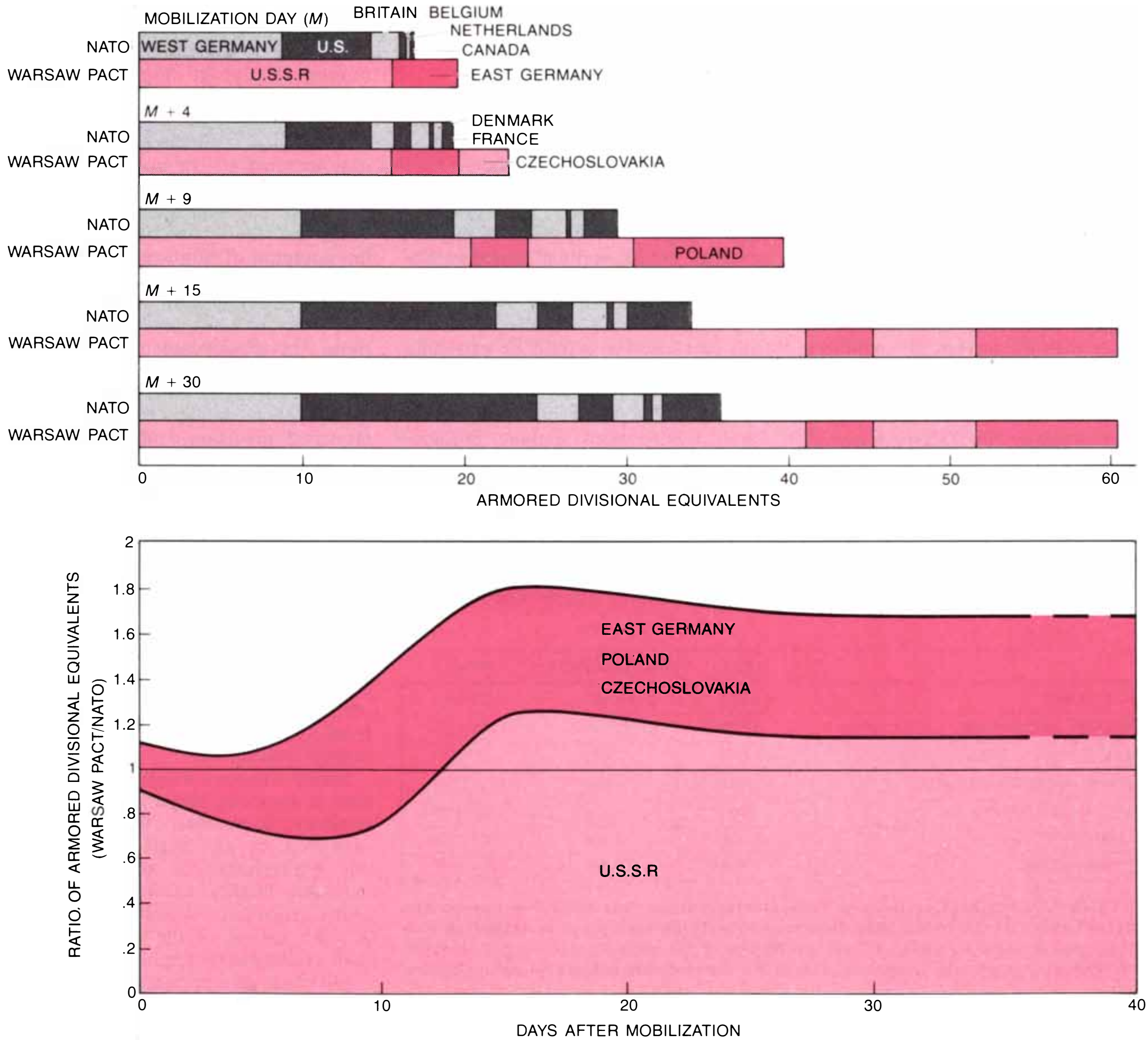
RATIO OF COMBAT-READY GROUND FORCES available to NATO and to the Warsaw Pact at the central European front would fluctuate somewhat during the 30-day period following mobilization, but at no point would it approach the critical two-to-one advantage generally thought to be necessary for a successful attack. The balance of forces is presented here in two ways: bar charts (*top*) and graphs (*bottom*). The left side of the illustration gives the balance in terms of combat personnel; the figures are only approximations, since the number of troops in major combat units varies in peacetime. The right side gives the balance in terms of "armored divisional equivalents," a standard measure of military strength developed by the U.S. Army that stresses weapons effectiveness rather than manpower. The

© 1984 SCIENTIFIC AMERICAN, INC

All the force ratios given above and their military significance are dependent on a number of assumptions: rapid and effective deployment on both sides, quick NATO decision making, high readiness of Warsaw Pact units on mobilization, excellent intelligence and so forth. Neither these factors nor others that could be cited prove what the outcome of a major conflict would be. They do establish, however, that far from having overwhelming superiority, the Warsaw Pact's forces should be considered barely adequate at most for a central European offensive.

There are a number of weaknesses and inadequacies in NATO's conventional defenses that should be corrected if NATO is to face the Warsaw Pact confidently without the option of first use of nuclear weapons. The steps that should be taken fall into four categories. First, NATO must be able to make quicker political decisions and to deploy supporting forces and reinforcements promptly from Europe and overseas. Second, a number of operational improvements should be made, including the construction of tank obstacles and field fortifications and better organization of reserve forces to cover the threat of a long war. Third, NATO should improve its present advantage in advanced technology. Finally, NATO's ability to sustain military operations should be strengthened.

Traditionally NATO has reached decisions unanimously. Consequently the decision to move to a high state of readiness, to deploy forces in Europe and to mobilize reinforcements may not come in time. Indecision on these points is a crucial weakness that cannot be corrected by improvisation in a crisis. NATO should improve its ability to respond to a Warsaw Pact attack after a short period of mobilization, even though that is not a likely scenario for the outbreak of war. If NATO detects the Warsaw Pact's preparations and



differences between the two measurements are a rough indication of the uncertainty in the estimates. The U.S.S.R. has about 20 low-readiness divisions that would take three to four months to mobilize. The U.S. could make a comparable late addition to NATO's forces if it included the units now earmarked for the Rapid Deployment Forces or the National Guard. Hence the force ratio at 30 days after mobilization ($M + 30$) would probably persist for another two months (until $M + 90$), and it could then become either slightly more or slightly less favorable to NATO. The Warsaw Pact forces would include substantial numbers of troops from East Germany, Czechoslovakia and Poland. The questionable reliability of these troops, not reflected in such numerical assessments, could present a serious problem for U.S.S.R.

© 1984 SCIENTIFIC AMERICAN, INC

mobilizes within three or four days, the balance of forces should allow for a successful defense. If NATO's decision is delayed, however, and its mobilization is a week or more behind that of the Warsaw Pact, the latter's superiority in deployed forces grows beyond the danger point and remains high.

NATO's ability to respond promptly and effectively would also be significantly enhanced if France would clarify what it would do if there were a Warsaw Pact mobilization or attack. For example, France could take tangible steps that would give the NATO command greater assurance that French facilities would be available in a timely fashion, and it could join in a publicly disclosed decision to create a secondary line of communications in France as a backup to the primary NATO line.

NATO should also improve its ability to move forces in Europe promptly from their peacetime locations to the central front. The peacetime garrisons of NATO's forces in Europe are not well positioned for rapid deployment. A practical solution would rely on light infantry, which could be quickly deployed along the central front. Armed with modern antitank weapons, these forces could be garrisoned close to or within their assigned sectors, and they could be bolstered rapidly with local reserves familiar with the terrain. By exploiting the defensive value of the terrain these troops would delay, wear down and channel the attack while masking the counterattack by NATO's more heavily equipped divisions. Unequivocally defensive, the light-infantry forces would have the additional virtue that in a crisis they could be brought to full readiness without exacerbating an already tense political situation. Light-infantry units now exist in a number of allied armies, and many more could be formed from reserve manpower that can be quickly mobilized.

Readiness could also be significantly improved by completing the U.S. POMCUS program. POMCUS (Pre-Positioned Materiel Configured to Unit Sets) deploys equipment for a number of U.S.-based divisions and support units in central Europe so that the personnel can be flown directly to their equipment. The objective is currently placed at six division sets (plus support units) in Europe. Four sets are in the field, but two are yet to be authorized. The full program would provide for a prompt, powerful, mobile reserve in each army-group sector; that would substantially improve the balance of conventional forces at the most critical point in the buildup process: two weeks after the Warsaw Pact begins mobilization.

It is remarkable that NATO has no fortifications on any of its fronts. This is contrary to all previous military practice and is not justified by modern technology or the evidence of recent military history. Indeed, fortifications play a major role in current Russian military practice. Yet for political reasons obstacles and field fortifications have not been constructed in West Germany.

A system of tank obstacles backed by field fortifications would be extremely important. Obstacles such as ditches, minefields and concrete forms are comparatively cheap and simple to provide. Fixed fortifications include prepared emplacements for tanks and artillery, observation posts for controlling long-range artillery, clusters of antitank guns, remotely controlled, precision-guided munitions and numerous small concrete bunkers to provide shelter and support for light-infantry forces. The effectiveness of modern artillery fire, relying on modern reconnaissance, has made large permanent fortifications obsolete. Nevertheless, small, hardened installations that can be camouflaged and distributed randomly could still be effective. In some circumstances obstacles can be created as needed by means of projectable minefields or buried explosive hose.

The NATO alliance must also continue its efforts to meet the threat of a protracted conflict. This task need not entail costly increases in military personnel or a return to conscription in nations that do not have it. It can be achieved by increasing the number of U.S. and allied reserve divisions. Reserve forces can provide a powerful supplement to active-duty forces at approximately half the cost per unit.

NATO's present lead over the Warsaw Pact in a number of critical areas of military technology should be preserved and exploited. The 1973 Arab-Israeli war and the more recent clashes between Syrian and Israeli forces in Lebanon demonstrate the importance of innovative military technology. "Smart weapons," or precision-guided munitions, are an outstanding example of the potential of military technology. A smart weapon is one designed to hit and destroy its target with a probability of 50 percent or more under battle conditions. The effectiveness of smart weapons was demonstrated in the Falkland Islands operations, where both sides inflicted serious damage by means of air-launched, precision-guided munitions.

Smart weapons now being developed will carry multiple target-seeking and target-discriminating warheads. A single such round could destroy several tanks, artillery pieces and infantry-fighting vehicles. Coupled with tank obstacles and fortifications, smart weapons can provide a powerful antidote to tanks. The range and flexibility of these weapons and their control systems are such that their fire could be called in whenever needed anywhere along the front where advantage could be taken of prompt intelligence.

Smart weapons would also be valuable in attacking targets deep in enemy territory, where aerial bombardment is restricted by the high concentration and effectiveness of air and ground defenses. Highly accurate long-range cruise missiles, launched from the ground, the air or the sea and armed with conventional explosive warheads, could cause greater damage to deep targets with less exposure and lower losses than piloted aircraft can.

It would of course be a mistake to overemphasize technological solutions or to rely exclusively on any one system. In the case of precision-guided munitions, for example, their overall effectiveness is sensitive to the development of countermeasures. Nevertheless, it is

	NATO	WARSAW PACT	RATIO
HEAVY TANKS	17.629	46,500	1 TO 2.64
ARTILLERY AND ROCKET LAUNCHERS	9,829	20,300	1 TO 2.07
SURFACE-TO-AIR GUIDED MISSILES	1,662	6,293	1 TO 3.79
ANTITANK GUIDED MISSILES	4,644	1,822	2.55 TO 1
ARMED HELICOPTERS	795	1,406	1 TO 1.77
ANTIAIRCRAFT GUNS	5,207	6,486	1 TO 1.24
ANTITANK GUNS	996	3,724	1 TO 3.74

MILITARY EQUIPMENT available on both sides for a major conventional, or non-nuclear, conflict in Europe is listed in this table. The comparatively low numbers of most kinds of non-nuclear weapons deployed by NATO are a reflection of the alliance's dependence on a policy of "first use" of nuclear weapons to forestall a conventional military defeat in Europe.

	NATO	WARSAW PACT	RATIO
TACTICAL AIRCRAFT (*M* DAY)	2,100	2,700	1 TO 1.2
TACTICAL AIRCRAFT (*M* + 15)	3,700	4,900	1 TO 1.3

TACTICAL AIR FORCES of NATO and the Warsaw Pact are quite evenly matched now and are expected to remain so after mobilization. The numerical edge of the Warsaw Pact forces is apparently offset by the better-trained pilots and more versatile aircraft of the NATO forces.

© 1984 SCIENTIFIC AMERICAN, INC

important to recognize that advanced technology will continue to play a significant role in defense.

	COST (BILLIONS)
TANK OBSTACLES, FIELD FORTIFICATIONS, AIRBASE SHELTERS	$1
INCREASED STOCKS OF COMBAT SUPPLIES (FOR 29 DIVISIONS)	$40
IMPROVED *POMCUS* PROGRAM	$3
IMPROVED U.S. SEALIFT CAPACITY	$4
FIVE ADDITIONAL U.S. RESERVE DIVISIONS	$28
THREE ADDITIONAL ALLIED RESERVE DIVISIONS	$15
TOTAL	$97

ESTIMATED COSTS of improving the ability of NATO's conventional forces to withstand any invasion of Western Europe without resorting to nuclear weapons would add up to about $100 billion over six years. "POMCUS" stands for Pre-Positioned Materiel Configured to Unit Sets; under this program equipment for a number of U.S.-based divisions and support units is pre-positioned in Europe so that the personnel can be flown directly to their equipment.

Major hostilities between NATO and the Warsaw Pact would consume combat supplies at an unprecedented rate. A day's supply of fuel, spare parts, replacements and ammunition is much larger and more expensive than it ever was in the past. Investment in stockpiles has had a low priority because of the assumption that a major conflict between NATO and the Warsaw Pact would quickly escalate to nuclear warfare; as a result stockpiles of combat supplies for the central European area are at a low level. In war simulations some NATO units have exhausted their supplies in just two weeks of intensive operations. Under a no-first-use policy it would be essential to raise these stocks to an interim goal of 30 days of supply and, in the long term, to a level of not less than 45 days. As a first step NATO should increase the stockpiles for its standing divisions to cover an additional two weeks of combat.

A number of comparatively inexpensive steps can be taken to enhance the survivability of NATO's air bases, command centers and supply depots, which could be subject to heavy air attack by the Warsaw Pact. The construction of aircraft shelters on allied air bases in central Europe is one critical example. These facilities have not been built for the large numbers of air-force reinforcements that would arrive in the first days of the buildup. Failure to construct this highly effective and comparatively inexpensive defense for the vulnerable, costly and vital air-force reinforcements may be the most notable false economy in the NATO defense program.

Cost estimates have been made for the main improvements suggested above. Taken together, they would amount to less than $100 billion over a period of six years. We did not estimate the cost of additional measures, such as the creation of a light-infantry reserve force along the central front. It is reasonably clear, however, that these additional measures would not cause the overall cost of the improvements required by a no-first-use policy to exceed spending levels that are politically and economically feasible.

The $100 billion we project for the high-priority improvements would entail an annual increase of approximately 2 percent in real terms in NATO's total annual defense expenditures of $300 billion. This increase would be below the current level of 3 percent that NATO members agreed to in 1978 and considerably below the 4 percent recently advocated for the 1983–88 period by General Bernard W. Rogers, the Supreme Allied Commander in Europe.

Funds for implementing these conventional improvements could come from a restructuring of NATO's military priorities. In particular funds allocated to nuclear-weapons procurement could be shifted to meet the cost of many of the higher-priority conventional programs. To cite just one example, cancellation of the MX missile program could save more than $20 billion.

It is our thesis that the adoption of a no-first-use policy would be to NATO's military and political advantage, independent of actions by the U.S.S.R. and its allies, and that this decision is therefore not contingent on negotiations. Nevertheless, there are two other critical issues that are fundamental components of the European confrontation in their own right and can be addressed only by negotiations. If they could be resolved, they would significantly increase European security and diminish the political hurdles faced by a no-first-use policy. We refer here to the 10-year-old Mutual and Balanced Force Reduction (MBFR) negotiations, which seek to establish a stable conventional balance in Europe, and to proposals for a nuclear-weapon-free zone straddling the East-West border.

As currently conceived, the primary feature of an MBFR agreement would be the reduction of active-duty, ground-force manpower in central Europe to a ceiling of 700,000 men on each side, and a ceiling of 900,000 for ground- and air-force personnel combined. The agreement would also prohibit the introduction of additional military units into the regions affected, so that the force levels necessary for a successful attack would not be available. Proposals for verifying compliance and monitoring departures from normal patterns of military activity are under negotiation.

An MBFR treaty would reduce fears that the strengthening of NATO's conventional forces would simply trigger a conventional arms race. It would also provide a considerable measure of assurance that neither side was preparing for a conflict. If such a treaty were breached, that would give the NATO political leadership clear evidence of hostile intent and thereby justify intensified defensive efforts, including prompt mobilization.

The proposal for a nuclear-free zone in Europe that has received the broadest and most influential level of international support is the one made by the Independent Commission on Disarmament and Security Issues, which was chaired by Olof Palme, now prime minister of Sweden. The Palme commission issued a report in June, 1982, that called for a nuclear-weapon-free zone on each side of the central front, initially reaching almost to the midpoints of East Germany and West Germany. The zone would be extended gradually to include the entire European front between the two alliances [see "A Nuclear-Weapon-Free Zone in Europe," by Barry M. Blechman and Mark R. Moore; SCIENTIFIC AMERICAN, April, 1983]. In view of our earlier comments on the threat presented by forward-based nuclear systems, it is clear that the Palme commission's proposal would enhance the security of NATO and would form a natural complement to a no-first-use policy.

The essential requirement for completion of the MBFR negotiations is sustained support from top government officials on both sides. Strong political leadership on both the domestic and the international front would be necessary to turn a no-first-use declaration, an MBFR treaty and the Palme commission's recommendations into reality. Such a goal, if it could be achieved, would mark a major reduction in the risk of war in Europe.

The adoption of a no-first-use policy would have profound consequences. In contrast to arms-control measures, which rarely constrain the actual use of weapons, a no-first-use policy would transform the conceptual foundation on which military strategy and planning rest. A no-first-use regime, as defined in this article, would strengthen the cohesion of the NATO alliance and make relations with the U.S.S.R. less perilous in times of crisis, thereby markedly lowering the risk of nuclear war.

© 1984 SCIENTIFIC AMERICAN, INC

The Dynamic Abyss

Cold currents flowing toward the Equator in the deep ocean are often agitated by powerful storms. These disturbances transport huge volumes of sediment across the ocean bottom

by Charles D. Hollister, Arthur R. M. Nowell and Peter A. Jumars

"It was a dark and stormy night." This much-parodied opening line from a Victorian novel accurately describes conditions 4,800 meters below sea level at the base of the continental rise 450 miles off the coast of New England in the middle of September, 1979. It is always dark, of course, at depths below 1,000 meters. The instruments we had moored some weeks earlier at the much greater depth, at the base of the Scotian Rise, revealed the presence of a storm that lasted for about a week. Our instruments showed that a massive bottom current was flowing toward the southwest at more than half a meter per second (about one knot), heavily scouring the ocean floor as it passed. The forces on the moored instruments equaled those that above sea level would be generated by a fresh gale of between 34 and 40 knots.

Such an energetic event three miles below the ocean surface came as a revelation to those who grew up believing the abyss was a region as calm as it was dark. Recent studies have now demonstrated that on the western side of the great ocean basins, skirting the foot of the continental rise, periodic storms transport huge loads of fine sediment that dramatically modify the sea floor, scouring it in some areas and depositing large volumes of silt and clay in others. Much of the evidence for such bottom flows has come from a program sponsored by the Office of Naval Research and known as the High-Energy Benthic Boundary-Layer Experiment (HEBBLE).

The general picture that has emerged is that dynamic events in the deep sea broadly mimic the general circulation of the atmosphere, which gives rise to the highly active global weather systems. The mid-latitude eddies, often associated with strong winds and heavy precipitation, have their counterpart in energetic deep-sea eddies that last for from several days up to two weeks and that move a "blizzard" of silt and clay from one place to another. Such storms are not ubiquitous, however, even on the western side of the ocean basins. For example, only about 50 miles west of the HEBBLE site, where the wreckage of the nuclear submarine *Thresher* lies at a depth of 2,500 meters, photographs show no evidence of scouring around the hull where it rests on the sea floor, and very little silt has yet been deposited on the hull's surface. It seems that the storms follow certain well-marked paths, but their frequency at any given site is still poorly known. As a preliminary estimate, erosion-producing storms seem to come about once every two months in the most-studied region.

The periodic scouring of the sea bed and the occasional deposition of thick layers of fine material result in a complex stratigraphy much different from what would develop from the gentle rain of sediments that has been traditionally visualized by marine geologists. The episodic transport of sediment creates layered sequences that look much like those found in shallow seas where strong windstorms create what geologists call tempestites, whose simple layering of deposited material is dominated by overlapping beds of sediment graded into different particle sizes.

Our new findings may cause petroleum geologists to revise traditional ideas about the best places to search for unexploited deposits of oil. The discovery of strong abyssal currents and storms also has important implications for the conduct of submarine and antisubmarine operations, for the placement and design of sea-floor structures and for reliance on the deep ocean for the disposal of wastes.

The notion of a tranquil abyss had been so generally held that many investigators were initially reluctant to accept the evidence for strong currents and storms in the deep sea. The first argument for the existence of such currents came from theory. Cold water is denser than warm water, and models of ocean circulation showed that the sinking of cold water near the poles should generate strong, deep and steady currents flowing toward the Equator. Subsequent observations not only confirmed the presence of the deep currents but also disclosed the existence of eddies on the western side of ocean basins that can be some 300 times as energetic as the mean current. Photographs of the sea floor underlying the deep currents also revealed extensive graded beds indicative of the active transport of sediment. The final evidence for dynamic activity at great depths came from direct measurements of currents and sediments in the North Atlantic carried out in the HEBBLE program.

Before we describe the HEBBLE findings in some detail let us briefly review the sources and sinks of deep-sea sediments and the forces that activate the global pattern of ocean circulation. The sediments that end up on the ocean floor are of two main types. One component is the detritus whose source is the weathering of rocks on continents and islands. This detritus, together with decaying vegetable matter from land plants, is carried by rivers to the edge of the continent and out onto the continental shelf, where it is picked up by marine currents. Once the detritus reaches the edge of the shelf it is carried to the base of the continental rise by gravitational processes. A significant amount of terrestrial materiali s also blown out to sea in subtropical regions by strong desert winds. Every year some 15 billion tons of continental material reaches the outlets of streams and rivers. Most of it is trapped there or on the continental shelves; only a few billion tons escapes into the deep sea.

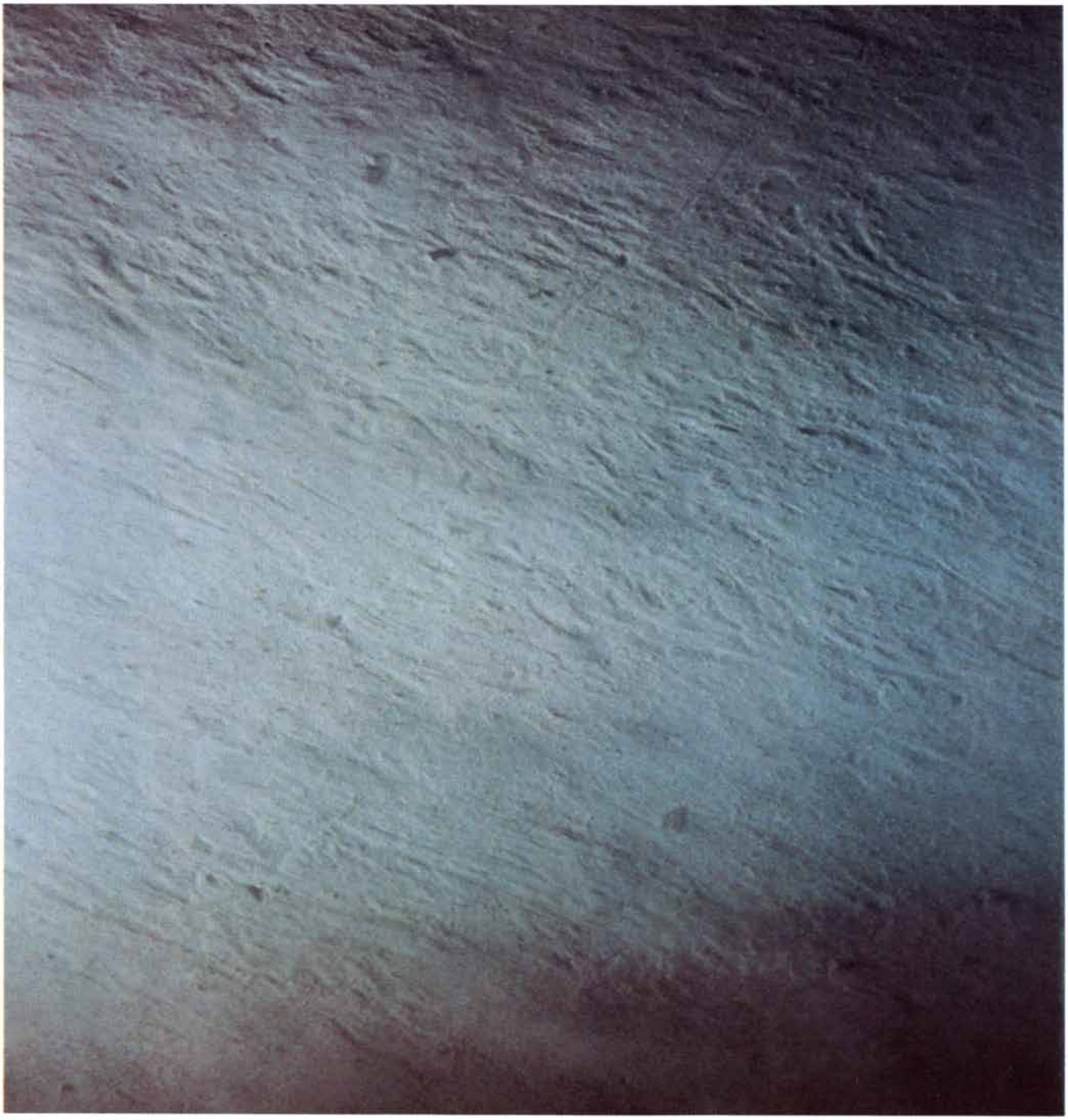
The second major component arriving at the sea floor consists of the shells and skeletons of dead microscopic organisms that flourish and die in the sunlit waters of the top 100 meters of the world's oceans. Such biological material contributes to the total inventory at the bottom about three billion tons per year. Rates of accumulation are governed by rates of biological productivity, which are controlled in part by surface currents. Where surface currents meet they are said to converge, and where they part they are said to diverge. Zones of

© 1984 SCIENTIFIC AMERICAN, INC

divergence of major water masses allow nutrient-rich deeper water to "outcrop" at the sunlit zone where photosynthesis and the resulting fixation of organic carbon take place. Such belts of high productivity and high rates of accumulation are normally around the major oceanic fronts (such as the region around the Antarctic) and along the edges of major currents (such as the Gulf Stream off New England and the Kuroshio current off Japan). Nutrient-rich water also outcrops in a zone along the Equator, where there is a divergence of two major, wind-driven gyres.

The rate of marine-life sedimentation is also strongly influenced by the depth of the water. The ability of a shell to survive being dissolved after its descent to the ocean floor is in large measure determined by how long it rests on the bottom before it is covered and thus protected from the corrosive action of deep-sea water, which is undersaturated in calcium carbonates. In deeper waters deposition is generally slower, shells have farther to fall and they fall through waters that are increasingly corrosive.

If the upper waters were uniformly productive and if there were no bottom currents, there would be a more or less even blanket of material settling onto

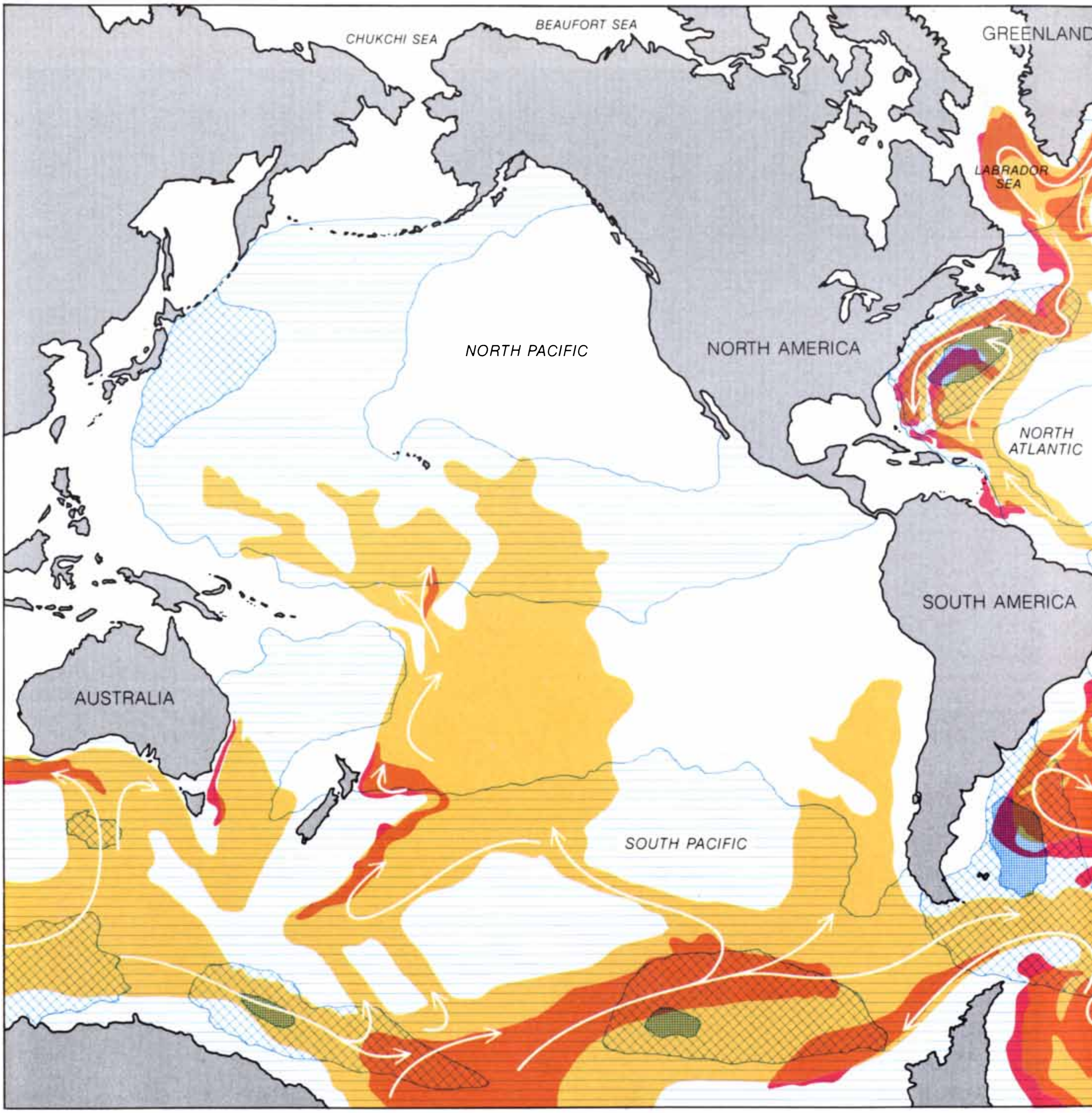


GROOVED AND SCOURED SEA FLOOR was photographed at a depth of 4,850 meters at the base of the continental rise some 450 miles off the New England coast. The site is being studied as part of the High-Energy Benthic Boundary-Layer Experiment (HEBBLE), in which the authors are participating. The project is financed by the Office of Naval Research. The absence of biological features on the sea floor confirms readings from moored flow meters showing that the bottom had been scoured by a deep-sea storm not long before.

© 1984 SCIENTIFIC AMERICAN, INC

the originally volcanic and rugged floor of the world's oceans. This, however, is not the case. As a glance at a world map will show, the largest rivers of North America and South America empty into the Atlantic; that ocean therefore receives considerably more river-borne material than the Pacific. Moreover, the deep-ocean trenches around the Pacific trap much of the material that reaches the ocean's western edge.

The Atlantic is also a smaller and shallower ocean than the Pacific, so that its biological product is buried more rapidly and is therefore more likely to survive than its Pacific counterpart. Thus on the average the floor of the Atlantic receives more sediment per unit area (about a centimeter per 1,000 years) than the floor of the Pacific. In addition the global pattern of strong near-bottom currents creates more potential for redistribution in the Atlantic than it does in the Pacific. For all these

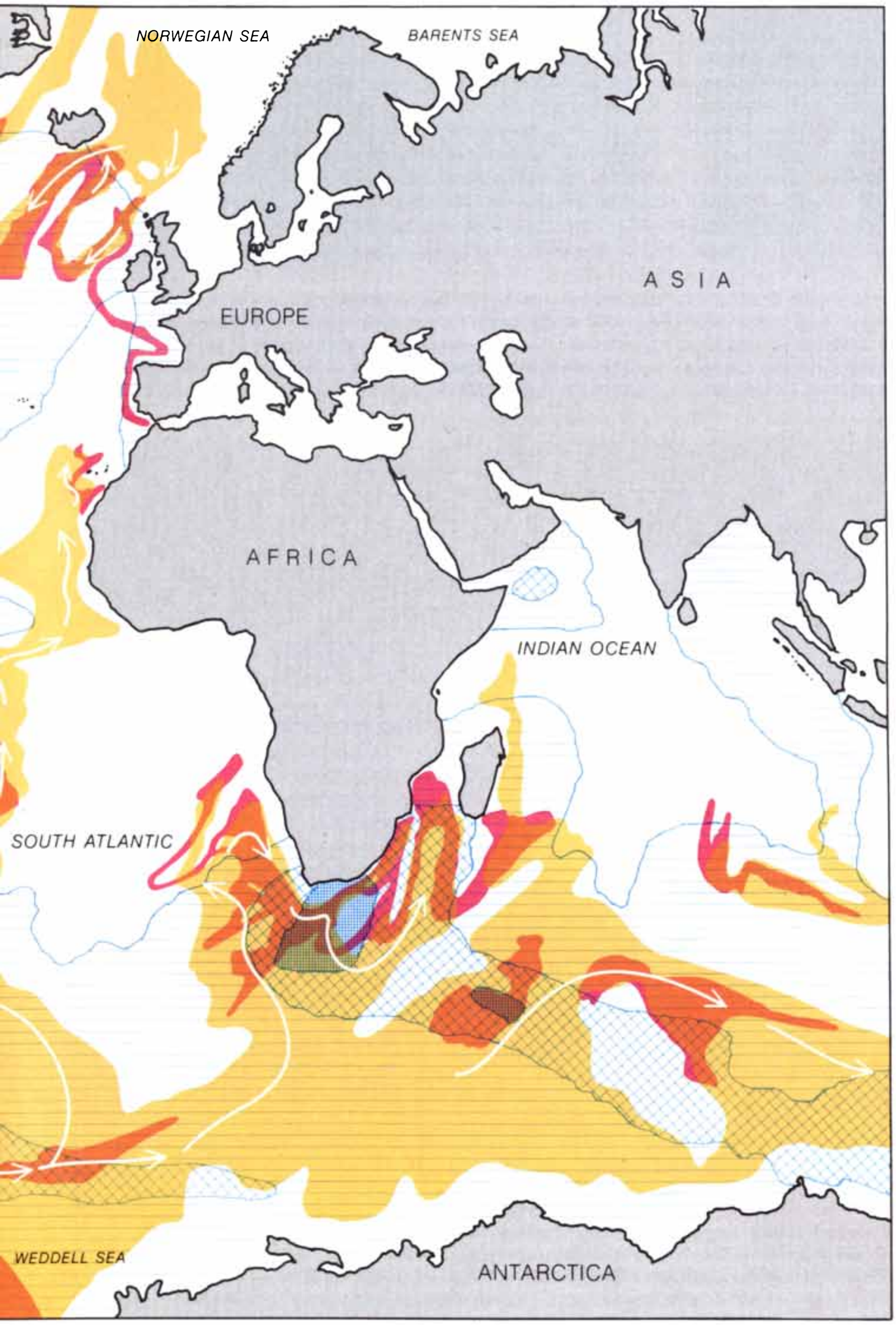


ABYSSAL CURRENTS (*yellow*) transport cold, dense water away from the polar regions into the world's three great ocean basins. Although both the Atlantic and the Pacific basins are connected to the Arctic Ocean, not much deep Arctic Ocean water can flow into the Pacific through the shallow Bering Strait. The Atlantic, however, receives strong flows of cold abyssal water from both polar regions. The currents are guided by the topography of the ocean bottom and are deflected westward by the Coriolis force, a consequence of the earth's eastward spin. The areas in shades of blue hatching represent variations in kinetic energy associated with near-surface eddies. The densest area corresponds to a kinetic energy (expressed in units showing the fluctuation of velocity around a mean) of more than 20 centi-

© 1984 SCIENTIFIC AMERICAN, INC

reasons the Atlantic is the ideal ocean in which to study the geologic effects of energetic deep circulation.

Let us now consider the nature of the bottom currents. In the polar regions the surface water is denser than the water at lower latitudes because it is both colder and saltier. (The salinity of surface water flowing poleward is increased by evaporation; it is further increased by the exclusion of salt from ice as it freezes.) As the water increases in density it sinks and spreads toward the Equator. Early theory suggested this thermohaline flow would take the form of deep currents that would be deflected westward as they flowed equatorward. Such deflection is a simple consequence of the earth's eastward spin, which gives rise to an apparent force: the Coriolis force. The precise path taken by the circulating water is also controlled by the distribution of landmasses and by the topography of the ocean bottom.



meters squared per second squared. The horizontal hatching represents values lying between four and 10 in the same units. The cross-hatching represents values between 10 and 20. The values of eddy kinetic energy are based partly on ship-speed observations and partly on satellite measurements of variations in ocean elevation. The eddy energy is transferred to the sea floor, where it interacts with the abyssal currents to erode and redistribute the bottom sediments, creating large areas (*red*) where the sediments lie in beds of graded particle size.

All three major oceans—the Atlantic, the Pacific and the Indian—are connected directly to the Antarctic continent, where they are fed by deep, cold currents. As these currents flow toward the Equator they are deflected westward by the Coriolis force. Thus a body of water moving to the north in the Southern Hemisphere is forced to the left looking downstream. Deep currents are pressed against the western side of all three ocean basins and the submarine ridges in them. Such squeezing against the continents causes the currents to pick up speed, just as a stream flows faster as its channel narrows. The Indian Ocean is unique, however, in that it is not connected to the northern polar region and therefore has only one source of cold bottom water.

Although the Pacific and the Atlantic are both joined to the Arctic Ocean, the flow of deep cold water into the Pacific is negligible because the North Pacific is effectively blocked off from the Arctic by the shallow Bering Strait that separates Alaska from the U.S.S.R. In addition the water of the Pacific, being less salty than Atlantic water because of a smaller input of salt in rivers, is more readily frozen by the cold subarctic wind, and the near-surface water that remains unfrozen is not sufficiently enriched in salt and so is not dense enough to sink to the bottom.

The situation in the Atlantic is quite different. The Atlantic has two substantial sources of salty water: the water from the Gulf of Mexico, which is carried north by the Gulf Stream, and the deep flow from the Mediterranean. These sources make the surface water of the North Atlantic saltier than that of the North Pacific. The water of the North Atlantic moves northward near the surface and enters the Norwegian Sea, where it is cooled to about zero degrees Celsius without freezing, gets dense and sinks. On reaching the bottom the cold, dense water reverses direction and flows back into the Atlantic through a series of narrow, deep sills in the submarine ridge that connects Greenland, Iceland and Scotland.

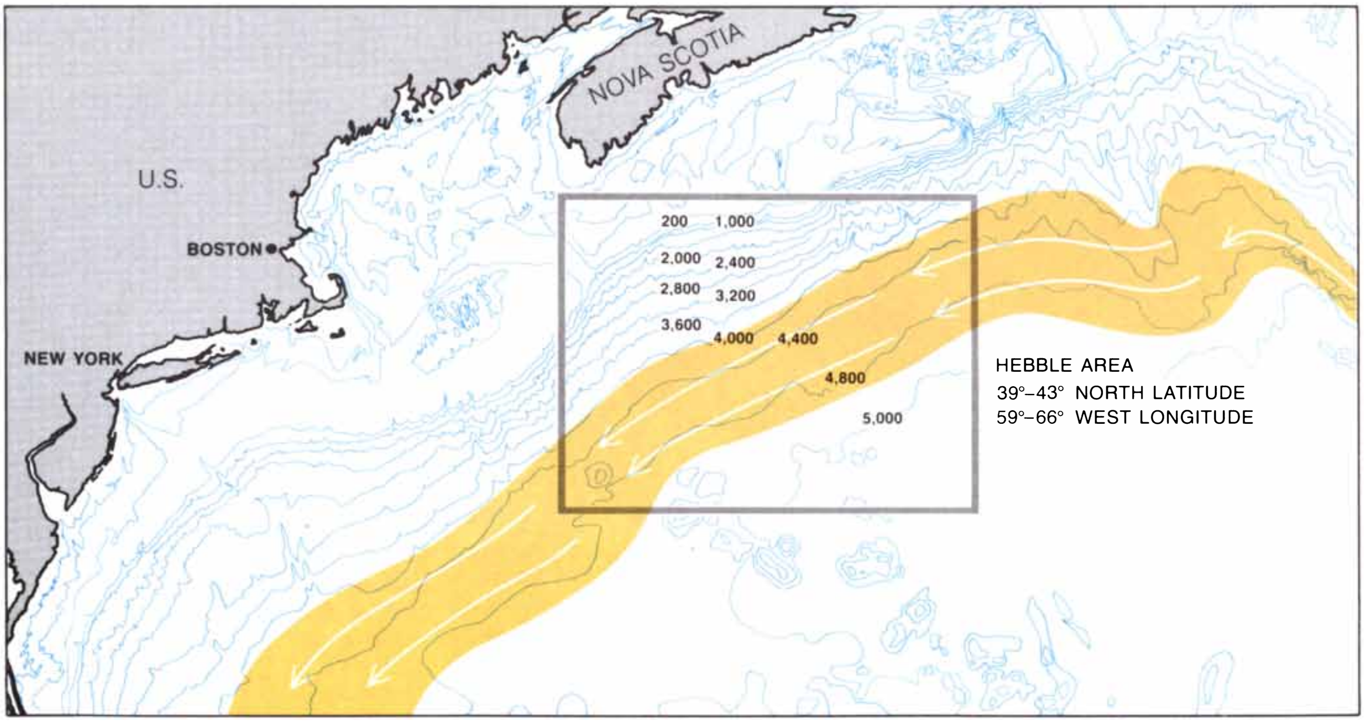
Constrained by topography, this large volume of deep water moves south and is forced to the right, looking downstream, against the continental margin of eastern Canada and the U.S., forming the Western Boundary Undercurrent. The existence of this current was first recognized two decades ago. It is now estimated that the current transports 10

© 1984 SCIENTIFIC AMERICAN, INC

QUIET AND ACTIVE SITES in the deep sea can readily be identified from photographs of the sea floor. In a quiet area (*left*) the surface is seen to be well worked by bottom-living organisms. Thin tubes that have been built by foraminifera, tiny shelled animals, rise as much as 10 centimeters above the sea floor. In an area with strong bottom currents (*right*) the sea floor is scoured and the direction of the flow of the current is revealed by tails of sediment deposited on the lee side of small mounds. The photograph at the left was taken on the continental slope at a depth of 1,000 meters, the one on the right on the New England continental rise at a depth of 5,000 meters.

million cubic meters of water per second along the East Coast of the U.S. During abyssal storms the current is able to carry about a ton of sediment per minute.

With this background one can look at the world oceans and try to identify where there may be current-deposited drifts of mud, possibly modified by abyssal storms. Clearly the greatest volume of inorganic mud and the strongest bottom currents are to be found in the western Atlantic. The high latitudes of the western side of both the North Atlantic and the South Atlantic offer a high potential for abyssal storms, and it is here that bottom photographs reveal a sea floor formed and shaped by strong bottom currents. These graded beds cover the largest drifts of sediment on the planet: up to 200 kilometers wide, 1,000 kilometers long and two kilometers thick. The Blake-Bahama outer ridges off Florida are outstanding examples. It seems plausible that the entire Atlantic continental rise at depths of between three and five kilometers off North America and South America has been shaped largely by contour-following abyssal currents. Deep-sea



SEA-FLOOR AREA STUDIED in the HEBBLE project lies at a depth of three miles centered 450 miles east of Boston at the base of the continental rise. The yellow band shows how cold bottom currents originating near Greenland and in the Norwegian Sea are deflected westward against the North American continental slope and rise as they flow toward the Equator. Periodic storms raise the current velocity at the HEBBLE site from a normal value of a tenth of a knot to one knot or more, stirring up "blizzards" of silt and mud.

© 1984 SCIENTIFIC AMERICAN, INC

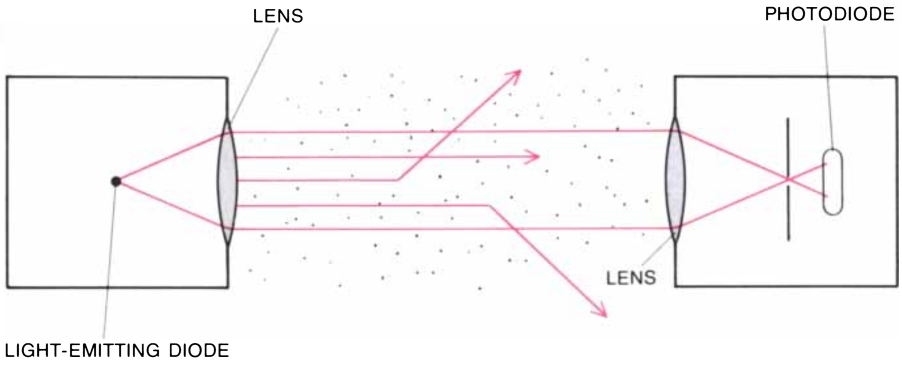
drilling has disclosed that bottom currents eroded large areas of the western Atlantic until some 20 million to 15 million years ago and that since then current-controlled deposition has prevailed, with the formation of large sedimentary drifts.

Elsewhere bottom currents have shaped the distribution of fine-grained material along the edges of Africa, Antarctica, Australia, India and New Zealand. In this entire region the formation of large drifts of mostly biogenic matter is favored by the combination of high biological productivity along the circumpolar divergence and the wind-driven circumpolar current. Although many of these drifts may not be attributable directly to abyssal storms, there is no doubt that truly vast areas of ocean bottom are modified by material entrained by deep currents. The major role of periodic storms may be to stir up sediment, which is subsequently picked up and carried downstream by the less energetic but more persistent bottom currents.

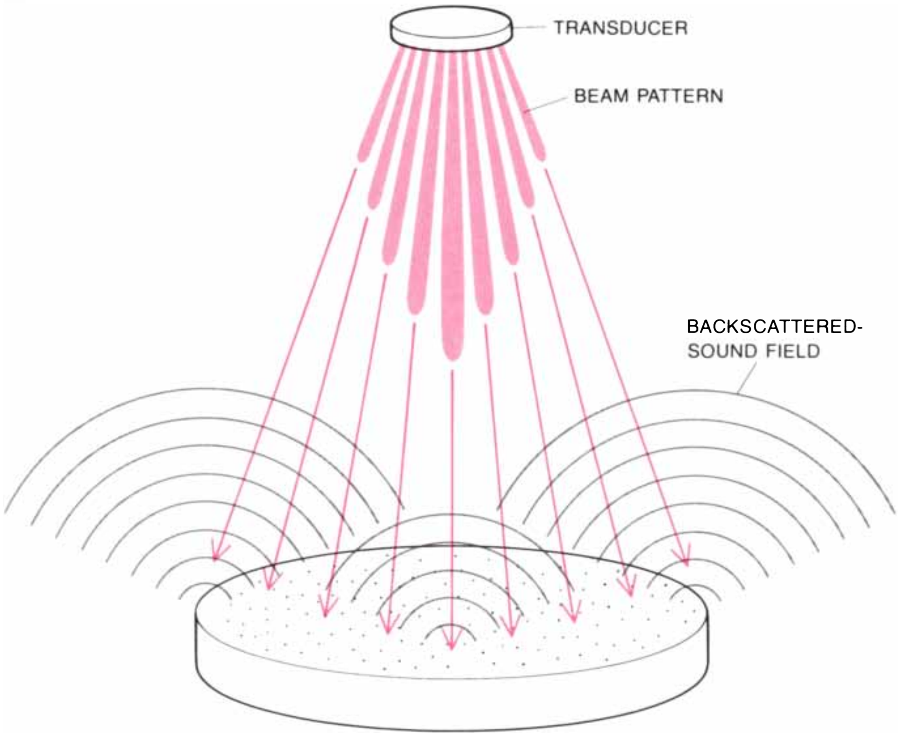
Instruments moored on the sea floor have added important details to the intuitive picture of how sediment is eroded, transported and redeposited. It turns out that the process is not readily duplicated in the laboratory. Fundamentally, to be sure, moving water exerts a stress on the bottom that is ultimately responsible for moving fine-grained material and shaping graded beds. Steady flow over a horizontal bed adopts a characteristic structure. At the interface between the water and the bed a layer of water molecules adheres to the bed and is essentially immobilized. Directly above this interface, if the flow is slow and the bottom is smooth, a thin layer of water will move over the bed in a laminar flow, in which the dominant force is the viscosity of the water. At the height of a centimeter or less the viscous forces are overcome by fluid inertia, and turbulent mixing dominates.

Within the viscous layer, when and where it exists, the velocity of flow increases linearly with height. In the bottom few meters of turbulent flow, mixing is intense and velocity increases exponentially with height. The region is defined as the logarithmic layer because within it a linear curve is obtained when the mean velocity of flow is plotted against a logarithmic scale of height above the bottom. In this region the mass properties of the water, such as salinity, temperature and content of suspended solids, are actively mixed.

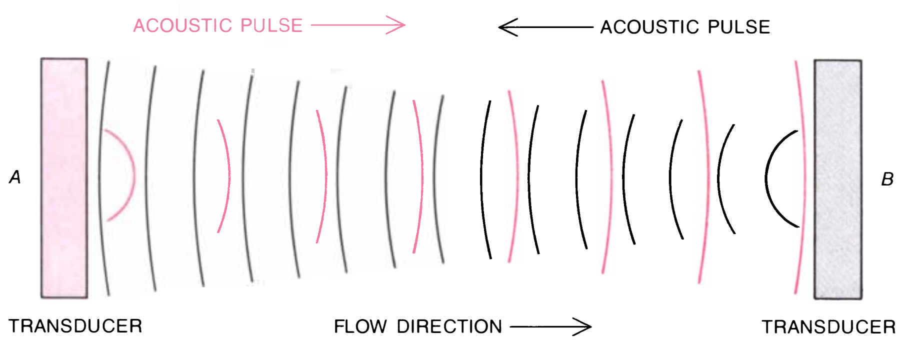
At large vertical scales water motions are influenced by the Coriolis force: in the Northern Hemisphere accelerating fluids are deflected to the west and decelerating fluids are deflected to the east. As currents are frictionally retarded by slower-moving fluid under them and ultimately by the bottom, the current



TRANSMISSOMETER is one of several types of instrument deployed by HEBBLE investigators. The amount of light reaching the photodiode at the right is closely related to the concentration of particles in the light path. Light source emits at a wavelength of 660 nanometers.



ACOUSTIC-BACKSCATTER SYSTEM is another source of information about the concentration of particles in water flowing near the ocean floor. The device emits pulses of sound and measures the amount of energy reflected back from the particles in a given volume of water.



FLOW VELOCITY AND STRESS can be determined by measuring the time required for a sound pulse to travel a known distance. HEBBLE workers call the device shown schematically the benthic acoustic-stress sensor. If the current is flowing to the right, the time a sound pulse takes to travel from *A* to *B* will be less than the time it takes to travel in the opposite direction. The difference in time gives the velocity of the fluid. The quantity known as Reynolds stress in the scanned body of water is calculated from two measurements at right angles to each other.

© 1984 SCIENTIFIC AMERICAN, INC

direction veers to the east, forming a characteristic pattern known as the Ekman spiral (after the pioneering Swedish oceanographer V. Walfrid Ekman). The shear between layers of fluid moving at different velocities continues to fuel turbulent mixing.

The effects of bottom friction do not extend more than 100 meters upward into the water column. This region, in which properties are well mixed by turbulence, is usually called the bottom mixed layer. Its thickness is limited by the attenuation of turbulent mixing, a process that extracts energy from the flow. At some height, proportional to both the velocity of the overlying current and the magnitude of the Coriolis force, shear is no longer sufficient to enhance turbulent mixing above ocean background levels.

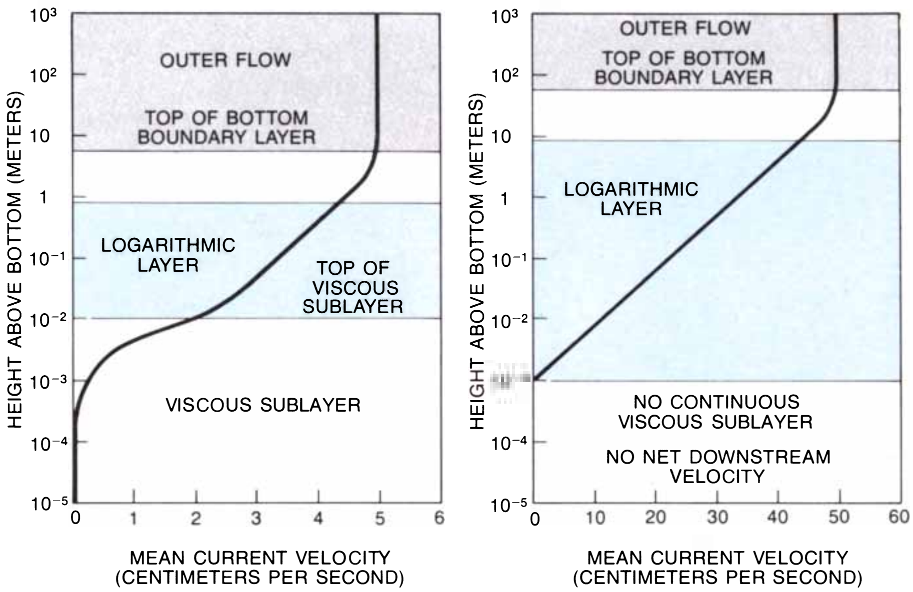
The top of the bottom mixed layer is usually considered to be the limit of thickness of the bottom boundary layer. Unlike the atmosphere, the bottom waters of the ocean either are unstratified in density or are at most stably stratified to a slight degree (saltier, colder water lying under warmer, fresher water). As a result perturbations do not grow as they do in the atmosphere. Any stable stratification would serve to decrease the thickness of the mixed layer by requiring more energy to achieve mixing.

Erosion can be caused only by energy that reaches the sea floor. The rougher the small-scale topography of the sea bed and the faster the overlying flow, the greater the shear stress acting on the bottom sediments. Above the critical level of shear that is just capable of initiating the erosion process the rate of erosion rises exponentially with additional shear. Once sediments are freed from the sea bed they are acted on by opposing forces: turbulence tends to diffuse the particles upward and gravity causes them to settle. The greater the settling velocity, the stronger the concentration gradient close to the bottom. The faster the current is and the rougher the sea bed is, the more uniformly the concentration varies with height above the bottom. Deep-sea sediments typically consist of clays and fine silts whose particle size is less than 30 micrometers. Such small particles tend to settle so slowly that the concentrations of sediment are nearly uniform throughout the bottom mixed layer in the presence of moderate currents. As a result very fine material, once it is suspended by an intense abyssal storm, can be carried long distances by comparatively weak currents.

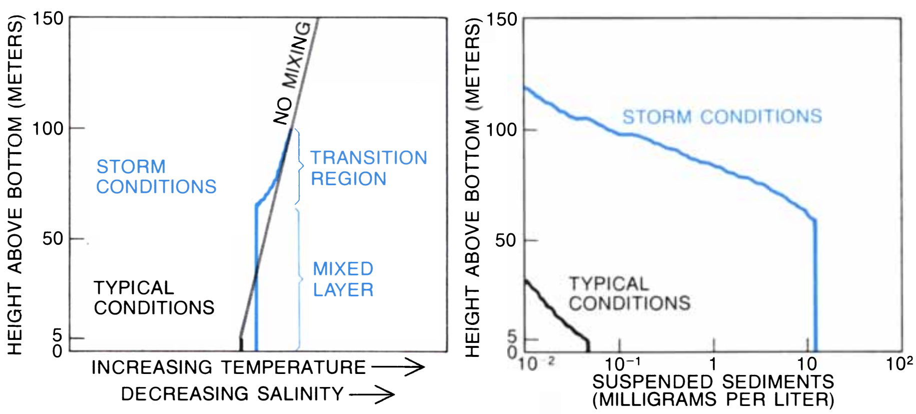
Fine sedimentary particles have a large ratio of surface to volume, so that they are extremely efficient at adsorbing dissolved substances from the water into which they are stirred. Natural clays, for example, have from 10 to 100 square meters of surface area per gram of sediment. Among the substances scavenged by the bottom currents are both natural and artificial radioactive isotopes of the elements. If the rate of input of isotopes from natural radioactive sources and the rate of sedimentation are both roughly constant, as appears to be the case for many isotopes in a number of deep-sea environments, the rate of decrease of radioactivity with depth in the sediments serves as a natural clock for estimating rates of sedimentation.

Provided the sediments are not mixed after deposition, the concentration of an adsorbed radioactive isotope will decrease exponentially with depth at a rate inversely proportional to its half-life. Natural isotopes commonly studied in such sedimentary chronologies have half-lives ranging from 24.1 days (for thorium 234) to 5,730 years (for carbon 14). Also useful in providing absolute dating of sedimentary layers are several radioactive isotopes introduced by atomic-bomb explosions in the atmosphere. These include cesium 137 (half-life 30.1 years), plutonium 239 (24,000 years) and plutonium 240 (6,580 years).

Although the mixing of sediments complicates the estimates of net sedimentation rate, it allows other inferen-



PROPERTIES OF WATER MASS (such as salinity, temperature and content of suspended solids) are actively mixed in the "logarithmic" layer, where mean velocity increases linearly with height when the height above the bottom is plotted on a logarithmic scale. When the flow is slow and the bottom is smooth (*left*), a thin layer of water will move over the floor in a viscous, or laminar, flow. If the flow is rapid or the bottom is rough (*right*), the logarithmic layer will extend all the way from one millimeter to about 10 meters above the ocean floor.



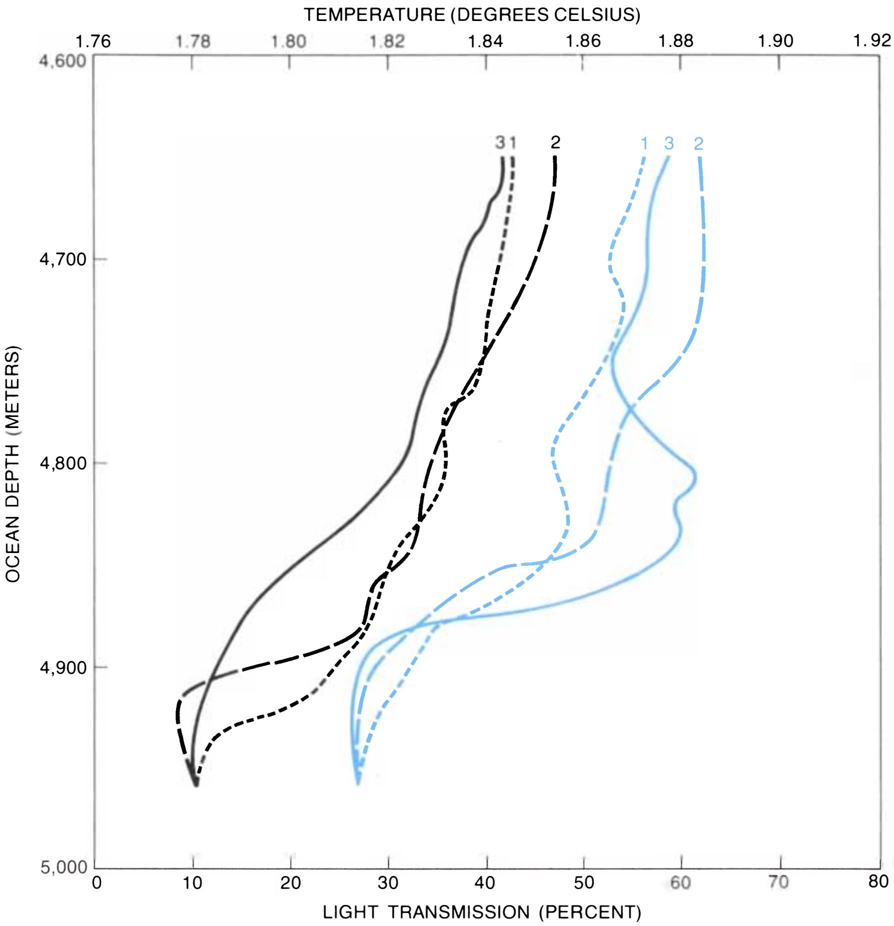
PRESENCE OF STORMS affects the way salinity, temperature and amount of suspended matter are distributed with height above the sea floor. Since salinity and temperature respond similarly, they are represented by the single curve at the left. Under typical abyssal flow conditions (*black*) the two properties increase linearly with height above the bottom except for the lowest five meters or so, where they are mixed to an essentially constant value. Under storm conditions (*color*) enough energy is introduced into the abyssal flow to extend the mixed layer to a height of 60 or 70 meters, with a transition region extending roughly 30 meters above that. The effect of abyssal storms on the amount of suspended sediment in bottom waters is depicted at the right. Typical conditions are shown in black, storm conditions in color.

© 1984 SCIENTIFIC AMERICAN, INC

ces to be drawn. The top 10 centimeters or so of most marine sediments at all water depths are mixed more or less continuously by the feeding and burrowing activities of the organisms that live in them. Bottom-dwelling organisms are also the source of most of the fine-scale bottom topography and of the biological "glues" (such as mucus) that affect both the amount of fluid shear applied to the bottom and the effectiveness of such shear in eroding sediments. Most of the animals that agitate the bottom feed by digesting bacterial and other organic components of sediments, as common earthworms do.

The abundance of such organisms generally decreases with the depth of the water and the distance from the shore, closely following the pattern of the sedimentation rate itself. The net accumulation of sediment typically decreases from a few centimeters per 1,000 years at water depths of 1,000 meters to a few millimeters per 1,000 years in the open ocean. By judicious choice of radioactive isotopes with appropriate half-lives one can get estimates of both the depth of the biologically mixed layer and the intensity of biological mixing. Two isotopes with different half-lives will not as a rule yield identically shaped profiles under steady biological mixing.

On the time scales of the radioactive isotopes employed in stratigraphy, physical erosion and redeposition are nearly instantaneous. A redeposited layer should therefore show uniform concentrations of all the isotopes over its entire thickness. It is in principle thereby possible to estimate the depth of a deposit affected by abyssal storms and, by comparing the profiles of two or more isotopes, to distinguish pulses of physical mixing from the steadier biological mixing.



RAPID CHANGES in temperature (*black*) and light transmission (*color*) were recorded within the space of an hour at the HEBBLE site under stormy conditions. Water transparency of course varies with the concentration of suspended solids and thus indicates changes in the amount of sediment stirred up by the current. Profiles were recorded at 20-minute intervals.

Such interpretations are complicated, however, by the fact that physical redeposition often leaves graded layers of sediment, with the coarser and faster-settling particles having less surface area and therefore a lower concentration of adsorbed isotopes. Because it is also difficult to get a sample immediately after a storm, biological mixing of the storm deposit will usually have modified the storm trace considerably. One can infer mixing of the sediment from the appearance of fecal mounds, feeding traces and crawling tracks left by the bottom organisms. Because such activities rapidly destroy flow-produced graded beds, one can conclude that a bed form (for example a ripple) seen in a deep-sea photograph is not an old feature but one that has recently been created.

At deep-sea depths observations of any kind are difficult to make. The tools used to observe the intense abyssal storms must be not only strong enough to withstand the pressure of some 7,000 pounds per square inch at a depth of 5,000 meters but also rugged enough to withstand the substantial drag forces exerted by the bottom currents. In order to characterize the abyssal storms and their consequences investigators in the HEBBLE program have designed sturdy instruments capable of measuring the following variables: the velocity and the stresses in the fluid that act to pick up the sediment from the bottom, the concentration of sediment in the water and the temperature and salinity of the water. The last two values enable us to calculate the density of a particular mass of water and to deduce its origin. Photographs made at intervals reveal how the microtopography of the sea floor has been altered by the passage of a storm. The effect of storms on the biological and chemical characteristics of the sediments are determined by the analysis of material core samples.

For measuring current velocities we rely on rotor meters similar to but much ruggeder than the wind anemometers seen at airports. They record internally both the velocity and the direction of the current averaged over selected time intervals. Usually six or seven current meters are spaced at various levels on a vertical wire moored to the bottom and held taut by buoys. The entire assembly can be recovered by actuating an acoustically controlled release mechanism. The stresses in the fluid are obtained by measuring changes in the time required for a pulse of sound to travel a known distance. In the system developed by Albert J. Williams III of the Woods Hole Oceanographic Institution the timing of sound pulses establishes all three components of velocity five times per second, making it possible to resolve the velocity fluctuations that represent turbulence in the flow.

Temperature and salinity measurements near the deep-sea bottom are difficult to make because the variations in those properties are very small, typically less than 50 millidegrees Celsius in temperature and .1 part per 1,000 in salinity. Temperature is measured with carefully calibrated thermistors and salinity with sensitive probes of the water's electrical conductivity. The concentration of suspended sediment can be measured directly by taking water samples in bottles. For routine meas-

© 1984 SCIENTIFIC AMERICAN, INC

urements extending over days or weeks, however, we rely on acoustic and optical methods. The backscatter from high-frequency sound waves can resolve patterns of clouds of sediment. Measurements of light attenuation yield an accurate indication of the concentrations of sediment in the light path.

Detailed images of the sea floor are provided by a high-resolution stereo-camera system. Larger areas can be mapped by towing a side-scan sonar system 50 meters above the bottom. Other instruments used to examine the flow field and the sea bed include a laser Doppler velocimeter developed at Woods Hole by Yogesh C. Agrawal. It measures the velocity in the bottom 50 centimeters above the sea bed in a continuous profile by scanning optically with a laser beam and measuring the frequency shift in the light reflected back from particles moving through the beam.

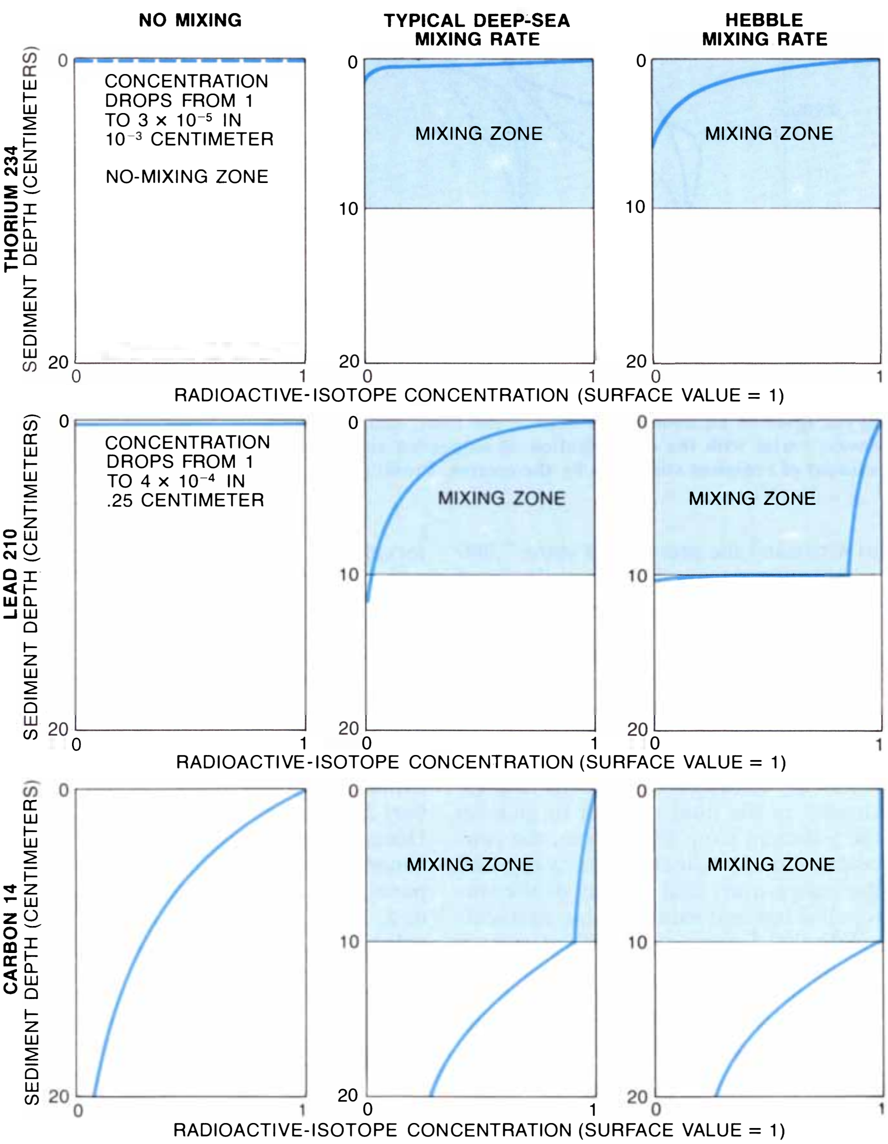
With the aid of such instruments we began studying deep-sea currents at the HEBBLE site in the North Atlantic in 1979. During abyssal storms the velocity of the bottom currents can increase from a tenth of a knot to one knot or more. Simultaneously the concentration of suspended sediment increases from 10 to 100 times. The elevated level of suspended material shows variable patterns in a storm, reminiscent of the variable gray patches seen in an active rain cloud. The moving cloud of suspended sediment exhibits apparently coherent patches with a lifetime of 20 minutes, corresponding to a linear dimension of 500 meters. This is about 10 times the thickness of the mixed layer at the bottom during a storm. On the basis of laboratory experiments a length scale of 500 meters is about what one would predict for large-scale turbulence in the boundary layer. During a storm the level of local mixing close to the sea bed rises steeply, accompanied by an increase in the turbulent energy throughout the bottom boundary layer, maintaining sediment in suspension.

Storms we have monitored usually last for from a few days to two weeks, at the end of which time the current velocity drops rapidly to about five centimeters per second (about .1 knot), with an accompanying drop in the level of suspended sediment. The rate of decrease in the concentration of suspended sediment implies that all but the finest particles settle out faster than might be expected from their individual settling velocities, apparently aided by the coagulation of particles in the cloud. The onset of abyssal storms is about as unpredictable as the arrival of storms in the atmosphere. It is surprising, in fact, that the circulation of the deep ocean, unlike the circulation of the atmosphere, does not show a strong seasonal pattern. At present we can only say that the abyssal storms are likely to come every two to three months.

The turbulence the storms generate in the boundary layer dissipates energy at a high rate. The rate at which mechanical energy is dissipated into heat reaches values of a hundredth of a watt per square meter of sea bed, values that are surpassed only in shallow estuaries and on the continental shelf when strong waves and currents interact near the bottom. Where does all the energy come from?

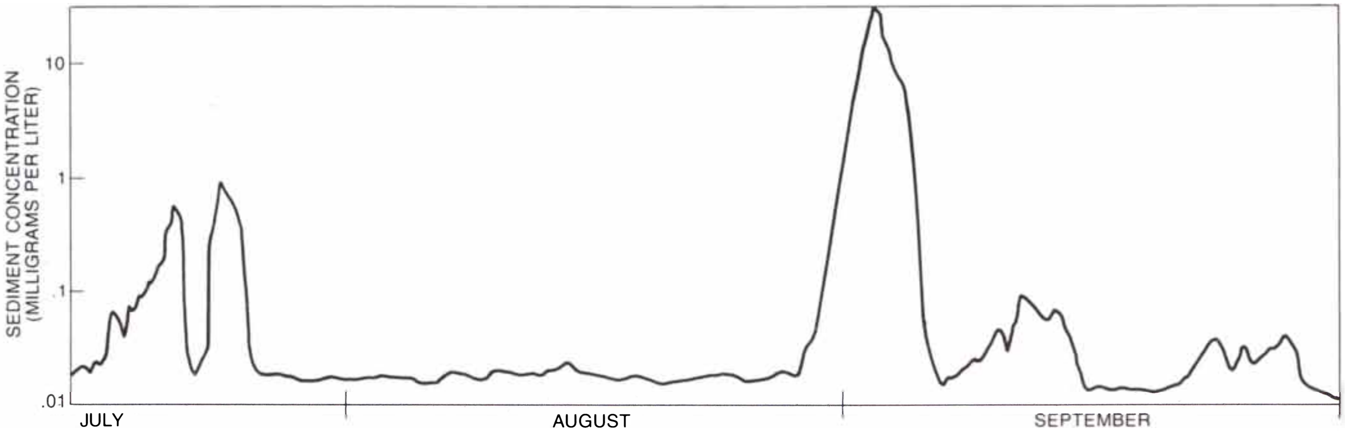
There are two sources: one is the nearly constant force associated with the thermohaline circulation and the other appears to be coupled to the eddies that are shed from the Gulf Stream at the surface. Georges L. Weatherly of Florida State University has recently gathered evidence that shows a strong correlation between the presence of warm-core eddies shed from the Gulf Stream and the presence of storms three miles down at the HEBBLE site. If these results can be generalized to other areas, it should be possible to use maps showing the distribution of surface eddies, which can be inferred from satellite measurements of variations in the height of the ocean surface, to identify other regions of the sea floor that are likely to be subject to active storm-driven erosion.

In the near future, in collaboration with Nelson G. Hogg of Woods Hole, we hope to deploy an array of current meters over a sufficiently large area to estimate directly the frequencies

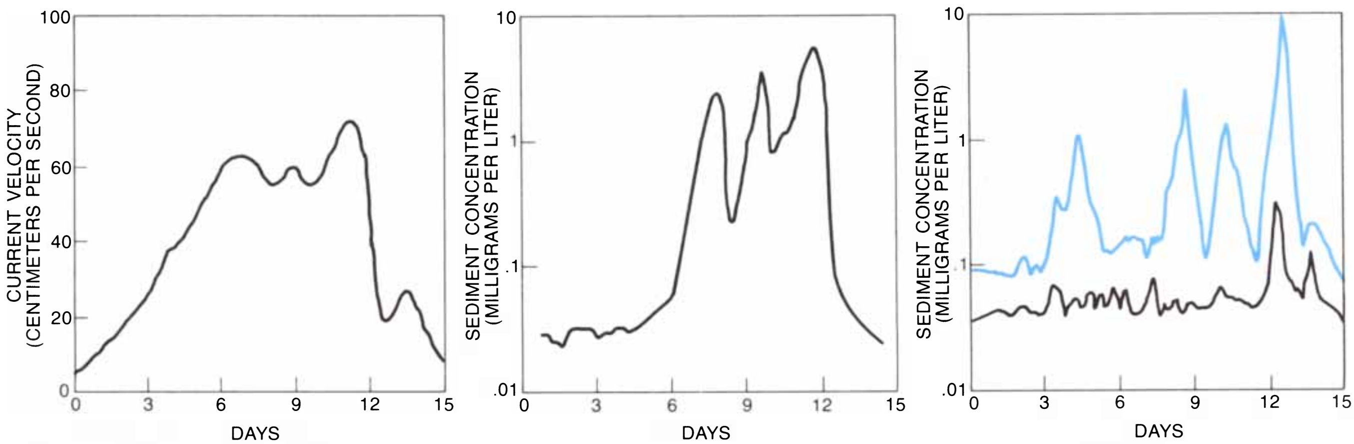


DISTRIBUTION OF RADIOACTIVE ISOTOPES in bottom sediments indicates the amount of mixing effected by currents and organisms. The nine curves show how three isotopes with extremely different half-lives are distributed with depth when there is no mixing (*left*), when the isotopes are subjected to a mixing rate typical of the deep sea (*middle*) and when they are subjected to the hundredfold higher mixing rate observed at the HEBBLE site (*right*). For thorium 234 the half-life is 24.1 days, for lead 210 it is 21 years and for carbon 14 it is 5,730 years. The curves assume that sediment accumulates at a rate of one centimeter per 1,000 years.

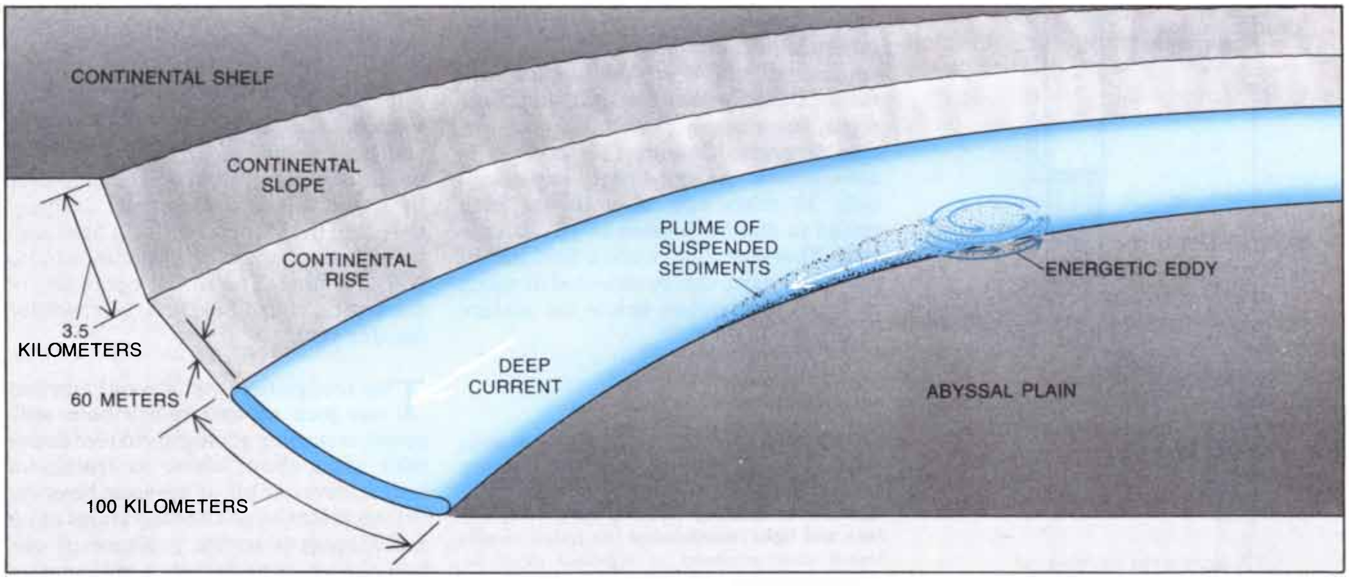
© 1984 SCIENTIFIC AMERICAN, INC



SEDIMENT CONCENTRATION varied by a factor of nearly 1,000 one meter above the ocean floor over a 10-week period at the HEBBLE site. Peak concentrations coincided with severe storms at the base of the continental rise in late July and early September.

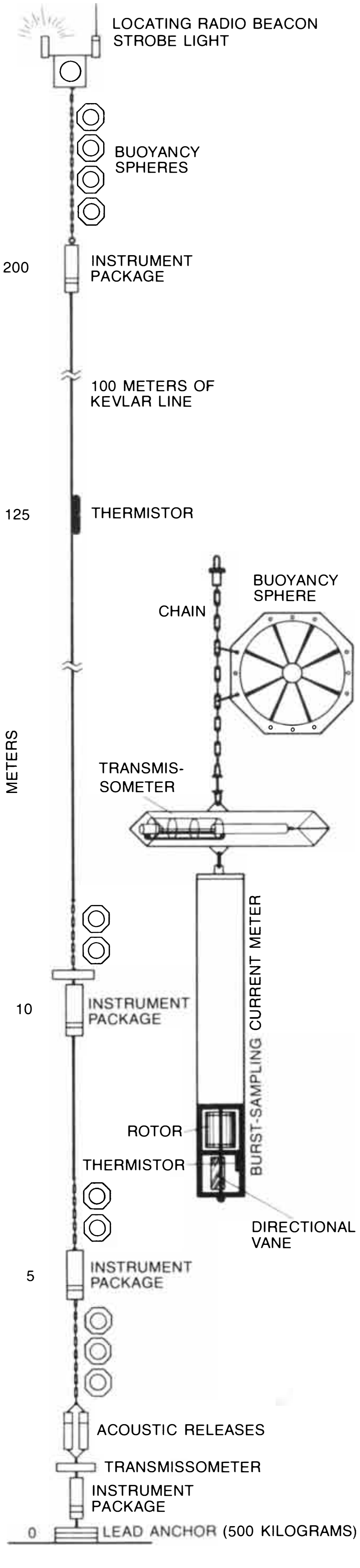


DURING A STORM lasting for about a week the velocity of the bottom flow (*left*) increased nearly tenfold, briefly exceeding 60 centimeters per second (1.17 knots). Simultaneously the concentration of sediment in the flow reaches a maximum of 100 times the background level (*middle*). Measurements over a short time scale (*right*) show how the concentration of sediment fluctuates rapidly two meters above the sea bed (*black*) and more widely half a meter above it (*color*) as sediment is picked up and carried past the transmissometers.



MESOSCALE EDDY is depicted schematically as it interacts with the southward-flowing abyssal current at the base of the continental rise 450 miles off the New England coast. The current is pressed against the rise by the Coriolis force. Such eddies, or storms, appear to be correlated with the presence at the sea surface of mesoscale eddies shed by the Gulf Stream. The abyssal eddy takes the form of an ellipse about 30 kilometers long and perhaps five kilometers wide; its height has not been established. The coupling of the eddy and the mean abyssal current scours the bottom, entraining mud that is swept downstream and is subsequently redeposited in sediment drifts.

© 1984 SCIENTIFIC AMERICAN, INC



and spatial dimensions of the eddies actually impinging on the bottom. For the present some limited turbidity studies by J. Ronald V. Zaneveld and Hasong J. Pak of Oregon State University suggest that the turbid patches of water stirred up by abyssal storms extend for a distance of perhaps 30 kilometers, a reasonable scale for ocean eddies. There are no corresponding estimates of the crosscurrent dimension of the storms. It is likely, however, to be less than 30 kilometers, because eddies tend to be somewhat elongated in the downstream dimension.

Although the passage of turbid water is clearly evident from turbidimeter measurements, one cannot say for any given site whether in the short term there is a net erosion of material, a net deposition or a negligible change. For example, substantial short-term erosion might be exactly balanced by deposition. The densest suspensions of sediments yet seen in the HEBBLE experiment reached 12 milligrams of solids per liter, or about 250 times the typical deep-ocean value. If one assumes that the turbid layer was 60 meters thick, the particles in suspension would have yielded a sedimentary layer half a centimeter thick if they were suddenly and uniformly deposited. Looked at another way, if one estimates the turbid stream to have been five kilometers wide, the amount of sediment transported through a cross section of the storm path per minute would have exceeded six metric tons.

If the storm had actually dumped a fresh layer of sediment half a centimeter thick at the HEBBLE site, it would have equaled the net amount normally deposited in 500 years, according to studies with long-lived radioactive isotopes. Such a sedimentation rate is in line with other estimates for deposition along the continental rise. When the HEBBLE sediments are examined for short-lived isotopes, however, a rather different picture emerges. Cesium 137, created in atmospheric atomic-bomb explosions only 30 years ago, already has been mixed to depths as great as 12 centimeters. Thorium 234, with a half-life of only 24.1 days, can be detected as much as seven centimeters below the surface of the sea floor.

MOORING SYSTEM for studying bottom currents in the HEBBLE program typically supports an array of seven instrument packages for measuring flow velocity, temperature and light transmission (an index of sediment concentration) at different distances from the bottom: less than one meter and at five, 10, 30, 60, 125 and 200 meters. The moorings are usually about 210 meters high and are spaced from five to 100 kilometers apart. They can be recovered by triggering an acoustic mechanism near the base of the mooring.

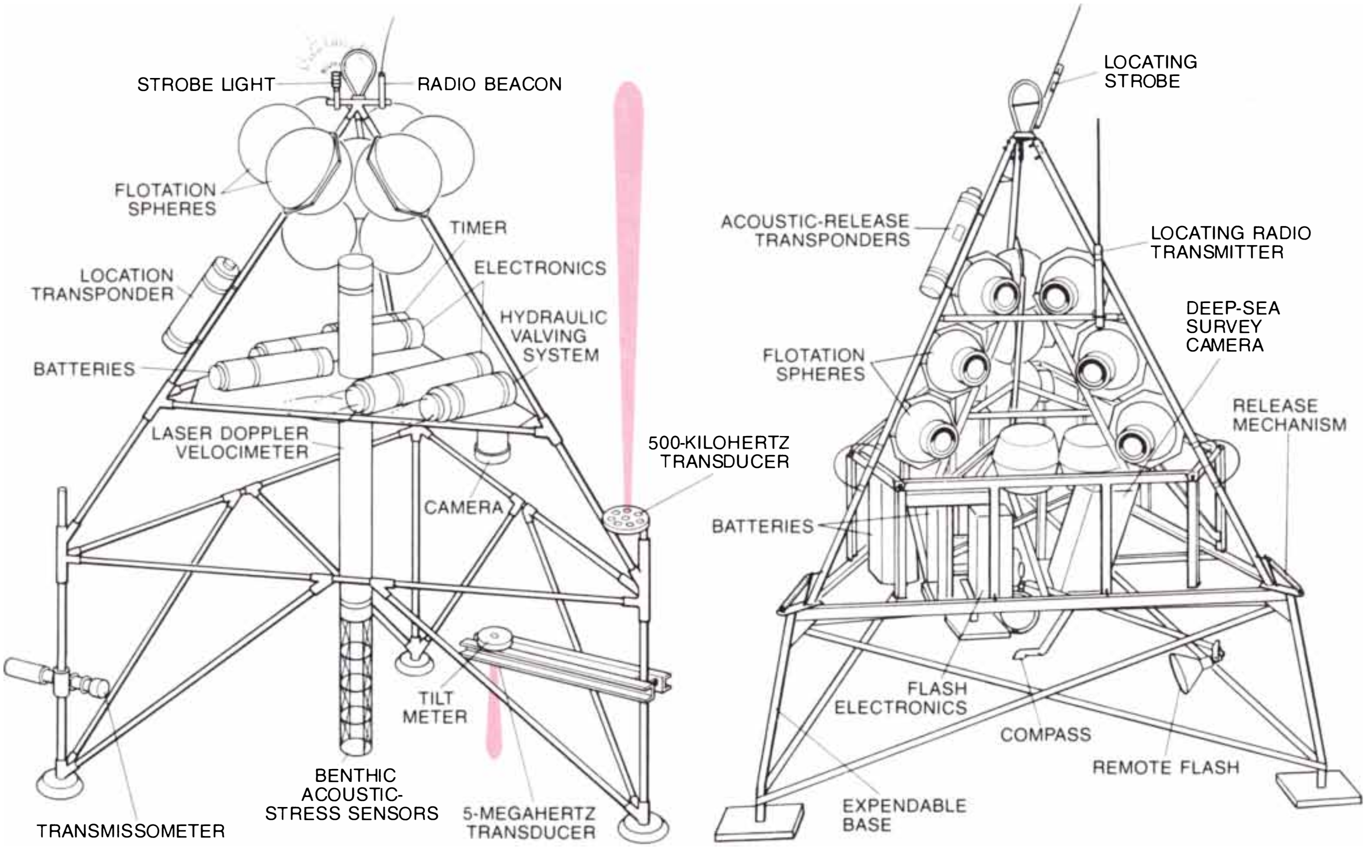
These studies, done by David J. DeMaster and Charles A. Nittrouer of North Carolina State University, suggest that the rates of mixing are about 100 times higher at the HEBBLE site than they are in typical abyssal areas free from storms. Evidently abyssal storms and bottom-living organisms act together to intensively mix the gradually accreted sediments. The relative contribution of the two forces remains to be worked out, but a preliminary estimate is that the storms account for more mixing than the organisms do. The organisms seem to be fairly sedentary, well adapted to collecting nutrients from the passing flow rather than to burrowing in search of food.

So far the HEBBLE program has been primarily observational rather than experimental. The provisional picture of the relations among flow strength, erosion and deposition must be tested by experiment before it can be held with any confidence or replaced with a more accurate one. At present erosional and depositional characteristics are known with confidence only for shallow-water sands. The response of deep-water mud beds to water flowing over them is sensitive to both the imposed stresses and the effects of organisms on their microtopography and adhesiveness. Since the phenomenon is not amenable to laboratory simulation, a field experiment is necessary. Only by in situ experimentation can one establish the complex interplay among flow strength, erosion rate, erosion depth, deposition rate and radioactive-isotope distribution.

To that end one of our colleagues at Woods Hole, Clifford L. Winget, has designed a special instrument that, lowered to the sea floor, can impose known and controlled stresses on bottom sediments. The device will circulate a sample of bottom water at various constant rates across an exposed but confined section of sea floor. The fluid stresses and the number of particles picked up by the circulating fluid will be measured by a special laser anemometer designed to record the Doppler shifts in light scattered from suspended particles moving with the flow. The initial operation of the device, called SeaDuct, is scheduled for this year.

The recognition that abyssal currents can pick up and redistribute sediments is causing geologists to reexamine their ideas about where to search for new reserves of oil. It has long been understood that large concentrations of oil are dependent on the presence of certain geologic conditions: a sedimentary source of the oil, a porous rock to serve as a reservoir and some kind of dam, or confining structure. Organic carbon in fine-grained, carbon-rich sediments generally provides the source. Porous and

© 1984 SCIENTIFIC AMERICAN, INC



TWO KINDS OF INSTRUMENT TRIPOD have been developed by the HEBBLE investigators for detailed study of bottom currents and sea-floor topography. The tripod depicted at the left carries three measuring systems. Acoustic-stress sensors determine the temporal and spatial variation in velocity and turbulent stress in the bottom current by measuring changes in the time required for a pulse of sound to travel a known distance. A laser Doppler velocimeter also determines flow velocities by measuring Doppler shifts in the light reflected from moving particles. A third system measures the concentration of suspended sediments by recording the backscattering of sound waves emitted at two frequencies: five megahertz and 500 kilohertz. The second type of tripod (*right*) carries a high-resolution stereo-camera system programmed to take pictures every 30 minutes. Both tripods are recalled to the surface by acoustic command.

permeable sedimentary rock, such as sandstone, often provides the reservoir. Diverse geologic structures can serve as dams to pool the oil.

According to the conventional view, the main force acting on sediments is gravity. It has long been assumed that all sediment moves inexorably downhill. Although this view probably remains true for coarse sand and gravel, it now appears that finer-grained material can be transported significant distances by abyssal currents. Presumably most of the organic matter capable of forming oil is concentrated in fine particulates that originate primarily in surface waters. When such material settled to the floor on the western side of ancient ocean basins, it could well have been transported parallel to the basin contours, or at right angles to the direction commonly assumed.

In order for organic material to be transformed into oil at least one of two conditions is thought to be necessary. Either the rate of accumulation must be so high or the oxygen in the bottom water must be so low that the material is not oxidized before it is buried. If petroleum geologists could learn where either, or preferably both, of these conditions prevailed in ancient seas, they might greatly facilitate their finding new deposits of oil. Hence areas where abyssal storms induced high rates of accumulation of sediments rich in organic material should provide favorable sites for the formation of oil-bearing rock. Accordingly the geologist might look down paleoslopes for coarse-grained reservoir rock and down paleocurrents for drifts bearing sediments rich in carbon.

The new knowledge of abyssal storms has immediate practical implications for engineers who are responsible for the stability and security of cables and structures emplaced on the sea floor. Underwater photographs have shown that currents are capable of washing sediments out from under submarine cables and leaving them suspended between uneroded areas of the sea bed. In a few instances the subsequent strumming of the cable in the current has caused the cable to fail.

The existence of abyssal storms also presents new problems for those concerned with antisubmarine strategies. The noise associated with large turbulent storms can confuse the passive listening arrays placed on the sea floor to detect the passage of submarines. Conceivably it might be possible to make submarines sound like the storms themselves, rendering the boats extremely difficult to detect in a high-energy region such as the western North Atlantic.

Perhaps the most important question raised by the existence of abyssal storms is their potential effect on toxic wastes that may be consigned to the sea bottom. Before the ocean floor is selected as a final dumping ground for any hazardous waste it is essential to have reliable predictions about long-term conditions at the sea-floor site chosen. If the waste-disposal method depends on containment, the disposal site should be a region where there is negligible transport of sediment. On the other hand, if the disposal strategy is to dilute and disperse the waste material, it might be better to seek out storm-prone areas. In any event one should be aware that some toxic wastes are readily adsorbed onto fine suspended particles and might thereby gain ready entry into the food web of bottom-feeding organisms. These are just a few of the issues raised by our study of abyssal currents and storms in one corner of the world ocean.

© 1984 SCIENTIFIC AMERICAN, INC

How Genes Control an Innate Behavior

The techniques of recombinant DNA are exploited to define a family of genes encoding a set of related neuropeptides whose coordinated release governs a fixed-action pattern: egg laying in a marine snail

by Richard H. Scheller and Richard Axel

Certain stereotyped patterns of animal behavior are innate. They are shaped by evolution and are inherited by successive generations; largely unmodified by experience or learning, they are displayed by all individuals of a species and not by other species. Ethologists have described such innate, stereotyped behavioral arrays as "fixed-action patterns": patterns of behavior consisting of several independent elements that either together form a coordinated sequence or do not take place at all. Each animal inherits a unique collection of such fixed-action patterns, which characterizes the behavior of its species.

How does an animal inherit such a behavioral repertoire? What does an animal inherit? It inherits DNA. The genes of the DNA can specify stereotyped behavior in two ways. First of all, they can specify a precise network of interconnected nerve and muscle cells that are put in place and "wired" together in the course of the animal's development. A stereotyped behavior is elicited, however, only in particular situations or at particular stages of an animal's life cycle, and only by the coordinated activity of particular parts of the network of neurons and muscle cells. In addition to the network, then, the genes must specify control elements: substances that excite specific preexisting connections in a rigidly determined way to generate a fixed-action pattern at the right time.

As molecular biologists studying the nervous system we seek to identify such control elements and the genes governing their synthesis. In so doing we hope eventually to learn something about the factors that play a role in the generation of innate behavioral repertoires, how they evolve in different species and how they develop in a given species.

Specific behaviors, however, unlike such traits as eye color or the inherited disorders of hemoglobin, are not likely to directly reflect the state of specific genes. The central nervous system integrates and filters the dictates of genes in ways that for the most part are inaccessible to experiment. The more complex the central nervous system, the more elusive the relation between a set of genes (a genotype) and observable traits (a phenotype). It is easier to study genes specifying behavior in an organism that is sophisticated enough to exhibit interesting behavioral repertoires but simple enough so that the behavior can be attributed to identifiable cells. We work with such an animal: the mollusk *Aplysia,* a shell-less marine snail that can weigh as much as five to 10 pounds.

The snail's central nervous system is numerically simple, consisting of about 20,000 nerve cells collected into four pairs of head ganglia and one abdominal ganglion. This is in striking contrast to the brain of a mammal, which has perhaps a million times as many neurons. Moreover, the neurons of *Aplysia* can be as much as a millimeter in diameter, more than 1,000 times the size of a typical human brain cell. Most of these huge cells contain correspondingly large quantities of DNA: as much as a microgram, or several hundred thousand times the DNA content of a typical mammalian neuron. Functions that would be carried out by a large collection of related neurons in a more complex nervous system may be handled in *Aplysia* by a single large cell. A series of investigations, notably the elegant work of Eric R. Kandel and his colleagues at the Columbia University College of Physicians and Surgeons, has related specific patterns of behavior in the snail to the functioning of particular cells. We and our colleagues have gone on to examine the activity of specific genes in a single cell and have attributed a behavior pattern to the activity of individual genes.

An adult *Aplysia* is largely occupied with feeding and reproduction. Aspects of the snail's reproductive behavior are highly ritualistic, involving a coordinated series of stereotyped patterns that accomplish courtship, mating and the deposition of fertilized eggs. *Aplysia* is a true hermaphrodite, an individual serving as both male and female, in most cases simultaneously, with the snails copulating in long chains of half a dozen animals or more; sometimes the chain is closed to form a circle. Fertilization takes place internally, in the reproductive duct, and then the fertilized eggs are laid to develop externally, in the sea. The genes we study are those implicated in the elaborate but completely stereotyped array of coordinated behaviors that accomplish the egg-laying process.

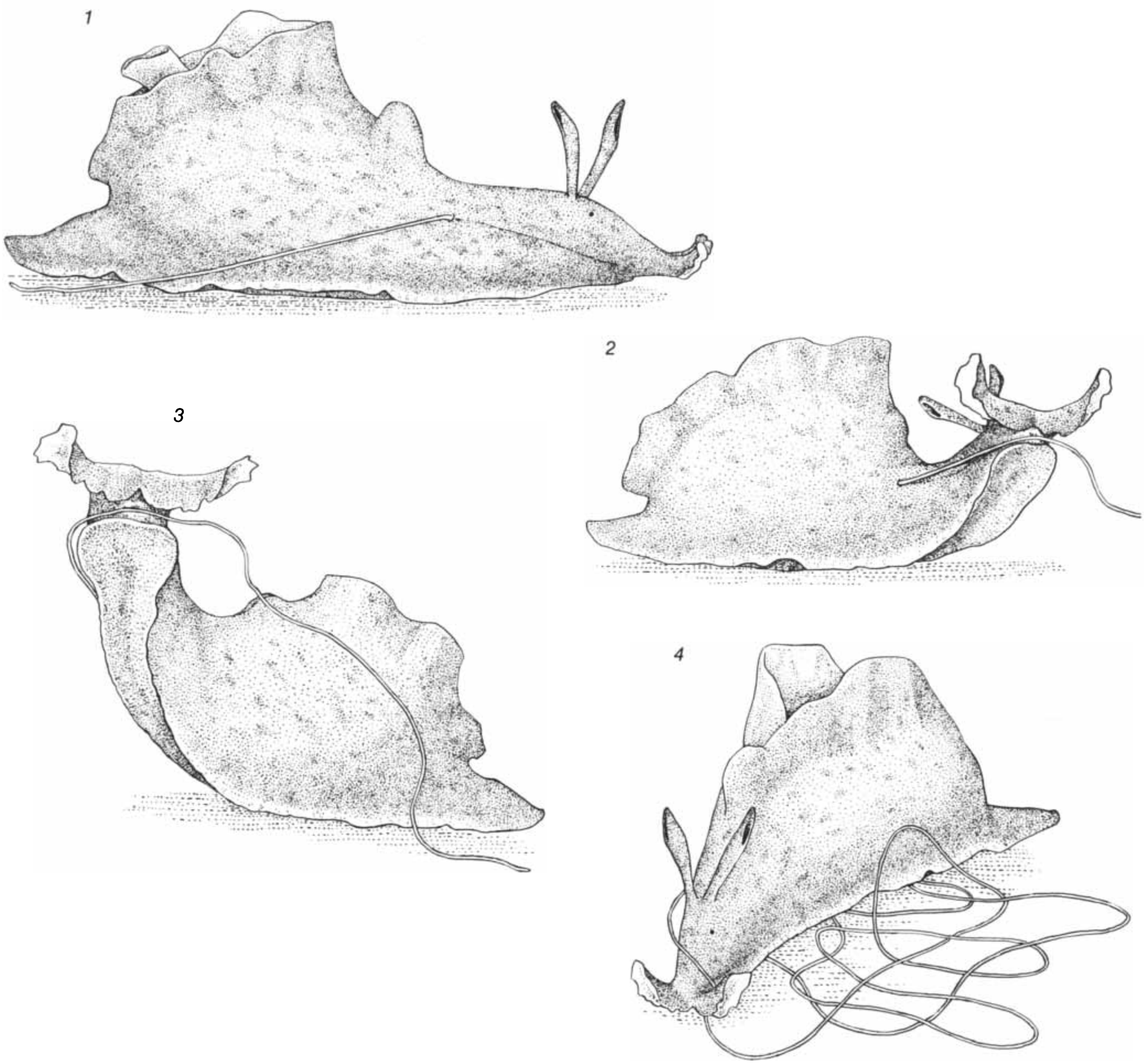
The eggs are laid in long strings of more than a million eggs. As the string of eggs is expelled by contraction of the reproductive-duct muscles, the snail stops walking. It stops eating. Its heart rate and respiratory rate increase. The snail grasps the egg string in its mouth. With a series of characteristic head-waving movements it helps to extract the string from the duct and winds the string into an irregular mass. A small gland in the mouth secretes a sticky mucoid substance that becomes attached to the mass. Then, with one forceful wave of the head, the animal affixes the entire mass of eggs to a solid support such as a rock. A number of disparate actions have come together in a rigidly coordinated sequence to serve a common function: the deposition of fertilized eggs in a way that will afford protection during their development.

What substances control this coordinated pattern of behavior and what genes encode these substances? First, where are the substances synthesized? Several years ago Kandel and Irving Kupfermann identified two clusters of neurons, the bag cells, at the top of the abdominal ganglion. When an extract of these cells was injected into live snails, it elicited the full repertoire of behaviors associated with egg laying even though

© 1984 SCIENTIFIC AMERICAN, INC

the animals had not mated and the eggs were unfertilized. In subsequent studies Stephen W. Arch of Reed College and Felix Strumwasser of the California Institute of Technology isolated one of the active bag-cell factors and identified it as a peptide, or small protein, consisting of 36 amino acids. They found that when the peptide was administered to snails, it elicited some but not all of the egg-laying behaviors, suggesting it was one of several factors controlling the total behavioral repertoire. Strumwasser determined the linear sequence of the amino acids in the peptide, which was designated egg-laying hormone (ELH). Because there is a linear relation between the amino acid sequence of a protein and the nucleotide sequence of the gene encoding it, the identification and sequencing of a peptide controlling behavior put a problem in behavioral biology well within the realm of molecular genetics.

In collaboration with Linda B. McAllister, James F. Jackson, James H. Schwartz and Kandel we set out to isolate the gene encoding ELH from the *Aplysia* genome: the total complement of DNA in the snail's chromosomes. The procedure for isolating specific genes depends on the techniques of recombinant DNA. In brief, one establishes a "library" of recombinant-DNA molecules, each of them carrying a small fragment of the *Aplysia* genome, and then scans the library with a probe that will detect the ELH gene. We cleaved the snail DNA into many thousands of fragments, each fragment calculated to include one gene or perhaps a few. We "recombined" the fragments of *Aplysia* DNA with the DNA of a bacterial virus, phage lambda, and packaged the recombinant DNA in the phage's protein coat. The hybrid phages served as vectors for introducing the small DNA into bacteria and cloning it. When the phages are added to a bacterial culture, each phage infects a single bacterial cell and multiplies, killing the cell; the phage progeny keep multiplying, killing adjacent cells and creating a plaque, or



EGG-LAYING BEHAVIOR is exhibited by *Aplysia*, a large marine snail, in a laboratory aquarium. Contraction of the muscles of the snail's reproductive duct expels a string of egg cases (*1*). The animal grasps the egg string in its mouth (*2*) and waves its head (*3*), thus helping to draw the string out of the duct, and eventually it affixes a tangle of string to a solid substrate (*4*). In the instance depicted in these drawings the behavior was elicited in an unmated snail by injecting the animal with an extract of *Aplysia* bag cells: neurons in which a "polyprotein" is synthesized from which an egg-laying hormone (ELH) and other peptides associated with egg-laying behavior are cleaved.

© 1984 SCIENTIFIC AMERICAN, INC

hole, in the culture. Each plaque is occupied by millions of phages descended from an individual phage and therefore contains a clone of millions of copies of a single *Aplysia* DNA fragment.

In selecting a probe with which to scan the clones and find the ELH gene we relied on the fact that ELH is synthesized in abundance in the bag cells, which must therefore contain a large amount of ELH messenger RNA. This calls for some explanation. DNA is a double-strand molecule each of whose strands is a chain of four different nucleotide subunits. Genetic information is encoded in the sequence of nucleotides defining a gene. Groups of three nucleotides are codons, or code words, each of which specifies a different one of the 20 amino acids that are the subunits of proteins. The DNA is not translated direct-

ABDOMINAL GANGLION of *Aplysia* is enlarged about 40 diameters in this photomicrograph. The two clusters of bag cells, where ELH is synthesized, are visible lying athwart the large nerve bundles at the top left and top right. Within the main body of the ganglion one can discern a number of very large neurons, or nerve cells, many of which have been individually identified and found to be invariant in all members of the species. ELH and its companion peptides have been shown to have specific effects on the firing of certain neurons.

© 1984 SCIENTIFIC AMERICAN, INC

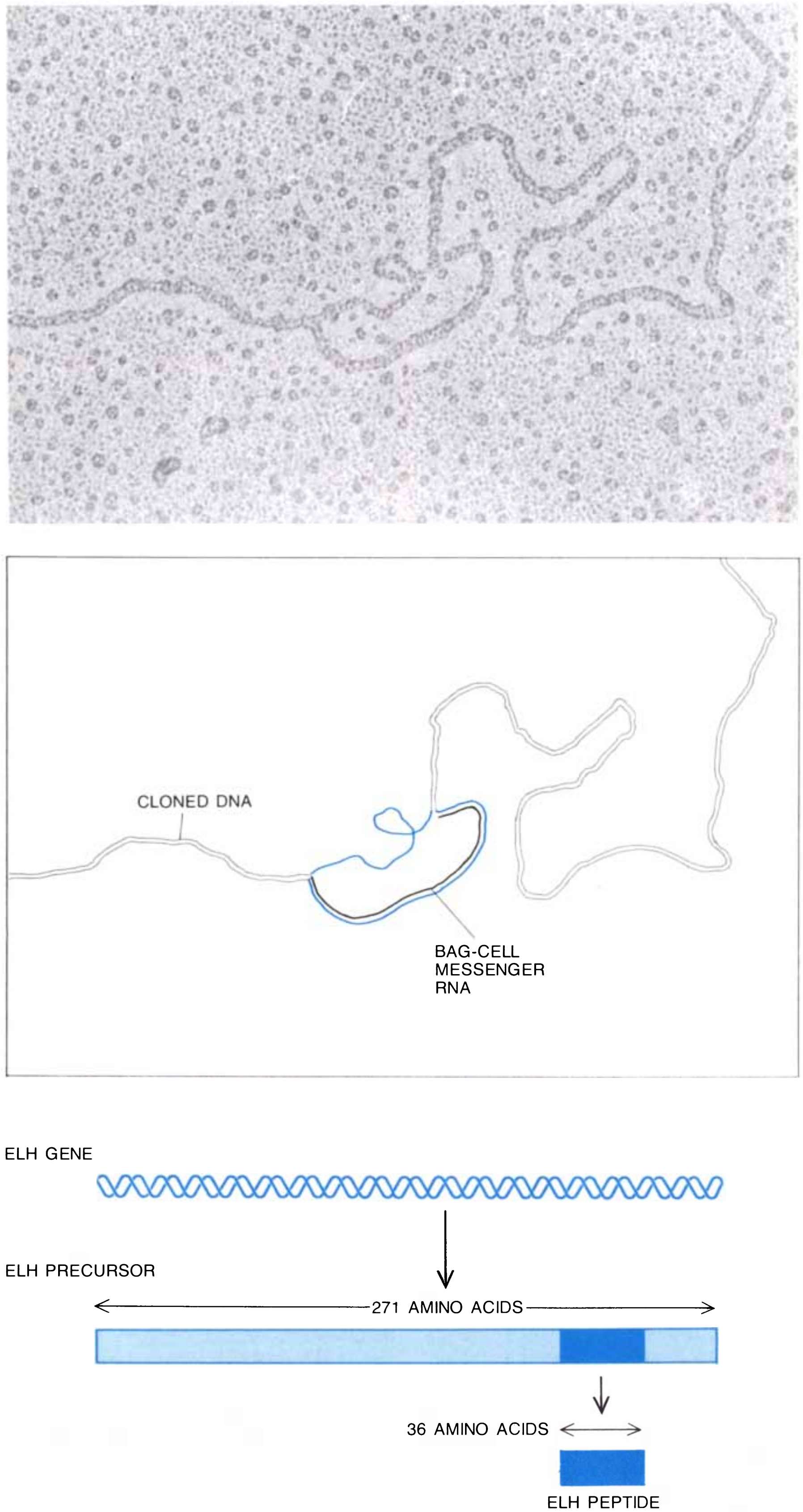
ly into protein, however. The coding strand of DNA is first transcribed into a complementary strand of the similar nucleic acid messenger RNA whose sequence reflects that of the gene, and the RNA is translated into protein according to the genetic code.

Since ELH messenger RNA is predominant in the bag cells but not in non-neural tissues, we isolated messenger RNA from bag cells and from non-neural cells. We "reverse-transcribed" it into complementary DNA and labeled the DNA with a radioactive isotope. The labeled DNA was then exposed to the library of recombinant clones under conditions such that the ELH complementary DNA would "hybridize" with, or bind to, stretches of cloned genes encoding ELH; the non-neural complementary DNA would not do so. Clones with which the radioactive ELH probe hybridized but with which the non-neural probe did not hybridize were revealed by autoradiography.

In this way we identified a number of clones containing ELH genes. (As we shall explain, there turned out to be three somewhat different genes, but they are so similar that the bag-cell probe hybridized with all of them.) Having identified clones containing the ELH gene, we grew some of them up into large cultures, isolated their recombinant DNA and exposed it to ELH messenger RNA. In electron micrographs we could detect the exact sites on the long recombinant-DNA molecule at which the RNA hybridized with the complementary DNA and so could close in on the precise stretch of DNA constituting the ELH gene.

Having isolated the ELH gene, we sequenced it, that is, we determined the order of its nucleotides. Given the nucleotide sequence, we could "reverse-translate" it according to the genetic code and thereby deduce the amino acid sequence of the protein chain encoded by the gene. The gene's protein product, it turned out, must have 271 amino acid subunits. Yet we knew that the ELH hormone itself has only 36 amino acids, and indeed we could see the small ELH sequence within the longer protein sequence. The ELH peptide is flanked on both sides by the same pair of amino acids: lysine and arginine. The lysine-arginine pair is known to be a cleavage signal. It serves as a site at which specific enzymes cut a large precursor protein chain to make an active smaller protein or peptide. The most obvious explanation, then, was that the 271-amino-acid chain is a precursor molecule from which ELH is cleaved.

It was surprising, however, that the ELH peptide accounted for such a small part of the ELH gene's protein product. Was the rest of the precursor



ELH GENE is demarcated in an electron micrograph (*top*). Fragments of snail DNA were recombined with the DNA of a bacterial virus and cloned. A clone containing the ELH gene was identified, expanded and then exposed to ELH messenger RNA under conditions promoting the formation of RNA-DNA hybrids. As is shown in the map, the RNA (*black*) hybridized with the complementary strand of the ELH gene (*color*), disrupting the cloned-DNA duplex. When the ELH gene was sequenced, it was seen to encode a precursor protein chain 271 amino acids long. The 36-amino-acid ELH peptide was identified within the precursor protein.

© 1984 SCIENTIFIC AMERICAN, INC

chain simply cast off unused or might it too contain biologically active peptides? We examined the sequence for additional cleavage signals. It turned out that there are 10 such signals in all; if each site were recognized and cut, 11 discrete peptides would be generated from the single ELH precursor protein. This raised the exciting possibility that such peptides might be other elements controlling egg-laying behavior.

Earl M. Mayeri and his colleagues at the University of California at San Francisco School of Medicine had been investigating the physiological properties of neuropeptides released by the bag cells, and so we joined them in an effort to see whether any of the peptides predicted from the ELH-gene sequence were present in bag-cell extracts and, if they were, whether they would prove to be active in the nervous system. Mayeri and Barry S. Rothman identified and sequenced three other small peptides in the bag-cell clusters: alpha bag-cell factor, beta bag-cell factor and acidic peptide. We found that each of them, bounded by cleavage sites, is encoded along with ELH in the gene we had isolated. Electrophysiological experiments showed that three of the four peptides (ELH and alpha and beta bag-cell factors) interact with specific identifiable neurons in the abdominal ganglion, where each acts as a neurotransmitter: a molecule that mediates the transfer of electrical activity from one neuron to another.

ELH acts locally as an excitatory transmitter, augmenting the firing of the abdominal-ganglion neuron designated *R*15. In addition to acting as a neurotransmitter the ELH peptide diffuses into the circulatory system and excites the smooth-muscle cells of the reproductive duct, causing them to contract and expel the egg string. In other words, ELH acts not only as a neurotransmitter but also as a hormone. (The peptide's effect on the duct muscle was noted before its nervous-system activity; hence the name egg-laying hormone.) The beta bag-cell factor is also an excitatory transmitter. It causes the firing of two symmetrical neurons, *L*1 and *R*1, whose functions are not known. The third neuropeptide, alpha bag-cell factor, is an inhibitory transmitter. It inhibits the firing of a cluster of four neurons, *L*2, *L*3, *L*4 and *L*6. In addition it appears to have a feedback-amplification effect: it is capable of exciting the bag cells from which it has been released.

The association of these three individual peptides, encoded by a single gene, with the activity of different sets of neurons (and muscle cells) suggests that an interesting mechanism may be responsible for generating the complex array of behaviors associated with egg laying. A single gene appears to specify a "polyprotein": a protein chain that is cut into a number of small, biologically active peptides. Perhaps all the components of the egg-laying behavior are mediated by peptides encoded by this one gene. If that is the case, there would be an all-or-nothing effect: no one component of the behavior would be displayed in the absence of the others. Moreover, the synthesis of the single polyprotein would rigidly coordinate the timing of a battery of related behaviors. A single gene encoding multiple neuroactive peptides might thus dictate a complex array of innate behaviors: a fixed-action pattern.

If indeed peptides derived from a single polyprotein mediate the egg-laying behaviors, what controls the release of the peptides to initiate the behavioral array at an appropriate time in the animal's life? The release of ELH and the other putative egg-laying peptides takes place after prolonged electrical excitation of the bag cells. Strumwasser's group has isolated and sequenced two peptides, designated *A* and *B*, that are synthesized in the atrial gland, an organ in the snail's reproductive system. Injection of the atrial-gland peptides into an animal results in the excitation of the bag-cell clusters, prompting the release of ELH and its companion bag-cell peptides. Whether the *A* and *B* peptides actually initiate the egg-laying process in nature has not yet been demonstrated, but in the laboratory they do appear to be factors controlling the release of the bag-cell peptides.

When we examined the amino acid sequence of the *A* and *B* peptides, we noticed that short blocks of amino acids in our ELH precursor were homologous to (the same as or almost the same as) stretches of the atrial-gland peptides. The reader will recall that in our cloning experiments the probe made from ELH messenger RNA had hybridized to three somewhat different DNA's, only one of which encoded ELH and its companion bag-cell peptides. It was possible that the three DNA's were members of a multigene family, with one of them encoding ELH and the other two encoding the *A* and *B* peptides.

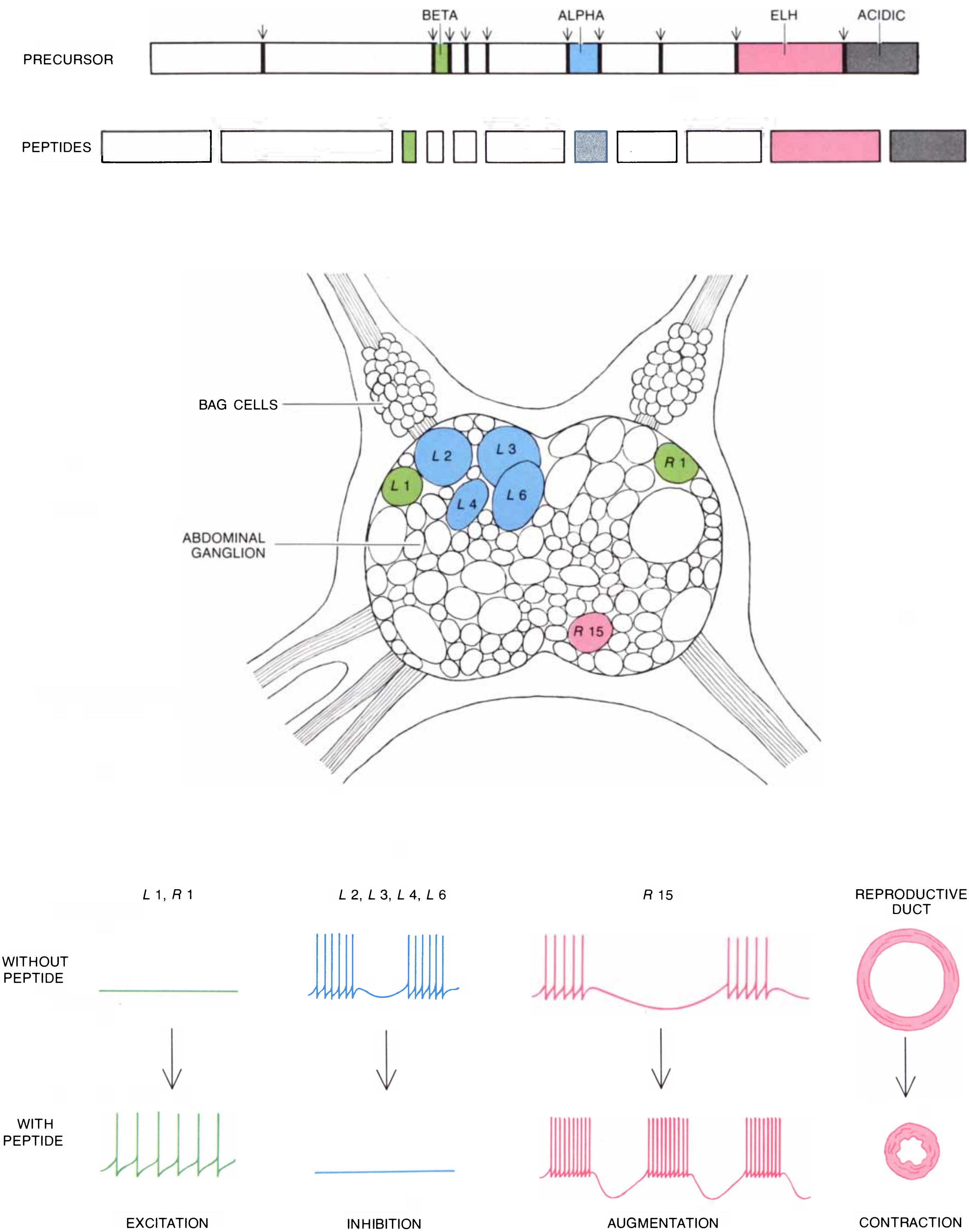
In a series of hybridization experiments we were able to show that messenger RNA derived from the two genes related to the ELH gene is indeed synthesized in the atrial gland, where Strumwasser had found the two peptides. We determined the nucleotide sequence of the two genes and, by reverse translation, the amino acid sequence each of them encodes. We saw immediately that although the two genes are closely related to the ELH gene, they do not encode the ELH peptide that is active in the bag cells. Instead they encode the *A* and *B* peptides of the atrial gland. The three genes are clearly members of a small multigene family with a common evolutionary origin, but they have diverged to generate different—yet functionally related—sets of peptides.

The similarities and the differences are revealed when the three nucleotide and amino acid sequences are compared in detail [*see illustrations on page 60*]. Each gene encodes a precursor protein in which lysine-arginine sites (or sometimes a single arginine or two adjacent arginines) delimit the blocks of amino acids that are cleaved to become active peptides. All three precursors begin with a characteristic signal sequence of about 25 amino acids that governs the processing of the protein chain. The newly translated chain enters the lumen of a membrane system called the rough endoplasmic reticulum, where it begins to be modified: the signal sequence is cut off and sugar and phosphate molecules are added to the protein, which proceeds to the organelle called the Golgi apparatus. There the precursor is cleaved and the component peptides are enclosed in small vesicles, or sacs. In response to appropriate stimuli the vesicles fuse with the outer membrane of the secreting cell and release their contents to interact with nearby cells, to diffuse through the ganglion or to enter the circulation.

The differences between the precursors incorporating the *A* and *B* peptides and the precursor incorporating ELH begin after the signal sequence. Let us first describe the *A* and *B* precursors. A single-arginine site signals the beginning of either the *A* or the *B* peptide, both of which consist of 34 amino acids. At the end of these peptides there is a glycine-lysine-arginine sequence, which serves as a signal not only for cleavage but also for transamidation: the addition of an amino group (NH_2) at the end of the peptide, replacing the usual hydroxyl group (OH). Transamidation "blocks" the end of the peptide, perhaps making it more resistant to degradation. There follows, in both the *A* and the *B* precursor, a stretch of 47 amino acids unrelated to any known peptide. Then comes another lysine-arginine cleavage site, followed by what looks like the beginning of the ELH peptide.

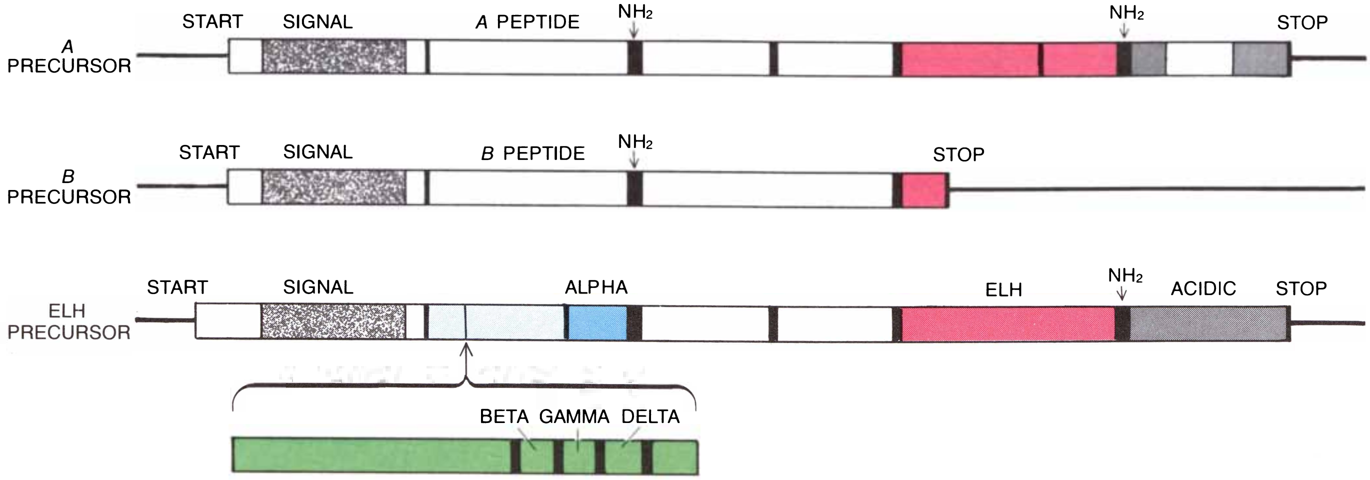
ELH is not synthesized in the atrial gland, however. Examination of the nucleotide sequence of the atrial-gland genes shows why. In the case of the *A*-peptide gene, the first 22 amino acids of ELH are encoded correctly. Then a single-nucleotide difference in the codon for the amino acid at position 23 generates an arginine-arginine-arginine sequence and thus establishes a potential cleavage site that could break up what would otherwise become the ELH peptide. A different kind of single-nucleotide change is seen in the *B*-peptide gene.

© 1984 SCIENTIFIC AMERICAN, INC

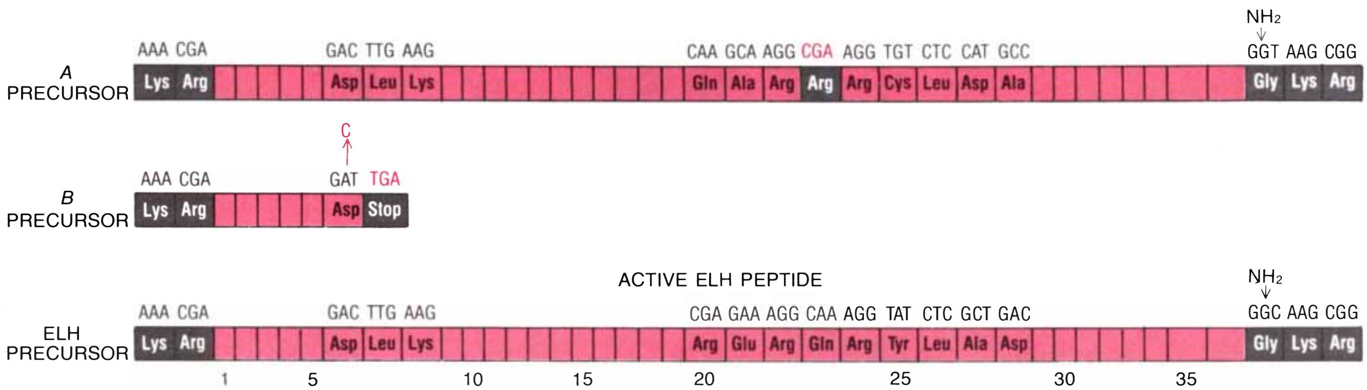


ELH PRECURSOR was found to be a polyprotein containing a number of active peptides. The precursor (*top*) is studded with 10 sites (*arrows*) at which a protein chain is cleaved by enzymes called endopeptidases. Cleavage at all the sites would release 11 peptides (*second from top*). Four of those peptides are known to be released by the bag cells: the beta and alpha bag-cell factors, ELH and acidic peptide. Three of them (*colored peptides*) have been shown to act as neurotransmitters, altering the activity of specific abdominal-ganglion neurons (*colored cells*) in specific ways (*bottom*). The beta factor excites cells *L*1 and *R*1. The alpha factor inhibits cells *L*2, *L*3, *L*4 and *L*6. ELH augments the firing of cell *R*15. ELH also enters the circulation and acts as a hormone, causing contraction of the reproductive duct.

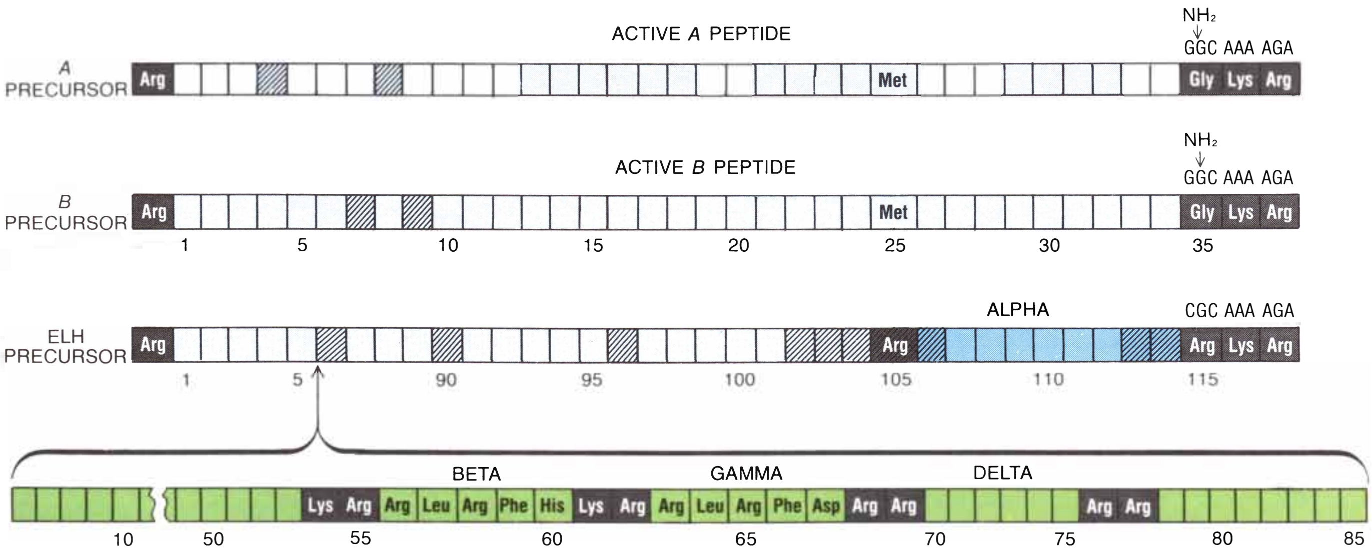
© 1984 SCIENTIFIC AMERICAN, INC



THREE PRECURSOR PROTEINS encoded by the three members of the ELH multigene family are compared. These precursors are cleaved at specific sites (*black bars*) to give rise to active peptides, some of which undergo amidation (NH_2). The three precursors have in common a signal sequence (*left*) and several regions of homology, or close similarity, that give rise in one precursor or another to the *A* or the *B* peptide or to ELH and acidic peptide. Single-nucleotide differences alter the *A* and *B* genes so that active ELH is not synthesized. An 80-amino-acid insertion in the ELH gene interrupts the *A*- and *B*-peptide sequences and gives rise to several different peptides.



ELH HOMOLOGY REGION is very similar in all three genes, that is, the sequence in which the four nucleotides (*A, G, C* and *T*) are arrayed is almost the same, so that the three-nucleotide codons specify the same amino acids at most positions. (Only some of the more significant codons and amino acids are shown here.) Whereas the ELH gene gives rise to an active ELH peptide, the other two genes do not. In the *A*-precursor gene the substitution of a *G* for an *A* nucleotide at position 23 gives rise to an arginine instead of a glutamine, generating a three-arginine sequence and thus a potential cleavage site. In the *B*-precursor gene a *C* nucleotide is deleted from the sixth codon, changing the "reading frame" so that a *TGA* is introduced at position 7. *TGA* serves as a "stop" codon, which terminates translation.



EIGHTY-AMINO-ACID INSERTION disrupts the *A*- and *B*-peptide homology region in the ELH gene. It is cleaved at lysine-arginine sites to make three peptides: the beta, gamma and delta bag-cell factors. (Note the near-identity of beta and gamma, which presumably arose by duplication of a short DNA sequence.) At position 105 in the homology region the substitution of an arginine in the ELH precursor for the methionine of the *A* and *B* precursors creates a cleavage site, giving rise to another active peptide: alpha bag-cell factor. Apart from the insertion the three precursors are very similar. The hatching marks sites where one of them has a different amino acid.

© 1984 SCIENTIFIC AMERICAN, INC

Here one nucleotide in the sixth codon is deleted. The "reading frame" of nucleotide triplets is thereby changed so that a "stop" codon is generated; translation is terminated after only a six-amino-acid stub of ELH has been synthesized.

Consider now the ELH precursor that is synthesized in the bag cells. The nucleotide sequence of its gene is very similar to that of the gene encoding the *A* and *B* peptides, and yet the gene does not specify those peptides; it specifies ELH and a number of other peptides involved in egg laying. It begins with the same signaling sequence seen in the *A* and *B* precursors. Then come the first five amino acids of the *B* peptide. At this point, however, the ELH precursor diverges dramatically from the *A* and *B* precursors. The ELH gene contains a 240-nucleotide sequence that is not present in the *A* or the *B* gene, encoding 80 amino acids. The insert includes four cleavage signals delimiting three bag-cell factors: beta, gamma and delta. The beta factor, as we have mentioned, is known to have a specific effect on abdominal-ganglion neurons *L*1 and *R*1.

After the insert the nucleotide sequence resumes, without any alteration of the reading frame, to encode the sixth amino acid of the *B* peptide, and then it continues through a sequence that is much like the sequence of the two atrial-gland genes. One divergence in this region is particularly significant. A cleavage site is introduced, generating the nine-amino-acid alpha bag-cell factor that, as described above, inhibits the firing of four neurons in the abdominal ganglion.

There follows a stretch of incomplete homology with the *A* and *B* genes, after which a cleavage site signals the beginning of the 36-amino-acid ELH peptide. The end of this peptide, as in the case of *A* and *B*, is followed by a signal for cleavage plus transamidation. Between the end of the ELH peptide and the stop codon that puts a halt to translation there remain 27 amino acids: those of acidic peptide, which is also released from the bag cells along with ELH but whose target is not yet known.

Here, then, are three genes that have in common three regions of homology: the *A* or *B* region, the ELH region and the acidic-peptide region. Moreover, within each of these three peptide regions there are near-identities of sequence at fixed positions. The implication is that all three peptides had their origin in a small ancestral peptide whose gene triplicated to generate a larger protein composed of at least three peptides. The gene encoding that larger protein apparently triplicated in turn, giving rise to three independent genes that diverged as they became specialized to satisfy different functional requirements. Then there may have been minor duplications of some regions of one or another gene, as is suggested by the fact that the adjacent beta and gamma peptides are almost identical. Such events presumably allowed for the evolution of variants without significant alteration of the original gene. The various versions of the gene may have been transposed to different sites in the genome, perhaps on different chromosomes.

The availability of cloned ELH genes enabled us, working with McAllister and Kandel, to ask whether the genes are expressed (transcribed into messenger RNA that is then translated into protein) not only in the bag cells and the atrial gland of adult snails but also in other parts of *Aplysia's* nervous system, and to trace the development of the neurons in which the gene is expressed. We do this by means of two techniques: in situ hybridization and immunocytochemistry. The former depends on the fact that wherever a particular gene is expressed one finds the messenger RNA transcribed from it. We cut a thin section of tissue, mount it on a microscope slide and treat it so that the RNA in its cells is accessible to molecular hybridization. When ELH genes labeled with a radioactive isotope are applied to the slide, they hybridize with the complementary messenger RNA, whose location is revealed by autoradiography. Immunocytochemistry, on the other hand, reveals the presence of the peptide. An antibody to the ELH peptide is applied to a slide and binds to the peptide if it is present. A second antibody, chosen to bind to the first one and linked to a fluorescent dye, is applied, and inspection under ultraviolet illumination reveals the cellular location of the peptide.

We have done hybridization and immunofluorescence studies of the snail's entire central nervous system. The two techniques detect individual neurons expressing an ELH gene even in ganglia with several thousand nonexpressing cells. The bag cells and the atrial gland are clearly the major sites of expression of the ELH genes, but the experiments show that ELH messenger RNA and the peptide itself are also synthesized in an extensive network of nerve cells, not only in the abdominal ganglion but also in three other ganglia. In similar experiments Strumwasser and Arlene Y. Chiu have also identified neurons outside the bag-cell clusters that synthesize ELH. It appears that ELH may have an extensive role as a neurotransmitter throughout the *Aplysia* nervous system.

The presence of scattered ELH-producing cells in four of the ganglia raises the question of how the cells arise during development. By in situ hybridization in developing animals we have learned when the ELH genes are first expressed and where the neurons expressing ELH originate. In all animals the nervous system develops as a specialization of the ectoderm, the embryonic body surface. The ELH genes are active very early in the larval stages of the snail's development, in a zone of primitive cells that line the body wall and are destined to become neurons. Later in development these cells leave the body wall and migrate, by crawling along strands of connective tissue, to their ultimate locations in the adult nervous system. It may be that a single primitive neuron divides to generate a cluster of ELH-producing cells, which migrate not only to the bag cells but also to scattered sites throughout the nervous system.

In *Aplysia* at least three genes encode a number of neuroactive peptides that have roles in the circuits governing a complex but stereotyped behavioral repertoire. In the mammalian brain too a number of peptides have been identified that seem to mediate specific behaviors [see "Neuropeptides," by Floyd E. Bloom; Scientific American, October, 1981]. What properties of neuropeptides, and in particular what characteristics of their mode of synthesis, suit them to a role as mediators of behavior?

The behavioral potential of an organism is at least in part dictated by a rigid network of connecting nerve cells. Much of the communication between neurons is local and is mediated by a neurotransmitter such as acetylcholine or norepinephrine, which is released from a neuron, crosses the narrow synaptic space and makes point-to-point contact with another neuron. Neuroactive peptides too can function locally as neurotransmitters. In addition, however, they can be secreted into the circulation to serve as neurohormones that act on several distant targets and thus generate several discrete activities. Peptides thereby make available an additional communication network that supplements the hard-wired network of interconnected nerve cells.

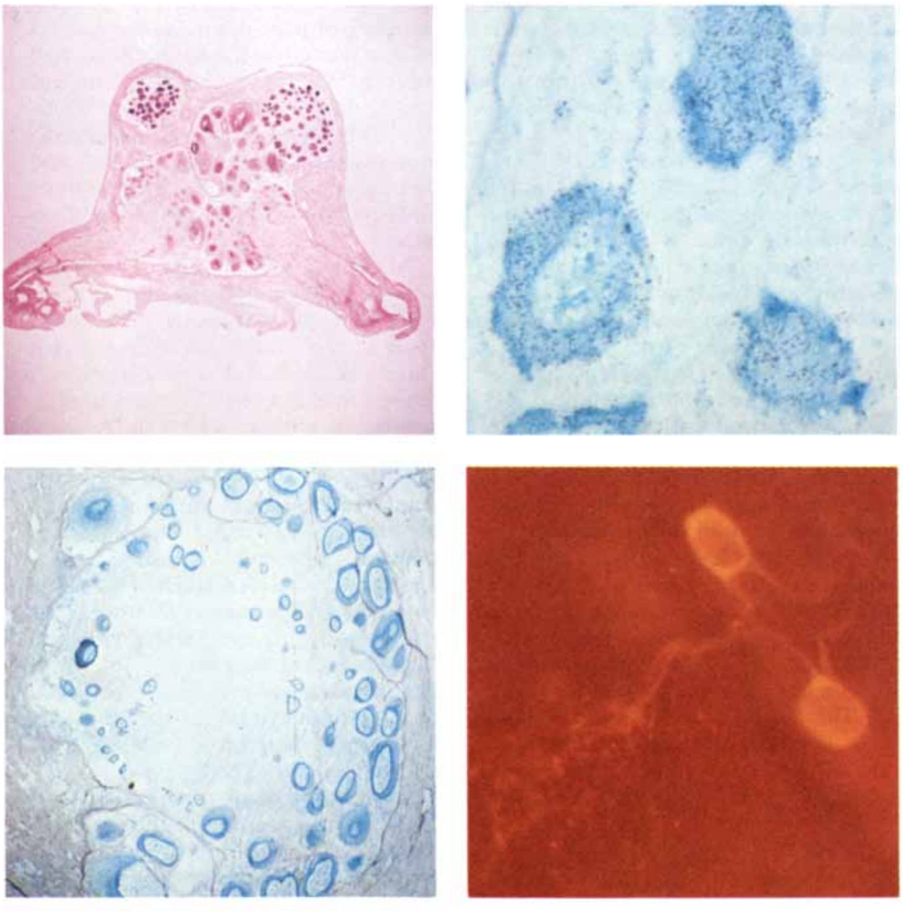
The diversity of a neuropeptide's target sites can enable the peptide to coordinate physiological events with particular behaviors. For example, the injection of the peptide angiotensin II elicits spontaneous drinking in vertebrates by acting on neurons of the hypothalamus in the brain. The peptide also acts indirectly on the kidney to promote the reabsorption of sodium and water into the bloodstream instead of their excretion. Both of these two quite different effects serve to rehydrate the animal. ELH provides another example. In *Aplysia* it acts locally to excite specific neurons in the ganglia that may cause such behaviors as head waving and such physiological

© 1984 SCIENTIFIC AMERICAN, INC

changes as an increase in heart rate. At the same time it acts at a distance, causing the reproductive duct to contract and expel eggs.

A number of egg-laying peptides of *Aplysia* are cleaved from a single polyprotein precursor. Some other neuropeptides are known to be synthesized in the same way. The polyprotein pathway would seem to have several advantages. For one thing, it provides a simple mechanism for the control of both peptide synthesis and release. Different peptides encoded by a single gene can be synthesized simultaneously under the control of a single regulatory agent. Moreover, the various small peptides of a polyprotein may, after cleavage from the precursor, be packaged in the same vesicle, to be released from the synthesizing cell simultaneously by a single stimulus.

The generation of multiple peptides from one precursor also provides a solution to a number problem: the number of genes in an animal genome is simply not large enough to specify the diversity of behaviors exhibited by a species. The information potential of a single gene can be increased if the gene's protein product is cleaved differently in different cells or in response to different stimuli. An example of alternate protein processing has been noted by Edward Herbert of the University of Oregon and James L. Roberts of the Columbia College of Physicians and Surgeons and also by Richard E. Mains and Betty A. Eipper of the Johns Hopkins University School of Medicine. They found that a single precursor is processed to make adrenocorticotrophic hormone (ACTH) in the anterior lobe of the pituitary gland and to make an endorphinlike peptide in the posterior lobe. We have observed one pattern of cleavage of the ELH precursor in a single cell. In principle, given the cleavage sites we have noted, the protein could be cut in different ways to generate more than 2,000 different combinations of peptides, and each combination could activate a different pattern of behavior. The potential for diversity is greater still because the ELH genes constitute a family of genes expressed in different tissues.



SITES OF ELH SYNTHESIS are revealed by in situ hybridization with the peptide's messenger RNA (*top left and top right, bottom left*) and the binding of antibodies to the peptide itself (*bottom right*). A section of the abdominal ganglion was exposed to an ELH-gene probe labeled with a radioactive isotope (*top left*). The probe hybridized with ELH messenger RNA, whose location was revealed by autoradiography. The black dots show hybridization in the bag cells (rounded clusters at the upper margin of the ganglion) and also in a single neuron near the middle of the ganglion. The magnification is about 17 diameters. Magnification to about 400 diameters shows ELH messenger RNA in the cytoplasm of bag cells and their processes (*top right*). ELH messenger RNA is also detected in a single cell of a different group of neurons, the pleural ganglion (*bottom left*). In an immunofluorescence study (*bottom right*) an antibody to ELH was applied to a section taken from the abdominal ganglion. A second antibody, directed against the anti-ELH one and labeled with a fluorescent dye, was applied. The bright orange fluorescence shows ELH in two abdominal-ganglion cells and their processes.

The polyprotein pathway offers temporal flexibility by allowing the various peptides to remain active for different lengths of time. The stability of a peptide can be influenced by certain postsynthetic modifications such as transamidation and even by size (because small peptides cannot easily fold into compact structures and are therefore likely to be degraded sooner than large ones). Both the pattern of amidation and the size of the various egg-laying peptides are consistent with their functional requirements. The peptides (ELH and the *A* and *B* peptides) that can act as hormones are consistently amidated and are longer, and so their active life is presumably extended. The peptides (such as the alpha and beta bag-cell factors) that seem to act only locally as neurotransmitters are shorter and are not amidated. They are presumably degraded faster, like most other neurotransmitters.

The organization of genes encoding polyproteins provides a striking degree of evolutionary flexibility. The interspersal of sequences encoding a set of active peptides in a gene that also encodes nonfunctional protein leaves room for the evolution of new active peptides without alteration of the original set. The ELH gene, for example, has a 240-nucleotide insertion that is not present in the homologous genes expressed in the atrial gland. The insertion encodes three peptides (two of which appear to reflect a small internal duplication), thus expanding the polyprotein's array of coordinated peptides without affecting the synthesis of active ELH.

Finally, the same peptide may be incorporated in several different precursors encoded by different genes. Consider head waving in *Aplysia*. A characteristic waving of the snail's head takes place during feeding as well as during egg laying. The same peptide or peptides could elicit the same behavioral component (head waving) in two very different contexts. To this end the head-waving peptide (or peptides) may be encoded in some other gene—one implicated in feeding behavior—as well as in the ELH gene. In this way complex behaviors could be assembled by the combination of simple units of behavior, each unit mediated by one peptide or a small number of peptides.

© 1984 SCIENTIFIC AMERICAN, INC

Strong medicine for feverish health-care costs.

A new blood analysis system which embodies Kodak's expertise in chemistry, optics, physics, and electronics is helping clinical labs control costs. Its flexible software design is bringing new ease and reliability to the process of blood analysis. It can improve laboratory efficiency, increase productivity, and help to put health-care costs on the road to recovery.

The new system incorporates numerous technological inventions with such intricate design and engineering that it can perform a full range of routine tests —including kinetic enzyme, as well as potentiometric and colorimetric analysis—in operator-preferred sequence. If this sounds like quite an accomplishment, it is.

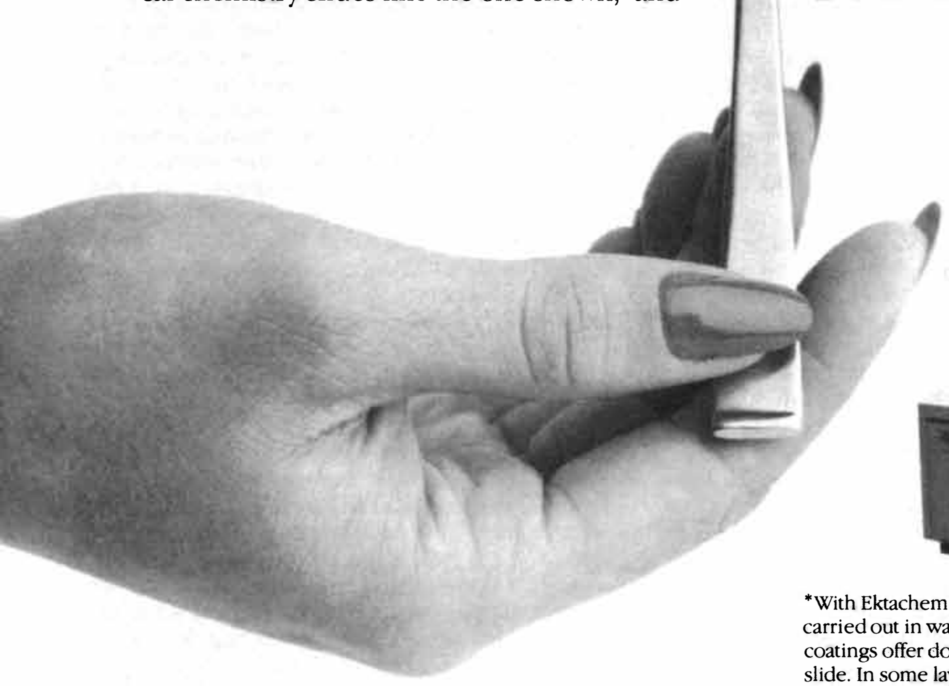
The Kodak Ektachem 700 analyzer incorporates 20 megabytes of hard-disk memory, relies upon Kodak's dry layer-coated Ektachem clinical chemistry slides like the one shown,* and it produces hard-copy results for physicians' evaluation. In just 5 minutes!

But its big advantage is selective testing. At the touch of a CRT screen, this analyzer performs any combination of one to 26 assays and has the potential to report up to 7 additional calculated results on a single patient sample. That's a big plus, because it helps to eliminate wasted tests and wasted time, and contributes to operating economy.

Surprised that we're so committed to health care? You shouldn't be. We've been serving diagnostic medicine with radiography products for more than 80 years.

If you'd like technical papers dealing with the technology involved in Kodak Ektachem products, write: Eastman Kodak Company, Dept. GBSA-10, 343 State Street, Rochester, NY 14650.

Kodak. Where technology anticipates need.



*With Ektachem slides, complex sequential reactions can be carried out in ways not possible in solution chemistry. Multilayer coatings offer domains for multiple reactions within the single-use slide. In some layers, interfering substances can be trapped or altered; in other layers, reactions can be run which produce measurable signals.

© Eastman Kodak Company, 1984

© 1984 SCIENTIFIC AMERICAN, INC

SCIENCE AND THE CITIZEN

Texts for Today

Fifty-nine years after the Scopes trial, 16 years after the U.S. Supreme Court finally invalidated on constitutional grounds a state law prohibiting the teaching of evolution in public schools and colleges and two years after a Federal judge struck down an Arkansas law mandating the balanced treatment of "creation science" along with "evolution science," creationists are successfully reducing the exposure of American high school students to the theory of evolution through natural selection, the organizing principle of modern biology. Their current strategy is to minimize or eliminate coverage of evolution in textbooks, and their primary tactic is to influence the adoption of textbooks by state and local boards of education.

The creationists' major success has been in Texas, where the State Board of Education does not simply publish a list of acceptable books for purchase by local boards, as is the case in many states, but itself buys the books. The Texas board is therefore the country's major purchaser of high school biology texts, accounting for some 10 percent of the total national market. In January the Texas board published Textbook Proclamation 60, which sets guidelines for the content of the science and biology books it will buy starting in 1985. The guidelines do not require any coverage of Charles Darwin or of evolution. That does not mean evolution may not be mentioned, but any discussion of the subject is circumscribed by the state's "general content rules" for textbooks. These require that books treating the theory of evolution "shall identify it as only one of several explanations of the origins of humankind." Books that treat the subject "substantively in explaining the historical origins of humankind shall be edited, if necessary, to clarify that the treatment is theoretical rather than factually verifiable," and they must carry a statement on an introductory page to the effect that any material on evolution is "presented as theory rather than fact."

Biology teachers and others have objected to the singling out of evolutionary theory as if it were more "theoretical" than any other theory. Scientists in general have objected that the rules show a fundamental misunderstanding of what science is and how it works. In a letter to the board proposing amendments to the proclamation in behalf of People for the American Way, a group of science educators wrote that the board "simply does not understand the difference between theory and fact in science." Theories are based on facts, the letter continued. "There is a slow progression toward increasing certainty" as theories are tested by newly observed facts, "but at no point do 'theories' become 'facts.' Evolution theory has been subjected to this kind of rigorous verification over the last 100 years and in its broad outlines is as certain as anything we have in science. We are as sure that biological evolution has occurred, and continues to occur, as we are that matter is composed of atoms." The educators proposed changes in the proclamation to require the inclusion in textbooks of material on the nature of scientific investigation, on evolution and on such details of modern biology as the role of DNA in heredity, but the proposals were ignored.

The importance of the unitary Texas market makes it likely that many publishers will edit their books to conform to the Texas board's taste. That seems to have been happening ever since the Texas content rules were first set forth in 1974, according to Barbara B. Parker of People for the American Way and Wayne A. Moyer, a former executive director of the National Association of Biology Teachers who is now a consultant to the organization. They quote one publisher's representative as saying: "We'll give them [Texas] whatever they want." Whereas a 1968 edition of *Biology,* a high school text published by Silver Burdett, devoted more than 22,000 words to topics related to evolution, the 1981 edition of the book had fewer than 4,400 words on the subject. The 1973 edition of *Modern Biology,* published by Holt, Rinehart and Winston, stated that scientists "do not doubt that organisms living today descended from species of previous ages," but subsequent editions have omitted the statement. According to Gerald D. Skoog of Texas Tech University, who has studied textbook coverage of the subject over the years, in 1981 Laidlaw Brothers, a division of Doubleday, omitted from its *Experiences in Biology* any explicit coverage of evolution and indeed failed to mention the word.

There are some signs of a counteroffensive. In 1982 the New York City Board of Education turned the tables on publishers attempting to meet the Texas requirements: it rejected three textbooks, including the Laidlaw one, for inadequate treatment of evolution. A Texas state senator has asked the state's attorney general for an opinion on the constitutionality of the textbook-content rule. Last year Governor Bruce E. Babbitt of Arizona vetoed a bill that would have required the origin of man to be taught as "theory," and creationist bills failed to pass in the legislatures of Connecticut, Mississippi and West Virginia. A creationist law that was enacted in Louisiana in 1982 is under attack in the courts.

Better Next Time

The development of nuclear power is proceeding more or less steadily outside the U.S. but is now at a virtual standstill in this country. One cause is self-fulfilling: public opposition to nuclear-power plants has changed their political and regulatory environment, leading to delays in construction and commissioning that increase their cost and make them less competitive with oil- and coal-fired plants. Even without this handicap, however, the U.S. nuclear-power industry would have trouble overcoming the escalating costs of all construction, the slower growth in the demand for electricity and the difficulty of financing long-term, capital-intensive projects of any kind in a prolonged period of high interest rates.

Since 1975 a total of 89 new reactors have been canceled by U.S. utilities, accounting for a net loss in future generating capacity of more than 85,000 megawatts. Only two nuclear-power plants ordered in the past nine years have not subsequently been written off, and work has yet to begin on them. The next increment in nuclear generating capacity will come entirely from plants ordered a decade or more ago. According to a survey conducted by the Atomic Industrial Forum, Inc., 14 such plants will be ready to receive operating licenses in 1984; added to the 81 plants now in operation, they would boost the nation's nuclear generating capacity to a total of 81,000 megawatts. Assuming the remaining 50 or so plants now under construction or on order are eventually completed, it is projected that nuclear-power production will double by 1990 and will then account for about 20 percent of the total U.S. electrical output. That would be a far cry from earlier projections, which as recently as the early 1970's set a goal of 1,200 nuclear-power plants for the U.S. by the end of the century.

Does that mean the nuclear option cannot be counted on for any additional contribution to generating capacity in the U.S.? Not necessarily, in the view of Richard Lester, a nuclear engineer on the faculty of the Massachusetts Institute of Technology. Lester suggests that the present interruption in the development of nuclear power in the U.S. may be a blessing in disguise for the beleaguered nuclear-power industry. He maintains that the industry's main problem all along has been "its inability to construct and operate the present generation of nuclear power plants with

© 1984 SCIENTIFIC AMERICAN, INC

PONTIAC FIERO

BURSTS ON THE SCENE TO RAVE REVIEWS!

From the moment the hot new mid-engine Fiero hit America's streets, it became the driving sensation of the year. And the automotive critics agree:

"The Fiero 2M4 is a thrill to look at, a joy to ride in, and a ball to drive." – CAR AND DRIVER September 1983

Fiero features fully-independent suspension, rack and pinion steering and power four-wheel disc brakes, all standard. But Fiero excitement isn't limited to the way it drives:

"The Fiero has its own brand of fresh technology: a space frame machined to tolerances normally reserved for engines and transmissions." – MOTOR TREND September 1983

The Fiero space frame is "milled and drilled" for precise fit of the Enduraflex™ body panels, which resist minor dents and will never rust. And what about Fiero's bottom line?

"We welcome its addition to the ranks of affordable machines for enthusiast drivers." – ROAD & TRACK September 1983

Fiero prices start at just $7,999.† Even at $8,499† (Fiero Sport Coupe shown), Fiero takes its rightful place among the great cars of the world. The mid-engine Pontiac Fiero: practical, durable, economical* and flat-out exciting – only from Pontiac!

†Manufacturer's Suggested Retail Price including dealer prep. Taxes, license, destination charges and optional equipment additional. Fiero Sport Coupe (shown), with WS6 perf. pkg. – $463 additional.

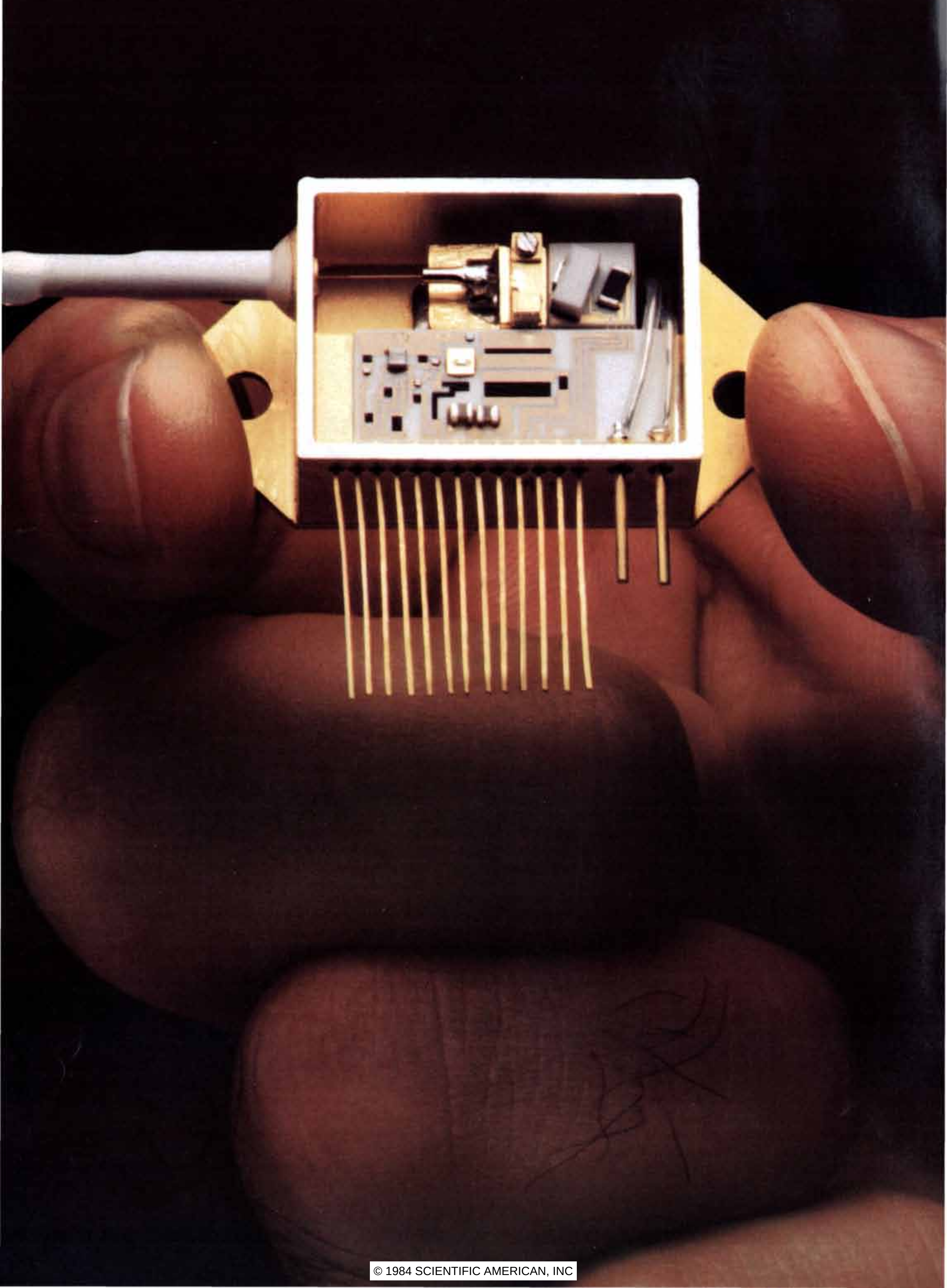
*Pontiac Fiero Sport Coupe offers an EPA EST MPG of 26 and a highway estimate of 40. Use estimated MPG for comparisons. Your mileage may differ depending on speed, distance, weather. Actual highway mileage lower.

Some Pontiacs are equipped with engines produced by other GM divisions, subsidiaries or affiliated companies worldwide. See your Pontiac dealer for details.

GM

PONTIAC WE BUILD EXCITEMENT

© 1984 SCIENTIFIC AMERICAN, INC



© 1984 SCIENTIFIC AMERICAN, INC

Rockwell International know-how: It's in lightwave systems for advanced telecommunications. It's in everything we do.

Optical hybrid circuits designed by Rockwell International for its lightwave communications systems reflect the advanced technology that's made Rockwell one of the major suppliers to the telecommunications industry. Rockwell's fiber optics research and development is helping the industry to benefit from the increased channel capacity available in lightwave communications.

Rockwell provides high technology products and systems to three telecommunications markets: communications transmission, communications switching, and telephone company special services.

Only a company with the know-how to combine high technology with outstanding engineering and management skills can meet such broad needs of the telecommunications industry. That know-how goes into our electronics business and into everything we do at Rockwell.

In aerospace: where we're responsible for the design, integration management and production of the U.S. Air Force's B-1B Long Range Combat Aircraft. Flight testing and production are ahead of schedule, with delivery of the first production aircraft scheduled for 1984.

In the automotive industry: where our technology and reputation for dependable, quality design for car and truck components have made us a major supplier of drivetrains for U.S. Army trucks.

In general industries: where advancements such as electronic press controls and total production systems have made our Goss presses a world leader in offset newspaper printing.

We're an $8 billion company where science gets down to business in four diverse areas. And that diversity has helped us achieve eight consecutive years of increased earnings and impressive growth.

If you are interested in any of our products or want to learn more about us, write: Rockwell International, Department 815S-3, 600 Grant Street, Pittsburgh, PA 15219.

Rockwell International

...where science gets down to business

Aerospace/Electronics
Automotive/General Industries

© 1984 SCIENTIFIC AMERICAN, INC

the consistently high standards the technology requires."

Writing in *The Wall Street Journal,* Lester points to several "lessons from abroad" that "can help the industry do better next time." In particular he cites the case of Japan, "the most recent former nuclear pupil of the U.S. to reach the front rank of civil nuclear nations." In contrast to the U.S., where efforts to improve the basic design of the light-water reactor—the mainstay of the U.S. nuclear-power program—"have almost ground to a halt," he says, Japan has been pursuing "an independent technology-improvement program, while simultaneously building a strong domestic reactor industry" based on light-water reactors originally developed in the U.S. Moreover, in the future improvements in light-water reactors "will receive an even higher priority in Japan's nuclear strategy."

Today's light-water reactors, Lester points out, "are vastly more complicated (and expensive) than those built a decade ago, and they tax the construction and operating expertise of even the most sophisticated utilities, let alone their weaker brethren. The main driving force behind this trend has been the growing concern over reactor safety. But the technical response has lacked a strategic purpose.... The present hiatus offers a chance for a fundamental reexamination of light-water reactor designs and regulations. The goal should be to develop advanced, much simplified plants with equivalent or greater safety at significantly lower cost."

Assuming a continuation of present economic trends, Lester predicts, "most utilities henceforth will prefer to build generating capacity in smaller increments. In such a market, there may be a sizable niche for 'modular' nuclear power stations—clusters of small reactors each with a capacity of, say, 200 electrical megawatts (vs. 1,000 megawatts in many of today's plants).... Gas-cooled reactors seem especially well suited to this application."

Finding that the time has come "to investigate seriously the possible contribution of advanced reactors—whether improved LWR's or more radical innovations—to a revival of the U.S. nuclear-power industry," Lester concludes that "the single most important requirement is for the electric utilities themselves to take the lead in distilling their collective nuclear experience and specifying their technical needs over the next few decades. The government has a key supporting role—not, certainly, in picking winners, but in conducting supporting safety-related tests and promptly addressing important new regulatory issues." The alternative to technological innovation is presented starkly: "A failure to improve significantly on present light-water reactors in itself might not kill the U.S. nuclear option. But the sort of thinking revealed by an unwillingness even to try very well might."

The Soliton Laser

Few experimental tools gained such rapid and widespread acceptance in the laboratory as the laser. The main reason for the success of the laser is its ability to deliver radiation of a sharply defined frequency at a high rate to a small volume of material. Perhaps less widely appreciated is that laser energy can be pulsed, or delivered to an experimental sample in short bursts. Such pulses make it possible to manipulate physical and chemical processes in time as well as in the domains of space and energy. The study of these processes on the atomic scale calls for very sharp pulses indeed. For example, the temporal resolution of the transfer of energy within a molecule or the evolution of the electronic energy states in a solid calls for pulses as short as a picosecond (10^{-12} second) or less.

Now a new laser has been built by Linn F. Mollenauer and Roger H. Stolen of AT&T Bell Laboratories that has generated pulses of infrared radiation lasting only 210 femtoseconds. (A femtosecond is a thousandth of a picosecond, or 10^{-15} second.) The laser is called a soliton laser because its output is controlled by a soliton, or solitary wave. The pulses emitted by the laser are by no means the shortest anyone has ever made; that distinction belongs to Charles V. Shank and his colleagues, also of Bell Laboratories, who have generated pulses of visible light lasting for only 30 femtoseconds. The infrared pulses, however, have properties quite different from those of the visible ones. Unlike the visible pulses, they have a shape and a duration that can be controlled to a high degree of accuracy.

The idea for the soliton laser can be traced to a chance observation of a water wave by the British naval architect John Scott Russell in 1844. Russell noticed that an unusual wave was generated by the bow of a boat in a channel. When the boat stopped, the bow wave continued down the channel without either cresting or dissipating.

The wave observed by Russell is now known mathematically as the fundamental soliton; it is a wave that propagates without changing its shape. Nevertheless, a soliton can have higher-order forms as well as the fundamental. The second-order soliton is a wave that begins as a bell-shaped disturbance and narrows as it propagates. As it narrows, the height of the wave increases until both the width and the height reach extreme values; the wave then returns to its original form and the process is repeated indefinitely. Solitons of an order higher than two exhibit more complicated motions but their behavior is always periodic. For example, the third-order soliton contracts to a sharp spike, splits into two smaller peaks, contracts back to the sharp spike and finally returns to its original form before the sequence is repeated.

The soliton laser exploits the second-order soliton. A pulse is initially generated in the soliton laser by a device that can function as a laser in its own right. The device is called a color-center laser, and it can be tuned to emit single-frequency infrared radiation across a wide range of frequencies. When the color-center laser operates independently, it emits a pulse at a wavelength of 1.5 micrometers that lasts for seven picoseconds. In the soliton laser the pulse emitted by the color-center laser travels through a beam splitter, which sends part of the energy of the pulse to the output of the soliton laser and the other part to the laser cavity, where the pulse undergoes its main transformation.

The laser cavity is a glass fiber that confines the pulse to a core about 10 micrometers in diameter. The pulse that enters the laser cavity propagates down the cavity fiber, is reflected by a mirror at the end of the fiber and returns along the fiber to the color-center laser. Because the core of the fiber is so small, the confinement of the pulse gives rise to a high intensity of light within the core. The speed of light in glass varies slightly with the light intensity, and so the high light intensity in the core leads to changes in the speed of the light that cause a change in the shape of the pulse.

If there were no such variation in the speed of light with intensity, the input pulse would disperse, or broaden, as it propagated. Moreover, if there were no tendency for the pulse to disperse, the dependence of the speed of light on intensity would have no effect on the shape of the pulse. The effects of light intensity, however, combined with the effects of dispersion cause the pulse to contract. The pulse that returns to the color-center laser along the fiber in the cavity is shorter than the pulse that was sent out. The shortened pulse sweeps up the energy stored in the color-center laser and stimulates it to emit a second pulse; in effect the color-center laser acts as a mirror that amplifies the pulse.

The process is repeated several times, and it results in a cascade of increasingly narrow pulses. There is a limit, however, to the narrowing of the pulse. When the pulse reaches a characteristic width, determined by the length of the cavity fiber, it becomes a second-order soliton. On its way out to the mirror at the end of the fiber the soliton pulse narrows, but it broadens on its return; in short, when the pulse becomes a soliton, the fiber no longer introduces any net change in the shape of the pulse. Although the soliton might therefore seem to be a limitation,

© 1984 SCIENTIFIC AMERICAN, INC

SCIENCE/SCOPE

Weather satellites help airlines save $700 million in fuel bills, thanks to timely pictures and other data. Meteorologists who plan trips for airlines use satellite information to find efficient travel routes. For example, normally the route between Los Angeles and Hawaii is about 2300 miles. In case of a headwind, the airline will select an alternate route. Even though the new route may be 2500 miles long, better flying conditions could save the airline $1500 in fuel. The latest three spacecraft in the Geostationary Operational Environmental Satellite (GOES) system were built by Hughes Aircraft Company. The sensing devices, which provide the daytime and nighttime pictures shown on TV and printed in newspapers, were built by a Hughes subsidiary, the Santa Barbara Research Center.

Extremely sensitive high-speed photodetectors, key components in microwave fiber-optic communications links, have been fabricated by Hughes research scientists. The devices are gallium arsenide Schottky barrier type photodetectors. They have demonstrated a flat frequency response to at least 20 GHz and have the potential to approach millimeter wavelengths. They also have quantum efficiencies as high as 70%, roughly a 2 to 5 db improvement over existing high-speed detectors.

Approximately 2900 U.S. and foreign patents have been granted to Hughes inventors in the last 15 years. To honor its best and brightest talent, the company each year presents the Lawrence A. Hyland Patent Award to a handful of its outstanding inventors. The award, which includes an honorarium, is named for the company's retired chairman of the board. A leading electronics inventor and radar pioneer, Mr. Hyland was granted nearly 40 patents. Among his earliest was the one for the first shielded spark plug that permitted radio equipment to be used effectively aboard aircraft. To date 104 inventors have received the Hyland award.

Over 20 nations protect their sovereign airspace with command, control and communications systems produced by Hughes, the world's most experienced developer of automated air defense systems. The systems are comprised of air defense radars, computers, displays, and other electronic subsystems. Sightings are transmitted through data links to data processing centers, where computers identify, automatically track, and report the aircraft's speed, altitude, and course. The systems are tailored to the requirements of each country based on geography, military equipment, and size and structure of military forces. Nations equipped with Hughes systems include Japan, Switzerland, the U.S., Spain, Canada, and European NATO members Belgium, Denmark, Greece, Italy, the Netherlands, Norway, Turkey, the United Kingdom, and West Germany.

An ultramodern facility spanning 1.75 million square feet is the showcase where outstanding Hughes engineering combines advanced manufacturing techniques and production processes. Our complex is complete, so we're looking for experienced and graduating engineers to work on such programs as: infrared thermal imaging systems, laser rangefinders and designators, and missile launching and guidance systems. Send your resume to Hughes Electro-Optical and Data Systems Group, Professional Employment, P.O. Box 913, E9/W101, Dept. SE, El Segundo, CA 90245. Equal opportunity employer.

Creating a new world with electronics

HUGHES

HUGHES AIRCRAFT COMPANY

For additional information please write to:
P.O. Box 11205, Marina del Rey, CA 90295

© 1984 SCIENTIFIC AMERICAN, INC

it is quite the reverse. The shape of the soliton is well understood and is determined only by the characteristics of the fiber. Moreover, the width of the soliton is proportional to the square root of the length of the fiber, and so the width of the soliton that finally evolves in the fiber can be narrowed simply by reducing the fiber length.

The energy output of the soliton laser can be adjusted by changing the reflective characteristics of the beam splitter. If the output pulse is focused into a second fiber, that pulse too can become a high-order soliton. The order of the output soliton depends on the ratio of the energies of the two split beams. The output fiber can then be cut at the length at which the high-order output soliton is most compressed. The scheme leads to an additional reduction in the width of the pulse. The 210-femtosecond pulse was generated in 67 centimeters of cavity fiber without such refinements in the output; according to Mollenauer, it should not be difficult to focus the pulse in time by another factor of 10.

Silent Partner

How frequent are multiple pregnancies? The answer to the question is less straightforward than it might seem. The twinning rate (which is taken to include all multiple conceptions) varies with race, country and even part of the country. Estimates range from 6.4 per 1,000 maternities in Japan to 45 per 1,000 in Nigeria. In the U.S., textbooks put the rate at around one in 90, but this is considered an uncertain figure. One difficulty in arriving at accurate estimates of the twinning rate has been that until the advent of ultrasonography there was no way of detecting multiple gestations early in pregnancy and hence no way of knowing how many twins are lost during that period. Over the past decade investigators armed with ultrasound equipment have shown there are many "vanishing twins," suggesting that estimates of the twinning rate have been far too low. Writing in *Acta Geneticae Medicae et Gemellologiae,* Helain J. Landy, then at the Northwestern University Medical School, and Louis and Donald Keith of the Center for the Study of Multiple Birth in Chicago summarize the results of this research.

According to Landy and the Keiths, the concept of a "vanishing" embryo should not be altogether surprising. Some workers have suggested that a large fraction of all pregnancies end before the sixth week, before spontaneous abortion is generally detected, and that this is nature's way of eliminating an abnormal embryo. What has not been widely recognized is that in many instances one twin may be lost at an early stage as well, with the surviving twin being misinterpreted as a single birth.

Ultrasound (usually with frequencies above one megahertz) is much better than X rays at imaging soft tissue, and at low doses it is much less likely to do harm to the fetus or the mother. With it diagnoses of multiple pregnancies have been made as early as the fifth week, well before the development of the fetal skeleton that X rays could pick up, and at a time when the embryo is no more than a centimeter long. The gestational sac, forerunner of the amniotic sac, appears as a ring-shaped image on the cathode-ray tube of the ultrasonograph. Landy and her colleagues report that nine different ultrasonographic studies have documented the disappearance of a gestational ring after more than one had been detected.

One of the largest studies examined a cohort of 8,362 women, of whom 159 were diagnosed as having a multiple pregnancy, giving a twinning rate of 19 per 1,000. For those women scanned ultrasonically before the 10th week (less than half of the cohort) who were definitely diagnosed as having a multiple pregnancy and who were followed to delivery the investigators calculated a twin-disappearance rate of 78 percent. On the other hand, all the 89 women who were diagnosed as having a multiple pregnancy after the 14th week went on to deliver multiple infants. These and other data suggest that disappearance tends to come either in the latter half of the first trimester or in the early part of the second.

Some workers question the accuracy of the high rates, partly because examination of the placenta and fetal membranes after the birth of a single infant whose twin has supposedly vanished rarely yields physical evidence of the twin. This may have to do with the physiological processes leading to disappearance, which are still unclear. The gestational sac, containing either an embryo or a "blighted ovum" (a fertilized egg that never developed normally), may be resorbed into the mother's tissue; resorption is known to occur in single pregnancies.

Even with ultrasound it is hard to detect a multiple pregnancy in the first trimester and easy to misinterpret the image in a way that could lead to overdiagnosis and therefore to an overestimation of the disappearance rate. On the other hand, it would also be easy to miss a gestational sac that may be visible for only one to two weeks before it disappears; furthermore, obstetricians may tend not to scan the rest of the uterus with ultrasound after finding the first embryo. Landy and her colleagues conclude that the phenomenon of the vanishing twin is well established but that it is too early to specify a rate.

The only clinical complication that seems to be associated with twin disappearance is a slight vaginal bleeding in some cases. The bleeding is of some significance, however, because vaginal bleeding in the first trimester is generally taken as a sign that spontaneous abortion is going to occur. Obstetricians may thereupon perform a dilation and curettage ("D&C") to remove from the uterus unexpelled products of conception that might otherwise give rise to an infection. If they do not first scan the uterus with ultrasound, Landy and her colleagues remark, they may inadvertently abort a healthy twin that has lost its partner.

Flipping Fields

The magnetic field of the earth is basically a dipole, so that the earth is like a bar magnet with a north and a south magnetic pole. On numerous occasions in geologic history the dipole has flipped: the north pole becomes the south and vice versa. The flip itself is fast; on a geologic time scale it seems instantaneous. On the other hand, the intervening periods when the dipole is stable are quite variable: they have ranged over three orders of magnitude, from 10,000 years to more than 10 million. The statistics of these periods show little pattern, and so it has been assumed that the flips come at random intervals; at least they seem to have nothing to do with anything happening on the earth. A new analysis by J. G. Negi and R. K. Tiwari of the National Geophysical Research Institute in India suggests that the flips tend instead to be correlated with periodic phenomena in the galaxy.

Negi and Tiwari report their findings in *Geophysical Research Letters.* They argue that the methods employed to examine the statistics of the flips (notably Fourier analysis) suit the study of signals that vary sinuously, whereas the record of the flips more nearly resembles a "binary telegraph signal": a sequence of positive and negative pulses of constant amplitude, with instantaneous transitions from one pulse to the next. For such a signal a different method is more appropriate. Like Fourier analysis it yields the power spectrum of the signal; that is, it decomposes the signal into a sum of simple waveforms. Here, however, the simple waveforms are square waves, not sine waves. Thus frequency is replaced by "sequency" (one-half the average number of crossings from positive to negative or negative to positive per unit time).

Negi and Tiwari apply their method to the flips in the earth's magnetic field over the past 570 million years. The flips are known because the direction of the earth's magnetic field is impressed into volcanic rock as it rises into place and solidifies to form new ocean floor. Five distinct spectral peaks emerge, suggesting that the flips tend to come at intervals of 285, 114, 64, 47 and 34 million years. The first of these numbers marks

© 1984 SCIENTIFIC AMERICAN, INC

EFFICIENTLY HYATT.

Efficiency comes naturally to the people of Hyatt. We understand the special needs of business travelers like no one else in the world. And we know the better we do our job, the better you can do yours.

The Hyatt attitude is unmistakable. From our charming concierges, who can make an unfamiliar city feel more accommodating. To our energetic bellmen, who help with all the things you don't have time to do. To our friendly people at the front desk, like Diana Proctor of Hyatt Regency Chicago, who handle your messages and requests with cheerfulness and warmth.

A little something extra from us can help make your trip a success. An efficient touch of Hyatt.

Don't you WISH YOU WERE HERE℠

ATLANTA

Discover magnificent Peachtree Center, 15 minutes from Atlanta International Airport.

CHICAGO

In Illinois Center overlooking Michigan Avenue and Lake Michigan.

HOUSTON

In the very heart of exciting Houston, walk to business through climate-controlled tunnel system.

NEW YORK

Grand Hyatt on fashionable Park Avenue at Grand Central.

PRINCETON

Two miles from Princeton University, midway between Philadelphia and New York City.

HYATT® HOTELS

For reservations, call your travel planner or 800 228 9000. © 1984 Hyatt Hotels Corp.

© 1984 SCIENTIFIC AMERICAN, INC

© 1984 SCIENTIFIC AMERICAN, INC

There's a simple reason why some cars refuse to become outdated, while others do it all too willingly.

The difference lies in how a car maker applies technology—whether for purposes that are substantial and enduring, or simply to bestow upon cars an annual patina of newness.

And nowhere is this distinction made more vivid than in the BMW 528e.

It is built by a company that developed its first independent suspension in 1932. That began exploring aerodynamics as early as 1923. And that has been incorporating fuel injection into its cars for over a decade.

These and other indications of BMW's foresight are abundantly apparent in the 528e.

Its computer-governed Eta power plant was engineered to reconcile performance and efficiency. It won't have to be rethought the next time the world convulses between energy shortage and energy glut.

Its Service Indicator informs the driver when routine service is warranted. It will remain relevant for as long as cars require maintenance.

And its easily-readable instrumentation provides a wealth of vital data, including seven different measures of the car's operational readiness, as well as accurate fuel-mileage figures. None of it was inspired by a passing infatuation with video games, but by the timeless need of drivers to receive and instantly comprehend information.

In short, the BMW 528e has been programmed to provide a service that most other cars can't:

The assurance that the future will come as a fulfillment, instead of a shock.

© 1984 BMW of North America, Inc. The BMW trademark and logo are registered. European Delivery can be arranged through your authorized U.S. BMW dealer

THE ULTIMATE DRIVING MACHINE.

© 1984 SCIENTIFIC AMERICAN, INC

the highest peak in the spectrum. It also marks the strongest correlation: it compares quite well with the time it takes the solar system to complete one orbit around the center of the galaxy. Each such "cosmic year" lasts for 280 million years. The second, third and fifth numbers compare fairly well with the intervals at which the solar system encounters wavelike variations of density in the plane of the galaxy. The fourth number compares fairly well with an oscillation of the sun above and below the plane of its galactic orbit.

How might the changing gravitational forces encountered by the solar system in its orbit around the center of the galaxy bring on reversals of the earth's magnetic field? The formulation of a mechanism "is not possible at this stage." For now Negi and Tiwari note that the sun's orbit takes it above and below the plane of the galaxy with a period of 85 million years. They hope, therefore, that better geologic data will reveal still another correlation.

Resonances

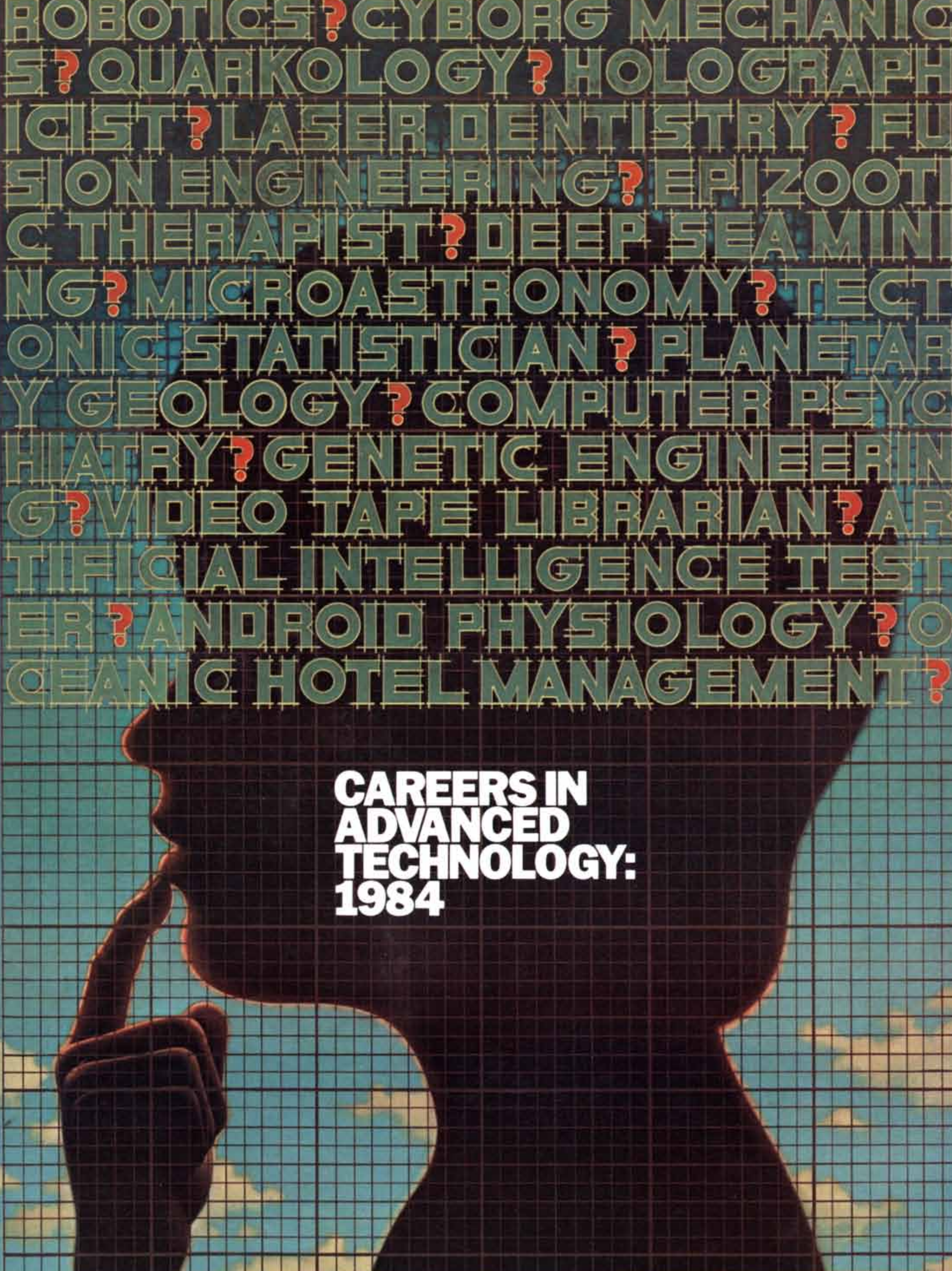
Recent atmospheric studies indicate that if fewer than a third of the nuclear bombs now in existence were detonated, even in an attack without retaliation, the dust and smoke that would be injected into the atmosphere would give rise to a catastrophic "nuclear winter." This discovery has had some curious echoes. Last year was the 100th anniversary of the eruption of the volcano Krakatau, whose atmospheric effects were observed around the world, and 1982 saw the eruption of the Mexican volcano El Chichón, which also has had worldwide effects.

In December, Senator Edward M. Kennedy (D., Mass.) and Senator Mark O. Hatfield (R., Ore.) organized a forum in the Senate Caucus Room on the effects of nuclear war. Among those who spoke was Sergei Kapitza, a well-known Russian physicist. Kapitza called attention to the 1815 eruption of the Indonesian volcano Tambora, which had the greatest atmospheric effects on record. Even though the eruption was in the Southern Hemisphere, the following year six inches of snow fell on New England in June, the summer in Europe was the coldest of any on record before or since and there were worldwide crop failures [see "The Year without a Summer," by Henry and Elizabeth Stommel; SCIENTIFIC AMERICAN, June, 1979].

The lost summer of 1816 seems to have depressed and disturbed many people in ways beyond the obvious. Kapitza cited a poem, "Darkness," written by Lord Byron in June, 1816. The poem is known to many Russians, Kapitza said, because it was translated into Russian by the novelist Ivan Turgenev. "Darkness" is here given in full:

I had a dream, which was not all a dream.
The bright sun was extinguish'd, and the stars
Did wander darkling in the eternal space,
Rayless, and pathless, and the icy earth
Swung blind and blackening in the moonless air;
Morn came and went—and came, and brought no day,
And men forgot their passions in the dread
Of this their desolation; and all hearts
Were chill'd into a selfish prayer for light.
And they did live by watch fires—and the thrones,
The palaces of crownèd kings—the huts,
The habitations of all things which dwell,
Were burnt for beacons; cities were consumed,
And men were gather'd round their blazing homes
To look once more into each other's face.
Happy were those who dwelt within the eye
Of the volcanos, and their mountain-torch:
A fearful hope was all the world contain'd;
Forests were set on fire—but hour by hour
They fell and faded—and the crackling trunks
Extinguish'd with a crash—and all was black.
The brows of men by the despairing light
Wore an unearthly aspect, as by fits
The flashes fell upon them; some lay down
And hid their eyes and wept; and some did rest
Their chins upon their clenchèd hands, and smiled;
And others hurried to and fro, and fed
Their funeral piles with fuel, and look'd up
With mad disquietude on the dull sky,
The pall of a past world; and then again
With curses cast them down upon the dust,
And gnash'd their teeth and howl'd. The wild birds shriek'd,
And, terrified, did flutter on the ground,
And flap their useless wings; the wildest brutes
Came tame and tremulous; and vipers crawl'd
And twined themselves among the multitude,
Hissing, but stingless—they were slain for food.
And War, which for a moment was no more,
Did glut himself again;—a meal was bought
With blood, and each sate sullenly apart
Gorging himself in gloom. No love was left;
All earth was but one thought—and that was death,
Immediate and inglorious; and the pang
Of famine fed upon all entrails—men
Died, and their bones were tombless as their flesh;
The meagre by the meagre were devour'd,
Even dogs assail'd their masters, all save one,
And he was faithful to a corse, and kept
The birds and beasts and famish'd men at bay,
Till hunger clung them, or the dropping dead
Lured their lank jaws. Himself sought out no food,
But with a piteous and perpetual moan,
And a quick desolate cry, licking the hand
Which answer'd not with a caress—he died.
The crowd was famish'd by degrees; but two
Of an enormous city did survive,
And they were enemies. They met beside
The dying embers of an altar-place,
Where had been heap'd a mass of holy things
For an unholy usage; they raked up,
And shivering scraped with their cold skeleton hands
The feeble ashes, and their feeble breath
Blew for a little life, and made a flame
Which was a mockery. Then they lifted up
Their eyes as it grew lighter, and beheld
Each other's aspects—saw, and shriek'd, and died—
Even of their mutual hideousness they died,
Unknowing who he was upon whose brow
Famine had written Fiend. The world was void,
The populous and the powerful was a lump,
Seasonless, herbless, treeless, manless, lifeless—
A lump of death—a chaos of hard clay.
The rivers, lakes, and ocean all stood still,
And nothing stirr'd within their silent depths;
Ships sailorless lay rotting on the sea,
And their masts fell down piecemeal; as they dropp'd
They slept on the abyss without a surge—
The waves were dead; the tides were in their grave,
The Moon, their mistress, had expired before;
The winds were wither'd in the stagnant air,
And the clouds perish'd; Darkness had no need
Of aid from them—She was the Universe.

© 1984 SCIENTIFIC AMERICAN, INC



SCIENTIFIC AMERICAN/ADVANCED TECHNOLOGY CENTERS/MARCH 1984

© 1984 SCIENTIFIC AMERICAN, INC

In Just Four Business Units, RCA Offers More Technical Challenge

RCA
Government Communications Systems
Camden, NJ

Join GCS and help create the world's most sophisticated communications systems. In our Camden facilities, a wide array of Command, Control, Communications and Intelligence Systems are designed, and transferred into full production programs.

Challenging opportunities exist to develop new communications security equipment, intelligence systems, radios, recorders/reproducers, transmission systems and integrated computer network communications systems.

RCA Camden employees enjoy a panoramic view of Philadelphia's Society Hill and Penn's landing.

RCA
Astro-Electronics
Princeton, NJ

Celebrating our 25th anniversary with 82 of our satellites launched, we have an outstanding record in space in communication, meteorological, navigation, and exploration satellites. Starting with the first radio broadcast from space by project SCORE in 1958 to today's all solid-state commercial communication satellites and the onboard computer-controlled meteorological satellites, RCA Astro has been a leader through the innovative application of state-of-the-art technology. Our future business in the space station, planetary explorer, direct broadcast, and meteorological satellite programs provide exciting opportunities at Astro-Electronics.

RCA
Missile and Surface Radar
Moorestown, NJ

RCA Missile and Surface Radar applies systems engineering from conception through manufacture and operation of major electronic defense systems. We have set new standards in the development of precision radar systems ranging from hand-held radars to the advanced AEGIS Guided Missile Cruisers which will form the Navy's Fleet Combat System for the 80's and 90's. We're proud of our advancements in VLSI technology, antenna-microwave engineering, and in the development of digital signal processing. Located in Moorestown, New Jersey, a beautiful suburb just 15 miles outside of Philadelphia, we're close to the shore and mountains.

RCA
Advanced Technology Laboratories
Camden, NJ

Do you enjoy working on the leading edge of new technology development? If so, our division would like you to join our quality staff. We cover a broad span of disciplines, including VLSI and submicron circuit technology, signal processing, advanced software, artificial intelligence, image processing and robotic vision, millimeter wave IC's, speech recognition and GaAs technology. We also have significant activity in electro-optics, optical disk systems, advanced sensors and lasers. Join us if you're interested in pursuing applied research and development in some of today's most challenging technical areas.

See For Yourself.

Explore the RCA spirit of new frontier technological development. For more information, write to the following address. Please be sure to tell us which business unit(s) you are interested in learning more about. RCA Corporation, Manager, Employment & Affirmative Action, Department DJI-S, Route 38, Bldg. 205, Cherry Hill, NJ 08358.

Equal Opportunity Employer
U.S. Citizenship Required

RCA

A Tradition On The Move!

© 1984 SCIENTIFIC AMERICAN, INC

As 1984 gets under way, advanced-technology industry is hiring slightly greater numbers than in the recent past. Technical-college enrollment, which picked up in 1974 as the appeal of science as a career-choice revived, is growing. Firms in biotechnology, in telecommunications, in business and professional computer hardware and software, in aerospace/defense, in process control, measurement, and analysis instrumentation, and in certain energy-related fields, all are seeking more good people. Even pure research, benefiting from more intercorporate cooperation permitted under a relaxed government antitrust posture, is "staffing up."

NO SHORTAGE, NO SURPLUS

"If we continue to go on today's track with the economy improving, with current college-enrollment levels, and with the present emphasis on defense, we'll find few serious shortages or surpluses in the engineering and science fields. We do have serious trouble in some of the biological sciences where there seems to be an oversupply of people," says Betty Vetter, executive director of the Scientific Manpower Commission. According to Bureau of Labor Statistics projections there should be balance between supply and requirements through the 1980s.

The statistical balance between the number of job opportunities and the number of scientists and engineers hasn't made high-tech companies complacent. Lacking a national clearinghouse that matches scientists and engineers to job openings in companies or areas where they are most needed, most high-tech companies expect recruitment will be more difficult in 1984 than in the past few years. The best recruits remain the scarcest.

"With the enormous proliferation of opportunities in small companies as well as large companies, a good person has a tremendous option now to go in many different directions," says Scott Sharpe, vice president of personnel at Computer Sciences Corporation. "So we have to compete aggressively—more aggressively than we ever did before—to get that good person.

"Companies are more sophisticated. In the 70s, when companies expanded, many would drag people in left and right. Now more companies are planning their needs better so they're prepared to get those people when they're needed," Sharpe adds.

"Today, technical professionals are saying 'I want to make a long-term commitment. I want a ten- to fifteen-year, or longer, career here,'" says Claude Kazanski, manager of corporate staffing at Prime Computer. "And they're saying, "I want the environment to support my creative skills, to rub shoulders with other professionals doing creative things.'"

'Rubbing shoulders' leads to a range of communication exchange. "We met 20% of our 1983 hiring requirements from referrals from existing employees," says a manager at a Washington-area electronics firm. "The remainder came about half from our advertising and half from personnel-search firms."

"Our own people are a steady source of recruits," says Pete Portanova, assistant manager for employment at Grumman Aerospace. "Our engineers attend symposia and present papers. They meet people, and say 'Why don't you give me a call,' and then resumes make their way to us."

VALUE OF EXPERIENCE

Companies will be concentrating more on attracting experienced people. Many smaller companies look for experienced people who can contribute to their goals immediately. As a result, salaries and other compensation for experienced people may rise as high-tech companies try to lure employees away from each other. "Good experienced people have always been difficult to recruit. This year —with more companies trying to find the same people—it's going to be even more competitive," says Don Diers, corporate manager of employment programs at Hughes Aircraft Company.

Paul Warren of ITT Dialcom, Inc., in Silver Spring, Maryland, says, "Our marketing people must be able to converse technologically, and they must be comfortable in the presence of important people at large companies. These qualities are rare among new graduates. To date, we've drawn from those who gained experience in office-automation procedures among our competitors. And they probably will be approaching our people the same way."

For established, profitable firms competition for high-demand employees can be overcome, through higher salaries, or, in some cases, lowered standards. For universities with generally uniform faculty compensation plans, and for government agencies with set civil-service pay scales, attracting qualified career people will be more difficult, and acceptance of lower standards more common.

Government agencies continue to appeal to new graduates despite their lower pay scales, since they are ready sources of the on-the-job experience that industry increasingly requires.

TEACHER SHORTAGE

A report recently released by the American Society for Engineering Education (ASEE) points out that between 1973 and 1981 undergraduate engineering enrollment increased 111 percent while faculty increased only 11 percent. From 1969 to 1982 the undergraduate student/faculty ratio went from about 14 to over 21. As a result, most colleges surveyed reported an increase in teaching assistants and part-time faculty. Many reported an inability to offer courses in certain subjects.

The ASEE study estimates slightly over 1,000 unfilled engineering faculty openings nationally. (There are about 18,000 filled full-time engineering faculty positions.) To meet their needs, colleges draw primarily from the annual crop of about 2,800 new Ph.D.'s, of which about 350 are real candidates for college teaching jobs.

Starting salaries in industry for B.S. graduates discourage post-graduate work. As one dean of engineering put it, "It costs a student over $45,000 to get a B.S. and students understandably want a return on that investment immediately. Convincing students to go on for advanced degrees is our number one problem."

The ASEE has set up a center for ideas and solutions from educators and industry. One suggested solution: limit the number of science/engineering students. Another: increase salaries for engineering and science professors beyond the standard professor-level, to better compete with industry. Yet another: look to industry for help, since it is in industry's best interest to maintain technical-education quality, and to ensure a steady supply of new technical talent.

In his 1982 book *Global Stakes*, Ray Stata, founder and president of Analog Devices, writes: "A capacity constraint is evident in the nation's engineering colleges and universities, and they are in no position themselves to alleviate it without considerable help from industry and government." Stata cites an example:

At (the University of Massachusetts) during the fall of 1981, a course entitled "Introduction to Problem Solving Using a Computer" was attended by 500 students. By spring of 1982, the enrollment jumped to 1,200 despite a warning in a student class evaluation book stating: "Nearly all the students complained of the large class size, poor facilities and/or inadequate terminal access."... a student reported... "I had the worst time fighting with 60 other students for the ONE terminal everyone had to use."

Analog Devices, and several other concerned manufacturers, are earmarking fixed percentages of corporate expenditures for direct support of higher education. They are urging Washington-level support as well, through legislators including Massachusetts Senator Paul

© 1984 SCIENTIFIC AMERICAN, INC

Tsongas, whose "High-Technology Morril Act" introduced last June was partly based on Stata's work.

Skilled and experienced people are wanted in advanced technology across the board.

The Electronic Industries Association, in Washington, has been involved in high-tech recruiting at the technician/operative level for years. The Association's membership includes consumer electronics giants like GE, Zenith, Delco, Panasonic, and RCA. "Our members encounter too many customer complaints, too much warranty-service claim activity, and too much undertraining among repair people. There is certainly no need for unemployment for anyone in the electronics technician business," states Association director of product services Don Hatton.

"We've shown our careers-in-electronics film more than 28,000 times," Hatton advises, "and our programs expand year after year. And still we're behind; some estimates call for a 95% increase in the number of computer technicians—people who can repair the things—in the next 10-12 years."

COMPENSATION PLANS IMPROVE

With more companies seeking to slow turnover, build stable technical staffs, and attract top-calibre experienced people, compensation plans are becoming more attractive.

Tuition reimbursement and comprehensive medical and profit-sharing plans, along with flexible work hours, have become standard. Stock plans, company cars, club memberships, bonuses, and sabbaticals are becoming popular rewards as salary increases among employees on similar experience-levels have become inadequate. As a way of narrowing the gap between the technical side of high-tech companies and the management side, the "dual-ladder" form of compensation has become more common.

"If you're a strong technical company, by definition you have to provide a career track for technical people. When technical people see that a company is structured totally toward the management side, they get frustrated and leave," says Bill Crean, vice president of personnel at Biosciences, a subsidiary of American Hospital Supply. "We're working very hard—as are other high-tech companies who are looking at the labor market realistically—to develop a dual career ladder to reward top technical performers to make sure that they have the appropriate career step to make also."

Another area of concern among advanced-technology companies is "salary compression," where inflating starting salaries match or exceed salaries of professionals already on-staff. "It may be too early to say that stabilized new graduate salaries are going to completely solve the compression problem," says Griff Baily, chief of personnel staffing at Martin Marietta Aerospace in Denver. "A new problem we're seeing is with hiring experienced people. There's so much competition for those good experienced people that we're paying higher salaries than we had expected. We're looking to bring 2000 people on board, and it's my view that those people are already employed. Higher salaries for experienced people may be a new source of compression. We've already made adjustments internally. And we're going to be looking at it very carefully, because compression is a major cause of turnover attrition."

Relocation reimbursement is one compensation area from which some companies are retreating. A Silicon Valley personnel director says, "The cost of housing has skyrocketed and so has the cost of moving people. Unless we need to find someone with very specialized skills who's impossible to get locally, we are not going to be relocating from another part of the country. We'd rather pay extra salary to someone already in our area." A challenge to companies in the newer advanced technology centers is convincing experienced people that they will, if they relocate, still remain in the mainstream of "new thinking" in their specialties.

"Obviously it's not as easy for an experienced person in this area to walk next door and start doing the same job with some other company as they can in an area like Route 128," says Tim Murray, supervisor of employment and employee relations at Hayes Microcomputer Products in Atlanta. "But when we are recruiting outside our area, we tell people about what has happened to Atlanta—especially in the last two years—in high technology growth. And there are a lot of exciting things going on here."

The Metropolitan Atlanta Council for Economic Development supports the efforts of firms like Hayes. Says Roy Cooper, director of economic development and research for the Atlanta Chamber of Commerce, "We've built a nucleus of over 150 different manufacturing firms in advanced technology. Governor Harris has established an Advanced Technology Development Center at Georgia Tech which already has scores of companies involved. And Tech itself is the fifth-largest producer of bachelor's-degree engineers in the nation.

"Mr. Murray's correct about the excitement here."

Grumman Aerospace is quite prepared to underwrite relocation expenses to get the experienced people it needs. "We hold 'open house' sessions a half-dozen times a year here in Bethpage, Long Island," says Grumman's Pete Portanova. "And we've had good results from 'job fairs' around the country, where local people arrange attendance, and we and others meet interested prospective recruits. The kinds of people we want are top people, and we visit the 'fairs' prepared to make an offer on the spot.

"There's a unique sort of family-type atmosphere at Grumman, which may explain why we experience only about 5% turnover among the 6,000-odd scientists and engineers among our 30,000 employees. But low turnover or not, we still needed about 300 people last year for work and research in advanced technology," Portanova observes. "We got them."

While new advanced technology centers throughout the country are giving scientists and engineers greater choice, overseas opportunities are becoming fewer. Arab projects are being scaled back. Companies with overseas operations are more inclined to recruit foreign nationals than to export Americans for extended stays.

"It becomes expensive to do business overseas when you bring someone over from the States and pay the high salaries and allowances people expect there," says Computer Sciences' Sharpe. "More important, however, career growth may be interrupted with a two- or three-year stay overseas. You'd like to tell people not to worry about their career paths when they return. But that promise can rarely be made."

U.S.-based recruitment for overseas technical jobs has changed, too. Engineers are no longer as motivated as they were by tax-free assignments with paid vacations and free trips home. "Now we attract them by offering opportunities to do truly advanced work," says a recruiter at Boyden Associates. "Water management promises to be a major engineering endeavor in the American southwest; we've recruited engineers for Middle Eastern assignments on the basis that they'll be obtaining invaluable experience there for use, in a few years, back here."

MANAGING (AND SELLING) ADVANCED TECHNOLOGY

In a 1982 speech, Texaco chairman and chief executive officer John McKinley described a human resource problem increasingly faced in his industry: "A more effective relationship between businessmen and technical people is necessary... in the future, the successful business manager must have a broad grasp and understanding of scientific and technical aspects and their impact on society."

© 1984 SCIENTIFIC AMERICAN, INC

Systems Engineers
Software Architects and Developers

Freedom to make a difference

AT&T Bell Laboratories has openings in the following areas of expertise:

- ☐ Communications planning
- ☐ Communications protocol
- ☐ Computer networking
- ☐ C^3
- ☐ Database management
- ☐ Digital services
- ☐ Distributed signal processing systems
- ☐ Local area networks
- ☐ Market definition for new technologies
- ☐ Measurement systems
- ☐ Network optimization tools
- ☐ Network planning
- ☐ Operations automation
- ☐ Packet switching
- ☐ Performance measurement and analysis
- ☐ Software development tools
- ☐ Subjective testing
- ☐ UNIX™ Operating System
- ☐ UNIX System applications

These openings are at our facilities in New Jersey, Illinois, Ohio, Pennsylvania and Massachusetts.

Having an impact, doing important work, creating something significant –even memorable–there's no feeling in the world quite like it.

We offer you an excellent opportunity to make this kind of technical contribution...

Because of the variety and interdisciplinary nature of our work.

Because of the considerable resources placed at your disposal.

Because of the caliber of the people around you.

Because of the freedom to pursue your own approaches to problems.

And because we are now free to take the knowledge behind the most extensive and sophisticated communications network in the world and apply it in new ways– to new markets.

With the development of new and expanding projects, we are seeking people who can evaluate alternative architectures, perform competitive analyses, assess service viability, define, implement and measure systems or develop strategies to get products to market.

If you have a Master's or PhD in Systems Engineering, Computer Science, Electrical Engineering, Mathematics, Physics, Statistics or Operations Research with hardware and/or software experience–or a BS in these fields with a minimum of 5 years highly relevant experience –and are looking for a chance to make a difference, please send your resume to:

Director, Technical Employment,
AT&T Bell Laboratories,
Dept. 340/4909/84
150 John F. Kennedy Parkway,
Short Hills, New Jersey 07078.

An equal opportunity employer

AT&T
Bell Laboratories

© 1984 SCIENTIFIC AMERICAN, INC

 CAREERS IN ADVANCED TECHNOLOGY: 1984

Who will these new managers and marketers be? Most experts agree they will be people who were first technically-trained. The personnel director of a large telecommunications company says: "It's much easier to introduce a technical person into the managerial or marketing side of the business than it is to try to acquaint a strictly business-oriented person with the nuances of technology. In sales, for example, the buyer is often a technical person, and it takes another technical person to be able to talk the same language."

For scientist/engineers who look to managerial and marketing opportunities, the outlook is bright. As a result of the increased call for technical managers and marketers, more universities offer degree programs and training sessions that address the special challenges of technical management. From only a handful of schools in the late '60s offering degree programs that blended technical and management courses, about 100 universities world-wide now do so. Carnegie-Mellon University's post-college professional education group offers a three-week "Program for Technical Managers;" its twentieth session is currently under way in Pittsburgh, and a one-week seminar, "Managing Technology," convenes in November.

One pioneer in technical management education is Dr. Bernard Sarchet of the University of Missouri-Rolla, who founded the school's Department of Engineering Management in 1967. "The failure of the traditional technical education to provide graduates with skills to handle other people and to deal with personnel matters has not changed since 1967," he says. "On top of that, technology is changing so rapidly that engineering schools are reluctant to take time away from the direct thrust of technical training."

Sarchet estimates that there are about 3,000 people with technical management degrees; a typical master's degree candidate has worked as a scientist or engineer for about ten years. A recent survey of UM-R graduates showed that both B.S. and M.S. graduates of the engineering management program doubled their salaries in six years. A typical comment: "(The curriculum) allowed me to speak fluently the different languages of the engineer and the general manager and to communicate well with either group."

Sarchet predicts that the next five to ten years will see an explosion of management training for technical people. "Industry is recognizing the need for their engineers to have some tools of management as they move up the ladder," he says.

THE CLASS OF '84: OPTIMISM AND STREET-SENSE

The final hiring results won't be in until summer, but the mood is 'upbeat' among seniors at Syracuse University's L.C. Smith College of Engineering.

"Some students are nervous, but not like last year. That was the worst," says Mark Simmons, Syracuse senior with a dual major in electrical engineering and bioengineering. Simmons had one job offer, as he started the year, from a medical instrument company where he had interned for two summers. He feels that work experience has given him an edge.

Linda Supon, assistant director of SU's placement center, agrees. "Companies have become much more selective over the past few years. They're looking at work experience and communication skills," she says. "Students know this. We've seen an increase in summer jobs, co-op programs, more counselling and job-search advice."

Although technical students are taking much less for granted in career planning, Dr. Brad Strait, Dean of the College, thinks they should do more. "Too many undergraduates don't think about employment until the second semester of senior year and maybe April at that," he observes. "Companies want good people and I worry that some student will miss the chance of a lifetime simply because he didn't 'have his act together.'

"We try hard to get students to the placement center early," Strait adds. He points to the College's internship and co-op programs, which begin at sophomore year, and to a junior-level course on "Engineering Ethics and Professional Responsibility."

"The Syracuse students I've interviewed this year have been practical-minded, articulate, and very much aware of the job-search process," says Hollis L. Holland, manager of professional recruitment at Eastman Kodak Company. Holland attended an "Engineering Career Weekend" at Syracuse in October 1983; "Dean Strait really pushes his upperclassmen to get out and see us."

Computer engineering major Matt Dewey uses SU's programs. By January he'd had 14 interviews and several invitations for follow-up meetings.

"I've got good grades and I'll have a job before the last minute. It won't be at one of those $30,000 or $40,000 starting salaries I used to read about. That was a fantasy," Dewey says. "But I didn't go into computer engineering for the money. I went into it because I like it. I just hope the job I get will be interesting and allow me to continue my education."

Kodak's Holland offers a word of advice to would-be recruits, experienced or not. "Talk to me in terms of value added to my company," he suggests. "Tell me how things are going to be different and better after you start working for us."

SMALL COMPANY OR LARGE?

For many engineers and scientists the prime question when thinking about a career choice is not what part of the country is most desirable or what the overseas opportunities are, but rather whether it is more desirable to work for a small, early-stage advanced technology company or a large, established one.

"To be part of a beginning is exciting. Ten years from now it will be a nice feeling to know that you were there at the start," says Jonnie Woodward, personnel manager at LSI Logic in California. Founded in 1981 with four people, LSI Logic is a high-tech success story. It now has over 350 employees, and plans to expand to nearly 1000 employees by the end of this year.

"Even though we've already gone public, there's till a 'start-up' feeling here," says Ms. Woodward. "Employees participate in the company's profits and they have the sense of direct responsibility for doing the best job possible."

The excitement of a ground-floor opportunity with potentially high financial reward makes some small companies attractive. Job security is a major appeal of larger companies, together with the possibility of more varied work-opportunities.

"We look for people who have skills and talents to grow over time, to change and to make different kinds of contributions over time," says Harry Portwood, corporate staffing and affirmative action manager at Hewlett-Packard.

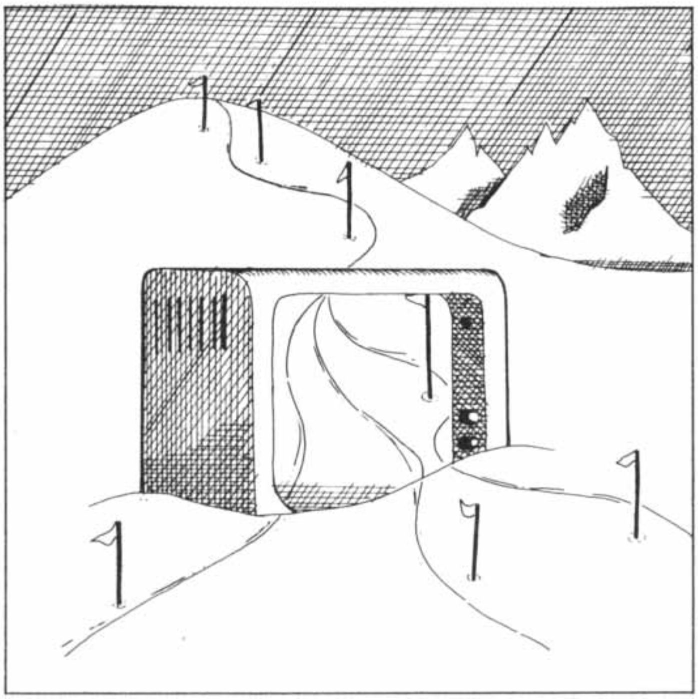
Says ITT Dialcom's Paul Warren, "A year ago, when we were a small firm making large gains in the electronic mail/database access area, we recruited pretty much locally. When we became a unit of ITT, we were able to draw on the resources of our parent company. And since we've doubled in size, to 200 people, in one year, we needed that help."

As both small and large high-tech companies increase their recruiting this year, there will no doubt be new and creative methods for attracting and keeping good technical people. Successful high-tech companies like Hewlett-Packard—which has 53,000 U.S. employees and an attrition rate of only about 5%—perhaps already know the secret.

"People are our most important resource," says Hewlett-Packard's Harry Portwood. "And that feeling is carried throughout every level of the company. It's what makes the company successful both in our business and in our recruiting of new people."

"CAREERS IN ADVANCED TECHNOLOGY: 1984" WAS WRITTEN BY BERNARD M. HODES AND RESEARCHED BY DON CHRISTIANSEN OF BERNARD HODES ADVERTISING, NYC. GRAPHICS: SHERIN & MATEJKA © 1984 SCIENTIFIC AMERICAN, INC COURTESY CHAMPION INT'L CORP. (ARTIST: DAVID WILCOX).

CAN YOU FIND THE SLALOM?



An Equal Opportunity Employer

You are skiing. You come to a point where three slalom courses branch off: A, B, and C.

To tell skiers whether the slopes are safe, there is a computer there with three data banks. But you know that saboteurs have rigged the computer so that one data bank always tells the truth, one always lies, and one alternates between truth and falsehood.

The three data banks say:

#1: At most, 2 of the slopes are safe.

#2: Slope B is dangerous.

#3: Slope B or Slope C is dangerous.

Which slope, if any, can you take?

This puzzle is part of our continuing effort to attract problem-solvers. We're United Technologies, a worldwide company that's ambitious, innovative and growing. We operate on the leading edge of technologies — such as microelectronics, aerodynamics, lasers, fiber optics, composite materials, and metallurgy. During the remainder of this decade, we plan to spend an average of $1 billion annually on R&D at our Research Center, Microelectronics Center and operating units to keep us out in front. Our operations include Pratt & Whitney, Otis, Carrier, Norden, Mostek, Sikorsky, Building Systems, Lexar, Inmont, Hamilton Standard, Essex, Elliot, and Automotive Group.

If you're an engineer or scientist interested in solving problems and growing with us, please write to Mr. Ray Lambert, United Technologies, Hartford, CT 06101. He'll send you more information about United Technologies — and, if you like, he'll also let you know whether your answer to our puzzle is correct.

UNITED TECHNOLOGIES

GGK

© 1984 SCIENTIFIC AMERICAN, INC

CHOOSE ANY KIND OF COMMUNICATIONS PROJECT IN THE WORLD.

YOU'LL FIND IT AT CSD.

From extreme environments on earth to remote areas of space — wherever the greatest challenges of modern communications are found, you'll find CSD. Our extensive resources, unmatched experience and total involvement in every area of communications technology makes CSD the pacesetter of the industry. If you can solve difficult communications problems by incorporating divergent technologies and innovations into a complete system design, join CSD on the threshold of new discoveries.

Because of our depth, leadership and emphasis on individual achievement, we can create a position specially suited to your interests and abilities. You'll be part of the stimulating atmosphere which sets CSD apart, where contributions and professional goals are recognized and rewarded with unequaled opportunity. Plus you'll enjoy the personal satisfaction of doing your best with the best in the industry. Wherever your career goals lie, you'll find the room to accomplish them at CSD — because we have the future for you.

TRANSMISSION SYSTEMS

We're engineering integrated radio systems using digital wide band processing. New technologies include unique fiber optic applications and systems operating in the VHF to EHF range.

MILSATCOM

Our emergent-technology EHF band systems involve AJ protection, rapid configuration and interoperability with satellite networks. Key systems elements include EHF antennas, feeds and high performance synthesizers.

DIGITAL PROCESSING SYSTEMS

These high performance systems cover the range of audio, ELF, VLF, LF and HF. We're developing algorithms with critical timing and throughput requirements, as well as applying high speed data processing in advanced digital communications products that include digital modems, voice and receivers, and systems controllers.

NETWORKING

We're pioneering totally integrated multinetwork interface systems pushing the state-of-the-art. Our extensive involvement in networking includes new concepts in such areas as message, packet and digital switching.

SECURE SYSTEMS

The security/intelligence community looks to us for leadership in systems knowledge. We're providing custom LSI/VLSI, dedicated secure communications networks, fault tolerant computer architectures and computer network security.

SURVIVABLE and ENDURING SYSTEMS

With our primary focus on C^3I propagation and network analysis under benign and stressed environments, we're expanding modeling and assessment software tools to determine vulnerability, survivability, endurability and/or exploitation of C^3I systems worldwide.

Along with a network of project opportunities offering new growth comes a strong system of support and rewards. Success here is recognized within GTE and across the communications industry. Internal advancement can take the direction of management or development of technical knowledge. Our educational opportunities are unusually sophisticated and include an in-house MASTERS DEGREE program in Systems Engineering and Computer Science, and complete tuition reimbursement. Our benefits include life, health and dental insurance coverage for dependents. And because we provide year after year of challenge, CSD has kept top quality professionals and maintained low turnover in a fiercely competitive market for engineers. For access to the important opportunities in communications, we can be the last place you'll ever have to look. U.S. Citizenship is required.

If you are interested in any of the above positions, send your resume, indicating areas of interest, to Aida T. Chaves, GTE's Sylvania Systems Group, Communication Systems Division, 77 "A" Street, Needham, MA 02194. Or call us at 1-800-225-3956.

An Equal Opportunity Employer M/F

IF IT'S HAPPENING IN COMMUNICATIONS, IT'S HAPPENING AT CSD.

Durham NC, Arlington VA, Colorado Springs CO, Rockville MD, Natick & Needham MA

GTE Systems

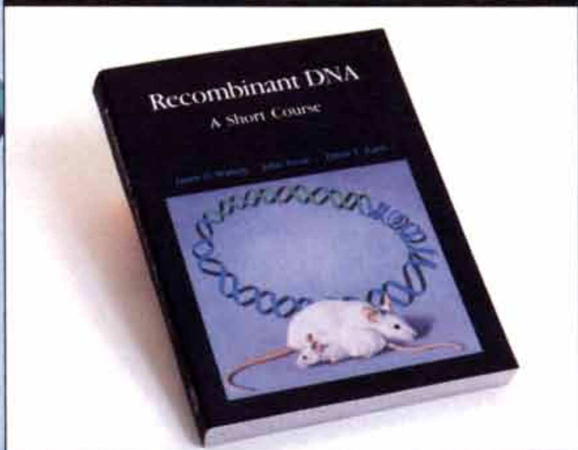
© 1984 SCIENTIFIC AMERICAN, INC

"An excellent, even extraordinary book"*

An up-to-the-minute overview of the state of recombinant DNA genetics, this eagerly awaited book summarizes the history of the recombinant DNA revolution. Nobel Laureate James Watson and his colleagues survey the recent techniques in recombinant molecular genetics, explaining both the experimental methods now in use and their dramatic results.

Enhanced by 174 exquisitely detailed illustrations, RECOMBINANT DNA: A SHORT COURSE explores such vital topics as tumor viruses, movable genes, the experimentally controlled introduction of DNA into yeast cells, and the genetic engineering of plants. Final chapters are devoted to viral vectors, recombinant DNA and genetic disease, and the science used in the recombinant DNA industry.

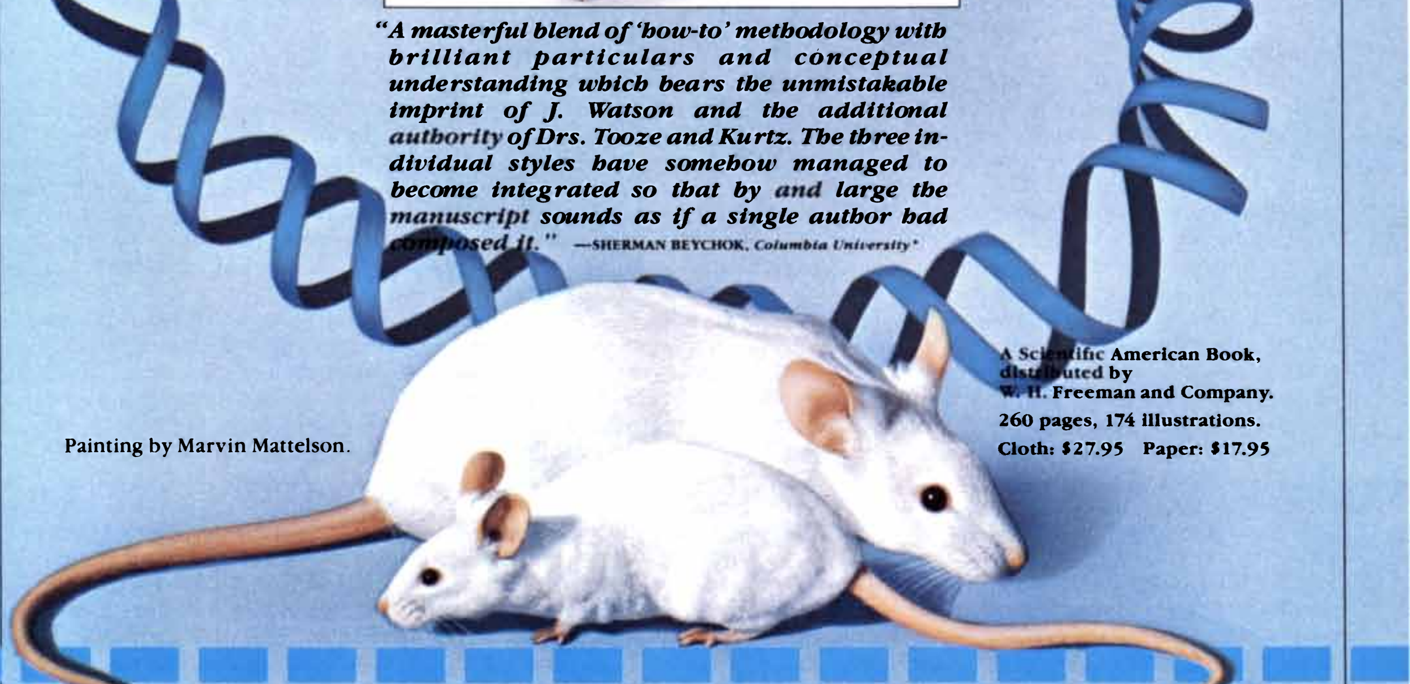
RECOMBINANT DNA: A SHORT COURSE



James D. Watson,
Cold Spring Harbor Laboratory

John Tooze,
European Molecular Biology Organization

David T. Kurtz,
Cold Spring Harbor Laboratory



"A masterful blend of 'how-to' methodology with brilliant particulars and conceptual understanding which bears the unmistakable imprint of J. Watson and the additional authority of Drs. Tooze and Kurtz. The three individual styles have somehow managed to become integrated so that by and large the manuscript sounds as if a single author had composed it." —SHERMAN BEYCHOK, *Columbia University**

Painting by Marvin Mattelson.

A Scientific American Book, distributed by W. H. Freeman and Company.
260 pages, 174 illustrations.
Cloth: $27.95 Paper: $17.95

Please send me ____ copy (copies) of RECOMBINANT DNA: A SHORT COURSE at ________ $27.95 (cloth: ISBN 1483-3)
________ $17.95 (paper: ISBN 1484-1)

☐ I enclose a check made payable to W. H. Freeman and Company (please add $1.50 for postage and handling; New York, California, and Utah residents include appropriate sales tax). Charge my ☐ MasterCard ☐ VISA.

Account # ____________ Expires ___ / ___

Signature ____________
(Credit card orders must be signed.)

Name ____________

Address ____________

City ____________ State/Zip ____________

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

0112

© 1984 SCIENTIFIC AMERICAN, INC

The Descent of Hominoids and Hominids

Findings over the past five years have made it possible to trace with greater assurance the divergence of the apes from the Old World monkeys and the later divergence of humans from apes

by David Pilbeam

Human understanding of human evolution has advanced greatly in the past generation and even in the past five years. Since the 1960's, for example, paleoanthropology has ceased being a discipline of concern only to physical anthropologists and a few interested paleontologists. It has evolved into a broad multidisciplinary pursuit that enlists the services of historical geologists, ecologists, molecular biologists, zoologists, students of animal behavior and even chemists and physicists.

In the Linnaean classification of living things human beings belong to the Order Primates, which also includes the apes, the monkeys and prosimians such as the lemurs. Within that order is the superfamily Hominoidea, including only humans and the apes. Within that superfamily, in turn, is the family Hominidae, including only humans and their extinct relatives closer than the apes.

Let me briefly review how the evolution of hominoids and hominids was understood some five years ago by a leading physical anthropologist, Sherwood L. Washburn of the University of California at Berkeley, and contrast that summary of the consensus then with the consensus now [see "The Evolution of Man," by Sherwood L. Washburn; SCIENTIFIC AMERICAN, September, 1978]. At that time revolutionary advances in geology and molecular biology had already challenged earlier ideas about the geography and timing of hominid evolution and the biological relations between humans and the other higher primates. For example, data from molecular biology made the reconstruction of primate descent a simpler task. Many physical anthropologists were skeptical, however, of the molecular biologists' suggestion that African apes (chimpanzees and gorillas) were closely related to humans but Asian apes (orangutans and gibbons) were not. The biggest surprise was the molecular biologists' proposal that humans had diverged from the African line of apes much less than 10 million years ago.

Washburn's summary also suggested that *Ramapithecus*, a Miocene hominoid widely distributed in the Old World, could have been the hominoid ancestral to all living great apes and to humankind. Further, the discovery in Ethiopia of *Australopithecus afarensis* (widely known from the partial skeleton called "Lucy") and in Tanzania of the famous Laetoli footprints showed that by more than 3.5 million years ago there had evolved in Africa primitive hominids: small-brained, ground-dwelling primates that walked on two legs and had canine teeth that were relatively small and did not project. The discoveries in Ethiopia and Tanzania showed bipedalism was the first and most basic of hominid evolutionary adaptations. By analogy with chimpanzees, these early hominids probably used wood and other naturally occurring materials as tools. Indeed, the contrast between the hominids' small canines and the projecting canines of the chimpanzee suggested that among the hominids the offensive role of canines had already been taken over by tools.

By the time another million years had passed, in Washburn's summary of the consensus, hominids with larger brains had evolved in Africa. They made stone tools and lived alongside a robust (that is, relatively heavy-boned and large-toothed) species of *Australopithecus* that became extinct in about another million years. Even before then, say by 1.5 million years ago, a more advanced hominid had appeared: *Homo erectus*. The association of its stone tools with animal bones was accepted as clear evidence that hunting behavior had become a hominid attribute long before the hominid brain expanded to its modern size.

Brain size and tool complexity increased only slowly between then and about 100,000 years ago, when a still more advanced hominid arose: *Homo sapiens*. These increases, however, are indirect evidence for the reorganization of the hominid brain. With the arrival of *H. sapiens* the rate of change in stone-tool types began to accelerate. The anatomical evolutionary changes are seen mainly in the skull. Then, about 40,000 years ago, anatomically modern humans appeared and spread rapidly throughout the Old World. At the same time the rate of behavioral change, as attested by changes in stone tools and other human artifacts, approached what we now expect from "normal" human cultural behavior. Washburn closed his discussion with the suggestion that fully modern human language, which possibly did not exist before the appearance of anatomically modern man, catalyzed this behavioral advance, although no direct anatomical evidence bears on the point.

In the five years since Washburn's summary important new fossil discoveries have been made, not all of them in the field. Fuller descriptions of earlier fossils have been published. Paleoanthropologists have become better able to interpret their data through the work of behavioral scientists, and they have become increasingly aware of relevant developments in the other natural sciences. These trends have shifted the discipline toward a greater integration with the rest of the biological community.

At the same time there is increasing doubt about the degree to which narratives of human evolution can be "brought to life." Just how detailed and reliable can such accounts be? Hard problems arise when the hominids of the past are interpreted in terms of the hominids of the present. The early hominids were, after all, markedly different from any living today. In many instances, however, these differences have been ignored and early hominids have been made to seem too much like modern human beings. Let me review some of

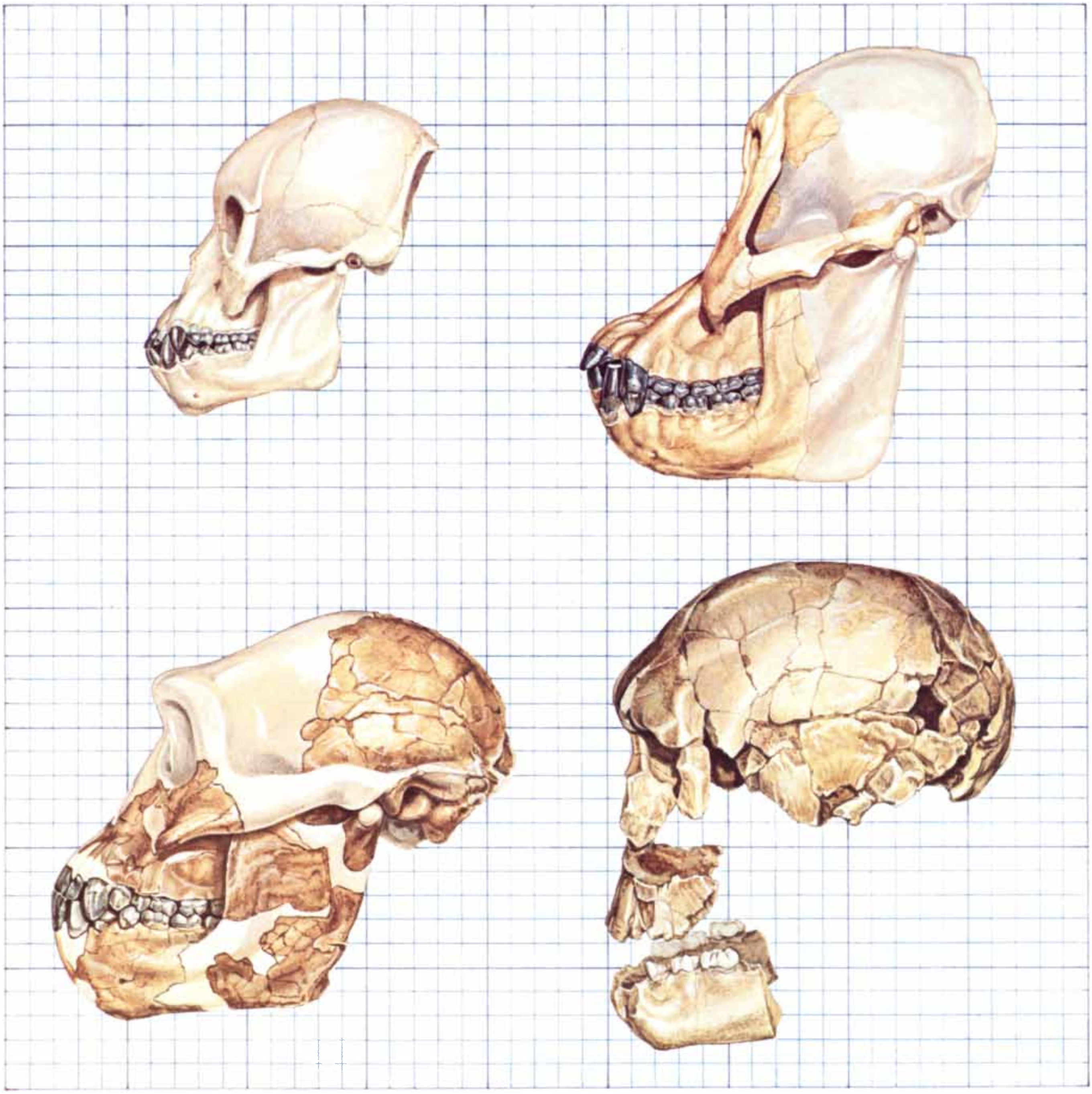
© 1984 SCIENTIFIC AMERICAN, INC

these problems, starting at the beginning with the evolution of hominoids, the group from which the hominids sprang.

Hominoid Evolution

The earlier debate between physical anthropologists and molecular biologists over the pattern and timing of hominoid evolution is now basically settled. Most paleontologists (and comparative anatomists) agree the molecular patterns showing the African apes are genetically very little different from humans and the Asian apes are about twice as different reflect the fact that the common ancestor hominids share with the chimpanzee and the gorilla was in existence only about half as long ago as the last common ancestor of all the larger hominoids.

Detailed comparative anatomy shows that the patterns of morphological resemblance among living hominoids confirm the molecular findings. The great apes and the hominids do not form what is called a monophyletic group, that is, they did not share a last common ancestor after the lineage leading to humans diverged. Instead African apes and humans are monophyletic with respect to the Asian apes. (That had in fact been proposed by certain earlier anatomists.)

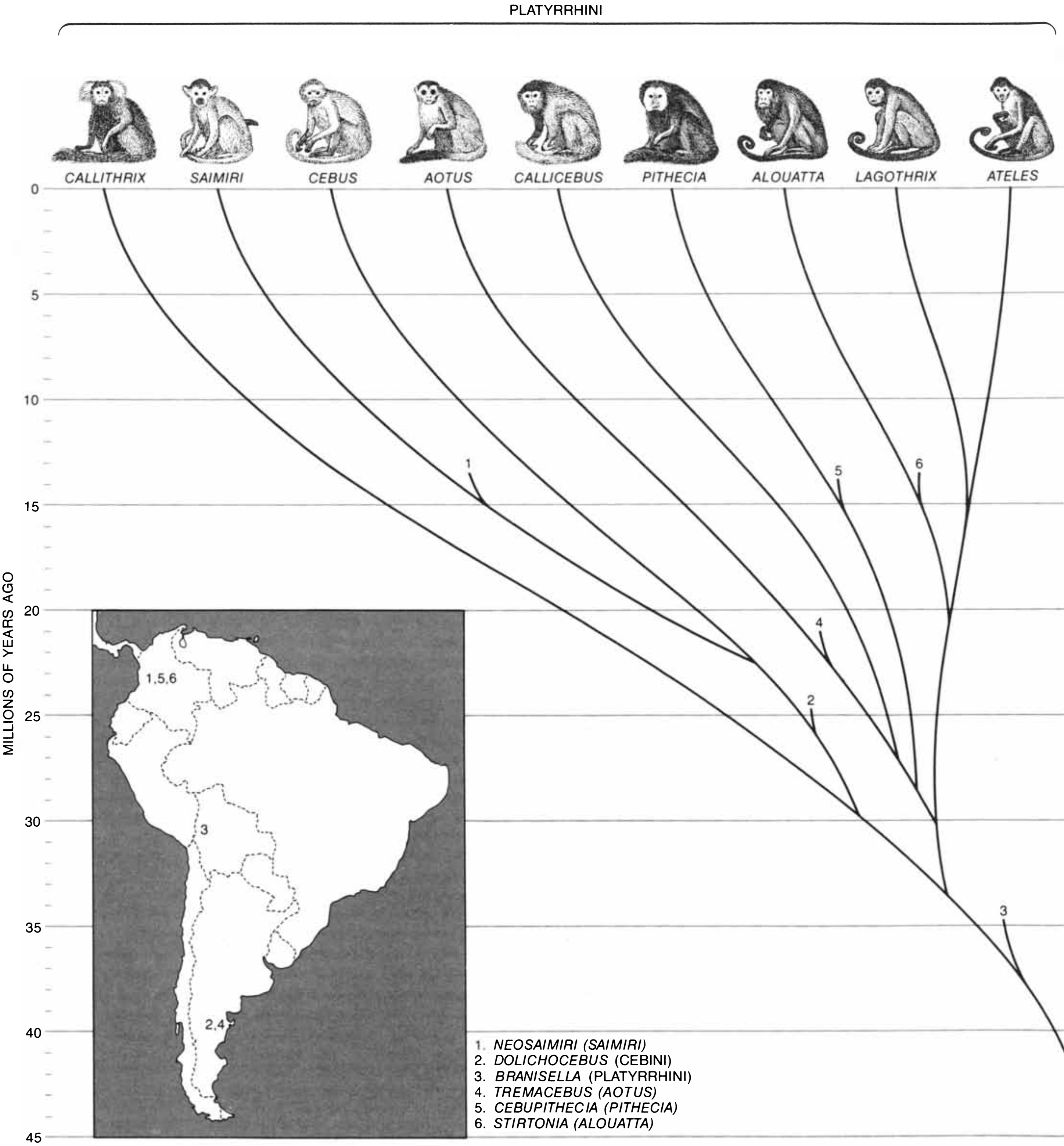


FOUR FOSSIL SKULLS, shown in profile in this painting, exemplify key stages in the branching off of hominoids from ancestral Old World monkey stock and the later branching off of hominids from ancestral hominoid stock. At the top left is the best-known of the early Miocene apes, *Proconsul africanus*, a baboon-size primate that is perhaps similar to the common ancestor of all apes and humans. At the top right is the substantially larger late Miocene ape *Sivapithecus*, first known from fossil teeth and jaw fragments found in foothills of the Himalayas in the 1900's. It may have been ancestral to the only living large hominoid of Asia, the orangutan. At the bottom left is the earliest-known African Pliocene hominid, *Australopithecus afarensis*, a species that walked erect in the period between four and two million years ago. At the bottom right is the much larger-brained hominid *Homo habilis*, which was first found in Kenya in 1972. Members of the species ranged southern and eastern Africa for a few hundred thousand years beginning about two million years ago.

© 1984 SCIENTIFIC AMERICAN, INC

This means features common to all apes, such as the absence of a tail, must either have been retained from the last ancestor of all large hominoids or have evolved separately in the different apes. As a consequence of the molecular findings morphologists are reexamining hominoid anatomical characteristics and similarities and looking more closely at the processes of embryonic development and their genetic controls.

These "new" molecular branching patterns are not incompatible with the hominoid fossil record. The time is past when many paleontologists placed the origin of the hominids in the middle Miocene, 16 or 15 million years ago, or when some molecular biologists held that the entire hominoid radiation took place within the past 10 million years, with the hominids diverging only a brief

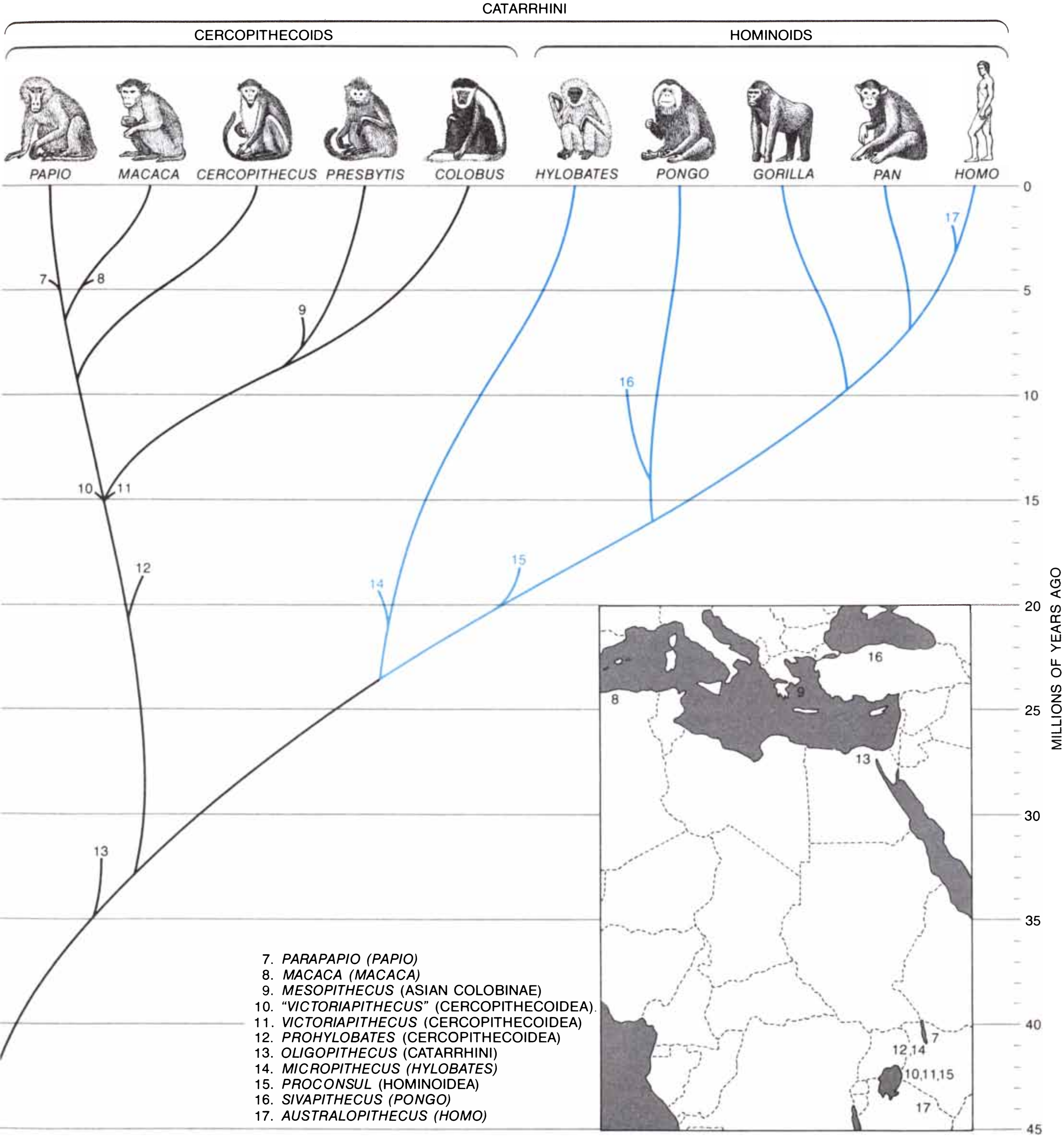


BRANCHING SEQUENCE of the higher primates from 45 million years ago to the present is shown under two major headings, the Platyrrhini, or the New World monkeys, at the left and the Catarrhini, or the Old World monkeys, at the right. The hominoid and hominid branches are in color at the far right. The sequences are based on such biochemical data as small differences in the amino acid sequence

© 1984 SCIENTIFIC AMERICAN, INC

four million years ago. A tentative consensus would now place the radiation of all the great apes from an ancestral stock in the middle Miocene, with the hominids splitting from an African ape lineage in the late Miocene, perhaps eight or seven million years ago.

Why was the hominoid fossil record misinterpreted, at least by dimmer paleontologists such as me? There are a number of reasons. First, far too much attention was being paid to the fossil record as a source of information about evolutionary branching sequences. It is now clear that the molecular record can tell more about hominoid branching patterns than the fossil record does. (The fossil record remains, however, the only source of direct knowledge of past evolutionary states.) Misinterpretation also resulted from the fact that the hom-



of the same protein in different living species. The specific dating of the branches is based on the age of fossils and on other kinds of geochronological data. The maps at left and right indicate where the earliest fossil representative of the genus was found. The name to the right of each fossil genus indicates its relation to living forms. Thus *Sivapithecus* pairs with *Pongo* but *Proconsul* pairs with all hominoids.

© 1984 SCIENTIFIC AMERICAN, INC

inoid fossil record going back before the last four million years of hominid evolution has until recently been both fragmentary and poorly dated. It consisted almost entirely of teeth and small pieces of jaw, which are parts of the body that are seldom very informative on questions of taxonomy. For example, the equivalent parts of the body in living mammals are generally not of much help in sorting individuals into species or clustering species into higher taxonomic groups.

A third cause of misinterpretation is the very diversity of middle and late Miocene hominoids. The living large hominoids are not notably diverse: only five genera have survived. When all the fossil evidence of past hominoids is taken into account, what emerges is a group that is quite heterogeneous; each hominoid species is in many respects unique. These factors have made it hard to reconstruct by means of fossils the branching sequences of living hominoids. It is even harder to link particular fossil species to particular living ones.

Significant new Miocene hominoid fossils, however, have changed the earlier state of affairs. Skeletal parts—faces, brain cases and limbs—that were misidentified or have only now been uncovered are yielding a clearer and often surprising picture of both evolutionary relations and the nature of the early hominoids. One can even catch glimpses of how they may have lived and what their ecological relations were.

At the same time newer molecular analyses are clarifying and expanding ideas about the pattern and timing of hominoid evolution; an example is the work of Charles G. Sibley and Jon Ahlquist of Yale University, in which DNA's from different primate species are hybridized to establish the differences and similarities of their sequences of nucleotides. Slowly paleoanthropology is moving from the time when the fossil record contributed nothing to an understanding of phylogenetic patterns to a time when it is one component of a synthesis that renders data of all kinds compatible: fossil, anatomical, molecular and developmental.

The Hominoid Radiation

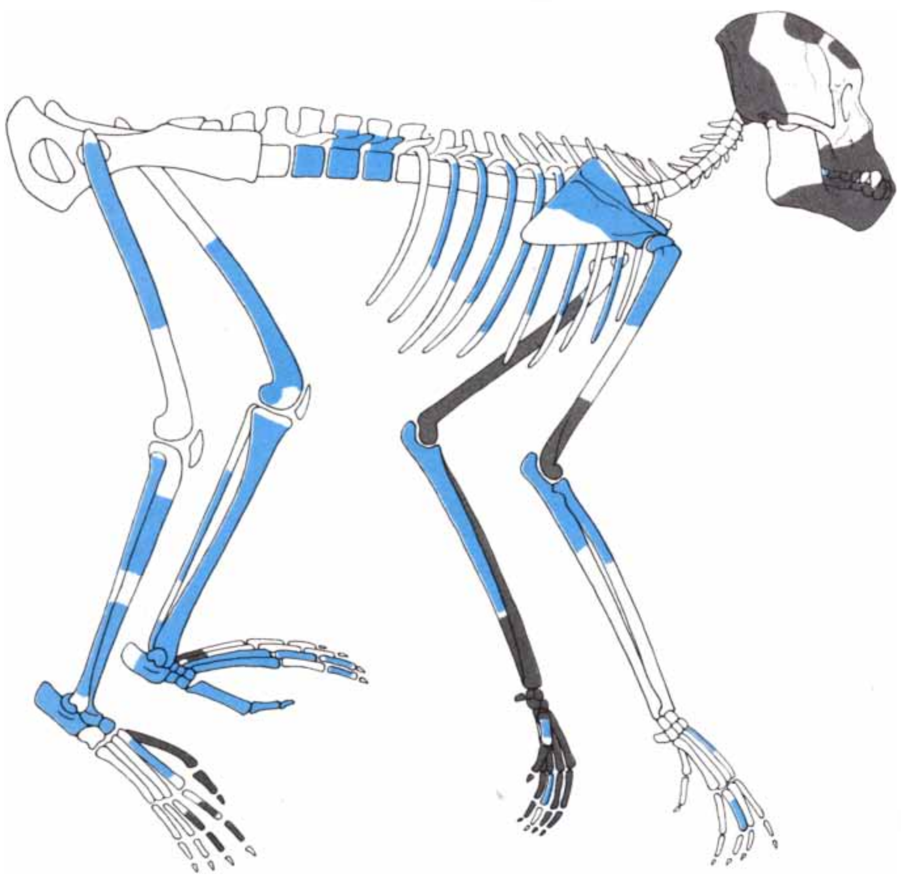
An account of the radiation of the hominoids that represents a consensus of current opinion can be summarized as follows. First, it is not clear when the great apes diverged from their relatives, the Old World monkeys. (The New World monkeys played no part in hominoid evolution.) It seems almost certain, however, that the divergence came after middle Oligocene times (some 30 million years ago) and before early Miocene times (some 20 million years ago). However that may be, the radiation of the hominoids was under way in the early or middle Miocene (20 to 15 million years ago) and certain lineages of this moderately successful group did not become extinct until the late Miocene (eight to six million years ago).

It is similarly not clear when the small hominoids, the gibbons of Asia, diverged from the larger ones. The split is possibly documented by some of the earlier Miocene primate fossils, although it is a controversial point. There is better documentation, however, for the divergence of the large apes into Asian and African species by middle Miocene times, say 16 million years ago. The orangutan is the odd and sole survivor of a group of middle and later Miocene Asian apes that were previously more diverse and widespread.

Sadly, for most of the Miocene the African hominoid branch is virtually without fossil representation. From the molecular record, however, one can infer a late Miocene splitting of first gorillas, then chimpanzees and hominids (between 10 and six million years ago), and it is known that bipedal hominids were present in eastern Africa by at least four to 3.5 million years ago. Until there is a better African fossil record (particularly a fossil record for the African apes) one can only guess about the timing and precise branching sequence of these lineages and about the nature of their early members, including the very earliest hominids.

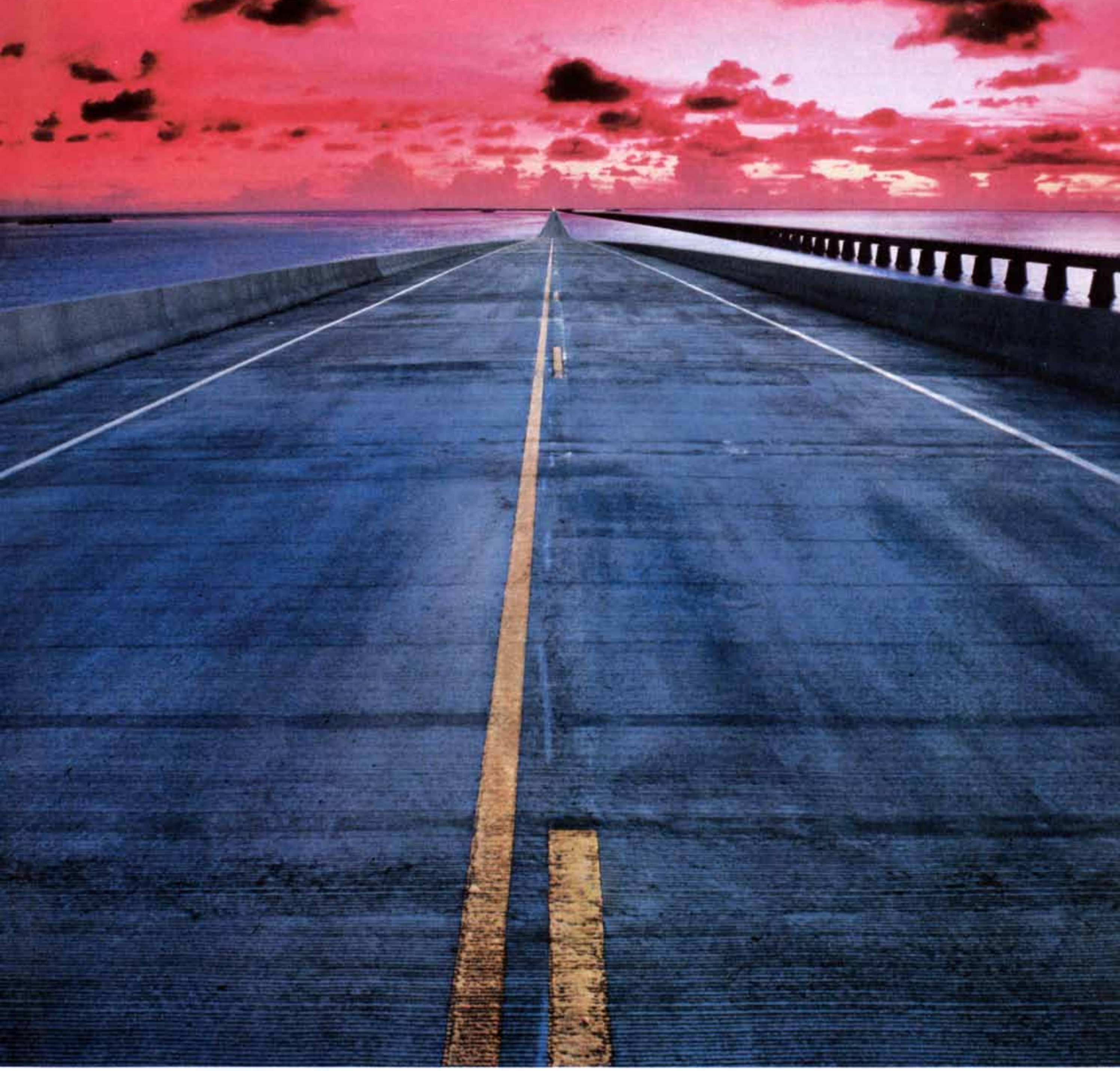
If one goes back to Oligocene times, the African fossil primate record is comparatively rich. Elwyn L. Simons of Duke University has been hunting successfully for primate fossils in the Fayum of Egypt on and off since 1961. Since 1977 he and his collaborators John G. Fleagle of the State University of New York at Stony Brook and Richard F. Kay of Duke have uncovered some truly remarkable material. Substantial parts of the skeleton of *Aegyptopithecus,* a species hitherto known only from a skull found earlier by Simons, and *Apidium* are now in hand. Even the very sparsely represented *Parapithecus* and *Propliopithecus* are now relatively well characterized.

These fossils represent a diverse group of unspecialized tree-living primates that had attained the biological status of monkeys at least as early as middle Oligocene times, 30 million



PARTS OF PROCONSUL found at the time of its initial discovery by Tom Whitworth in 1951 are shown in a shade of gray in this restoration of the four-footed hominoid. The additional parts of the skeleton shown in color are those recently found in museum collections by Alan Walker and Martin Pickford, who prepared this restoration. The parts shown in outline only, such as the pelvis, most of the spine and parts of the skull and jaw, have not been found.

© 1984 SCIENTIFIC AMERICAN, INC



Here's where we save taxpayers $6 million a mile.

The life of bridge deck surfaces in northern and seacoast areas can be as short as five years. Road de-icing salt and salt spray seep into the concrete and corrode the reinforcing bars inside. Corroded bars expand and rupture the roadway.

But with 3M "Scotchkote" Epoxy Coating protecting the bars, the problem is solved. And taxpayers can be spared the $6 million a mile it now costs to rebuild just a two-lane bridge.

We developed this solution for highway engineers. But the product also works on underground pipelines, in parking ramps and concrete buildings.

By listening to people's needs, we've pioneered over 700 products for the construction and maintenance industries. We now make everything from roofing materials and carpet care systems to fire barriers and electronic building monitors.

And it all began by listening.

3M hears you...

For your free brochure on 3M's construction and maintenance products and all other 3M products, write: Dept. 100703/3M, P.O. Box 22002, Robbinsdale, MN 55422.

Name ____________________

Address ____________________

City __________ State ______ Zip ______

Or call toll-free: **1-800-336-4327.**

©1984, 3M

3M

© 1984 SCIENTIFIC AMERICAN, INC

IBM PC 8087 SUPPORT FROM MicroWare

87FORTRAN/RTOS™ is a MicroWare adaptation of the Intel Fortran-86 compiler, a full ANSI-77 subset with 8087 extensions and overlays. It generates in line 8087 code allowing use of all 8087 data types, including 80 bit reals and 64 bit integers. The complete subset I/O is supported including Internal and External Files and List Directed I/O. 87FORTRAN/RTOS uses the Intel large memory model, allowing data/code structures which utilize the full megabyte. The compiler provides direct access to 8088 ports and supports logical operations on 8 and 16 bit operands normally treated in assembly language. It is ideal for large applications which are number intensive or control hardware.

87PASCAL/RTOS™ is Intel's ISO-Standard Pascal with 8087-8088 exceptions. These make it possible to use all the 8087 data types directly, while generating modules in one of the three Intel Memory Models.

87FORTRAN/RTOS and **87PASCAL/RTOS** include RTOS and support for one year............. each **$1350**

PC TECH JOURNAL REVIEW:
"The MicroWare package is preferable . . . it executes the basic operations more rapidly and MicroWare provides a free update service."

87BASIC™ includes patches to the IBM Basic Compiler and both runtime libraries for USER TRANSPARENT and COMPLETE 8087 support. Provides super fast performance for all numeric operations including trigonometrics, transcendentals, addition, subtraction, multiplication, and division............... **$150**

87MACRO™ - our complete 8087 software development package. It contains a "Pre-processor," source code for a set of 8087 macros, and a complete object library of commonly called numeric functions..... **$150**

RTOS DEVELOPMENT PACKAGE includes 87FORTRAN, 87PASCAL, PL/M-86, Utilities, TX Screen Editor, and RTOS................................. **$2500**

RTOS — Real Time Multi-Tasking/Multi-User Executive
RTOS is a MicroWare configured version of iRMX-86, Intel's legendary operating system. It includes the Intel Assembler, ASM-86, which supports the 8086, 8087, 8088 and 80186. All modules produced by the compilers or ASM-86 are linked, loaded and managed with LINK-86, LOC-86, LIB-86 and OH-86. These utilities support relocatable or absolute code, generate Intel Hexcode (when needed), and allow overlays. RTOS/ASM-86/LINK-86/LOC-86/LIB-86/OH-86 .... **$600**

iRMX-86, LINK-86, ASM-86, LOC-86, OH-86, LIB-86, Pascal-86, Fortran-86 and PL/M-86 are trademarks of Intel Corp.
IBM PC is a trademark of IBM Corp.

MICROWARE BEST SELLERS!

8087-3 CHIP in stock with 180 day warranty and 8088 exchange **$175**

64K RAM Upgrade **$64**

FASTPAK™+8087 includes one runtime library (87BASIC or 87MACRO), the 8087 chip, the 87/88GUIDE, and installation instructions.................................. **$325**

MATRIXPAK™ manages a **MEGABYTE!** Written in assembly language, our runtime package accurately manipulates large matrices at very fast speeds. Includes matrix inversion and the solution of simultaneous linear equations. Callable from MS Fortran 3.13, MS Pascal 3.13, and MicroWare 87MACRO, 87BASIC, and RTOS compilers.......................... each **$150**

MICROSOFT FORTRAN 3.13
MICROSOFT PASCAL 3.13 These new IEEE compatible compilers support both double precision and the 8087.... each **$259**

MICROSOFT C COMPILER includes Lattice C and the MS Librarian. .**$350**

Float 87 for MS C	$125
Multitool Word	269
Microsoft Business Basic Compiler	495
SuperSoft Fortran	340
SSS 8087 Support	50
SuperSoft Voice Drive	895
TRACE86 Utility	125
Computer Innovations C86	345
STSC APL★PLUS/PC	545
64K QUADRAM	349
SANDSTAR 64K Card	275
Sandstar WS2 Hard Disk System	1355
HALO Graphics	125
Graphmatic	95
Energraphics	225
Anderson-Bell Abstat	345
Draft-Aide	595

Micro Ware
P.O. Box 79
Kingston, MA
02364
(617) 746-7341

You Can Talk To Us!

HOW TO REACH THE PEOPLE WHO MAKE THE FUTURE HAPPEN IN SPAIN

INVESTIGACION Y CIENCIA is the Spanish edition of SCIENTIFIC AMERICAN which attracts a young audience of affluent professionals who have come to the top because of their technical expertise. Almost half of them hold top management job titles. The fortunes of major corporations in Spain increasingly depend on technically sophisticated people. INVESTIGACION Y CIENCIA keeps these people in touch with the advances in science that drive the growth of Spain's industry.

SCIENTIFIC AMERICAN speaks the languages of more than half the world's population. We are the one publication in the world today providing efficient coverage of technology-based management. We reach men and women in industry, whose qualification to make technical decisions places them in key positions in their country's government and industry.

SCIENTIFIC AMERICAN in eight languages: English, Spanish, Italian, French, German, Japanese, Chinese and now Russian has gathered in its audience three million people who make the future happen around the world.

For more information on our Spanish-language edition contact:

Elena Sanchez-Fabres
Prensa Scientifica, s. a.
Calabria 235-239
Barcelona 29, SPAIN
Telephone (011) 343-322-0551

or in New York
John Kirby,
V.P./Advertising Director
SCIENTIFIC AMERICAN
415 Madison Avenue
New York, New York 10017
Telephone 212-754-0262

© 1984 SCIENTIFIC AMERICAN, INC

years ago. They were not, however, primitive primates like the lemurs but bore a general resemblance to the living New World monkeys. They are nonetheless too primitive and too different from the living higher primates of the Old World for one to know whether they are representative of a time before the divergence of the hominoids from the Old World monkeys or a time after the divergence.

The early Miocene of eastern Africa is a happy exception to the fossil-poor record for Africa as a whole. Collecting over the past 50 years in Kenya and Uganda has yielded some 1,000 fossils of higher primates that are between 17 and 22 million years old and that almost certainly represent hominoid lineages of a period after their divergence from the Old World monkeys. Mainly through the work of Peter Andrews of the British Museum (Natural History) the paleoecology of the period is now well known. The landscape was a mixture of forest, open woodland and grassland, with the forest predominating. The forest and woodland were inhabited by at least six hominoid species. One of them, *Proconsul africanus,* has for some time been the best-known Miocene ape, and it has now become even better known.

A small animal the size of a baboon, *Proconsul* has played a central role in thinking about hominoid evolution. The species had been represented until recently only by an almost complete forelimb, a jaw, skull fragments and a few foot bones, all of one individual, an almost complete skull of another individual and other limb fragments, teeth and jaws. Then Alan C. Walker of Johns Hopkins University and Martin Pickford of the National Museums of Kenya discovered—in museum and laboratory collections—many more pieces of the "forelimb" individual, so that it is now represented by much of its skeleton.

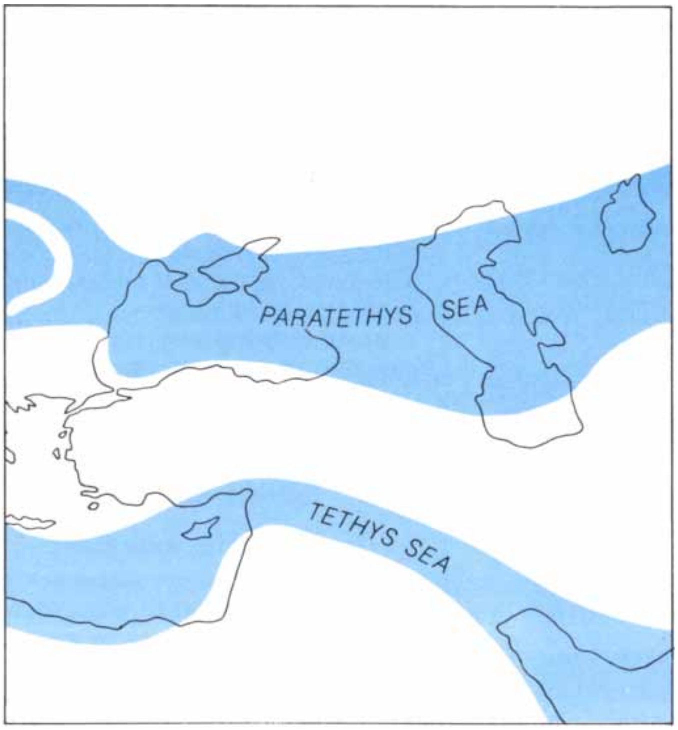
Until now *Proconsul* had been interpreted in several contradictory ways. At one extreme was the view held by some that its form was too specialized for it to be the ancestor of any living hominid and that it was possibly not even a hominoid. Others saw it as an unspecialized early hominoid and still others accepted it as being ancestral to the chimpanzee. The addition of the new material clarifies the issue to some extent. *Proconsul,* the consensus now holds, is a hominoid, albeit a very unspecialized and primitive one. It was a tree-dwelling, fruit-eating "formative ape" in which the sexes were distinctly different in size. In its total pattern it was unlike any living higher primate. Its elbow and shoulder joints and feet are like a chimpanzee's, its wrist is like a monkey's and its lumbar vertebrae are like a gibbon's. Many of its other features are unique, as is its total configuration. *Proconsul* provides a salutary lesson for students of evolution: the relations inferred for the animal have depended on what part of the body was being studied. When a fossil animal is found in fragments and over a period of time, the very order of discovery of its various parts will affect the phylogenetic interpretations, particularly in the case of a "mosaic" species such as *Proconsul.*

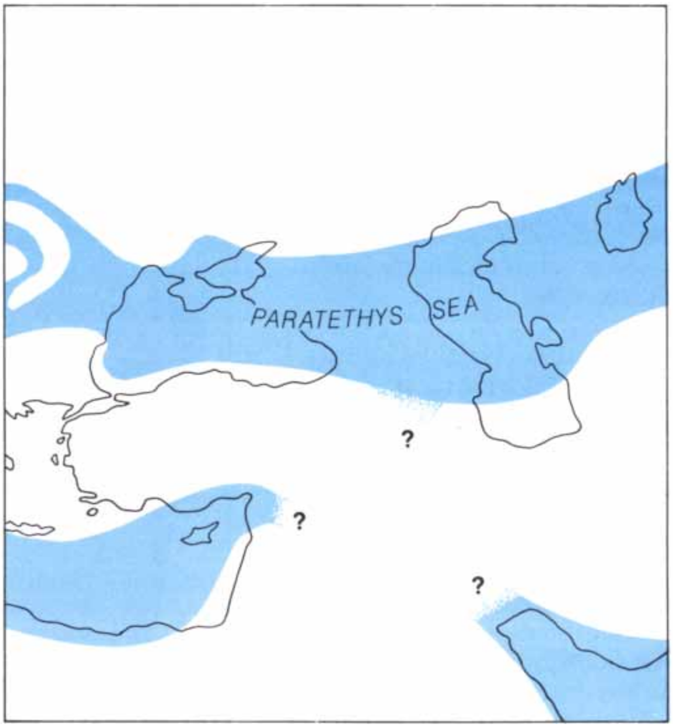
Dryomorphs and Ramamorphs

The end of the early Miocene, some 17 million years ago, was marked by a highly significant geologic event: continental drift linked Africa and Arabia with mainland Eurasia, thereby allowing the migration of African hominoids (and other mammals) into the rest of the Old World. The linkup initiated major mountain building: the Alps rose in Europe, the Taurus Mountains in Turkey and the Zagros Mountains in Iran. The Tethys Sea disappeared and oceanic and atmospheric circulation patterns shifted, bringing about large-scale changes in climate and habitats.

The Eurasian climate was seasonal in the middle and late Miocene, from 17 to five million years ago. The winters were milder, however, than they are today. Much of what is now grassland and even desert was then woodland, and the forests were more widespread. The mammalian fauna reflected these conditions: grassland grazers were scarce but woodland and forest browsers abundant.

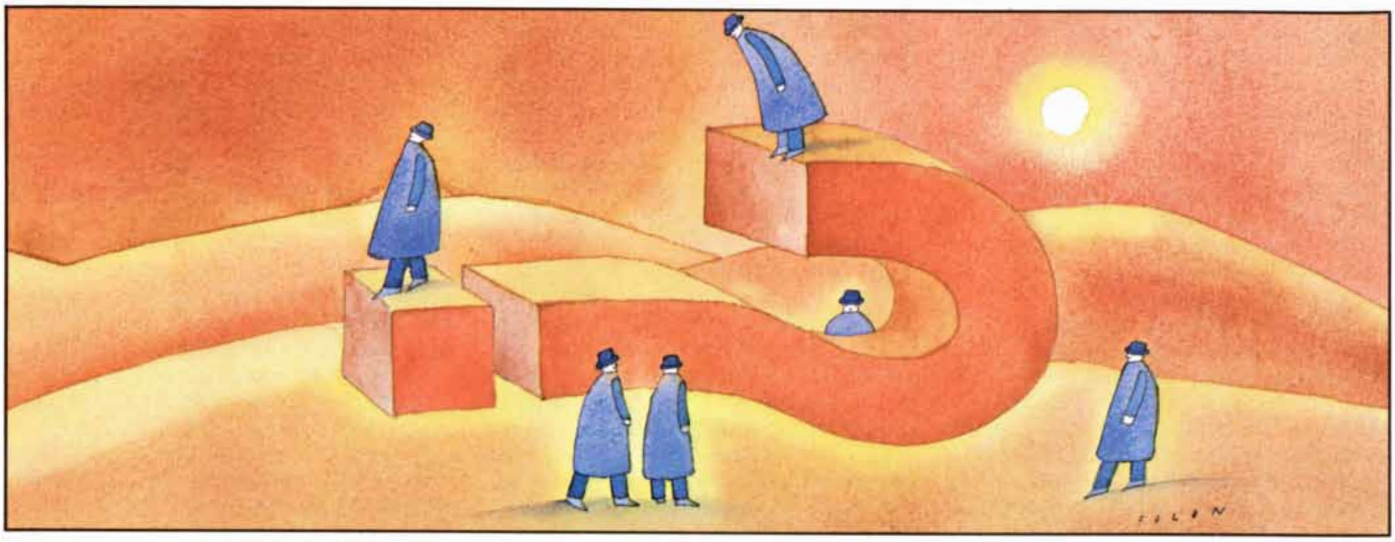
The hominoid fossil record for this period in Eurasia consists mainly of jaws and teeth, so that the phylogenetic relations are still not clear. As an approximate summary, however, the hom-





JOINING OF AFRICA AND EURASIA followed the time 20 million years ago when the Tethys Sea (*color*) separated the two landmasses, as is shown at the left. In early Miocene times, some 17 million years ago, the sea shrank, as is shown at the right, leaving a land bridge that allowed animals to migrate between Africa and Eurasia. Among migrants to Eurasia were some of the African hominoids.

© 1984 SCIENTIFIC AMERICAN, INC



NUCLEAR WASTE DISPOSAL
Scientists have an answer

Scientists and engineers agree that high-level waste from nuclear-electric power plants can be disposed of without harming people or the environment. They favor placing the sealed-up waste deep within natural geologic formations that have been stable for millions of years.

In nuclear plants, the uranium fuel pellets that are used to generate energy accumulate a small percentage of highly radioactive byproducts called high-level waste.

A solid majority of scientists have concluded that the best way to dispose of this waste is sealing it in long-lasting canisters, which are then placed in mined repositories deep underground.

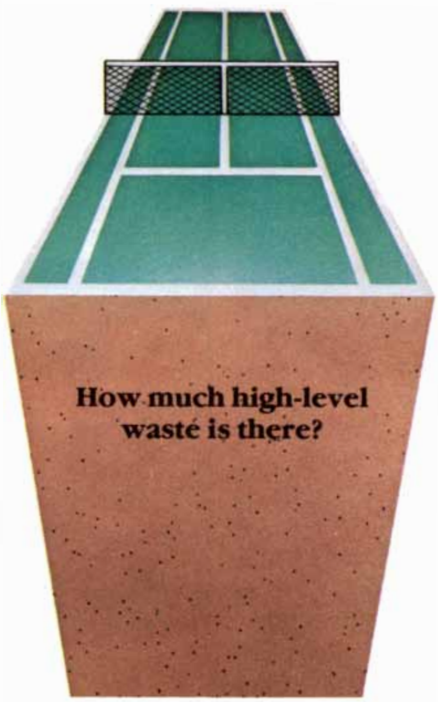
Scientists worldwide say disposal is practical and safe

Over 25 years of experience in managing nuclear waste have led to an international scientific consensus: The many natural and engineered barriers of a deep geologic repository will keep high-level waste safely isolated from our environment and our descendants. Countries pursuing this disposal method include West Germany, Canada, Switzerland, and Sweden.

The respected National Academy of Sciences and other eminent scientific bodies have reaffirmed that high-level nuclear waste can be disposed of thousands of feet underground, in geologic formations that have been stable for millions of years.

America's new law on nuclear waste

With this scientific consensus as background, Congress has passed and the President has signed into law the Nuclear Waste Policy Act of 1982.



The actual volume of high-level radioactive waste from nuclear power plants is small. The volume taken up by all the used uranium fuel produced to date by commercial nuclear power in this country is the same as the volume of a single tennis court about 45 feet deep.

It calls for lowering the sealed-up waste (in a glass- or ceramic-like form) about 2500 feet underground, into geologic formations such as granite, basalt, or salt.

Geologic disposal is designed to keep the waste safely isolated from our environment for many thousands of years. Eventually, the waste hazard decreases to the level of natural uranium ore.

The clear and present benefits of nuclear electricity

America's factories, businesses, and homes are relying more and more on an assured supply of electricity. Nuclear energy is a valuable, secure contributor to that supply. It supplements coal, replaces oil, and does not pollute the air or water.

Waste disposal need not be an obstacle in the use of such an important energy source. The scientific community has shown us that we do have the knowledge to handle nuclear waste—safely and confidently.

Free booklet tells more

For a free booklet that covers nuclear electricity in more detail, write to the U.S. Committee for Energy Awareness, P.O. Box 37012 (A6), Washington, D.C. 20013.

Information about energy America can count on today
U.S. COMMITTEE FOR ENERGY AWARENESS

© 1984 SCIENTIFIC AMERICAN, INC

inoids can be divided into two broad groups. One group can informally be called the dryomorphs, after the various European species of the genus *Dryopithecus,* a primate with many primitive hominoid characteristics. The other can be called the ramamorphs, after *Ramapithecus* and similar genera. The ramamorphs were mainly Asian in distribution; they differ from other Miocene hominoids in certain characteristics that resemble those of later hominoids.

Edward Lewis, a Yale graduate student who was later with the U.S. Geological Survey, gave the name *Ramapithecus* to a fossilized fragment of a primate upper jaw from India he had found in 1932. The fossil is now known to be a little more than seven million years old. By the 1960's more primate jaws and teeth from India, Pakistan and other parts of the Old World had been found and assigned to the same genus. The animal was seen as being different from its supposedly more apelike contemporaries, *Dryopithecus* and another Indian fossil find, *Sivapithecus.* Its various remains showed that its jaws were robust and its cheek teeth large, with thick enamel caps. These features and a few others resemble those of the later African hominids of the genus *Australopithecus.* The resemblances led many, including me, to argue that *Ramapithecus* was in fact an early hominid, that the hominids had diverged from the hominoids of Africa at least 15 million years ago and therefore that the divergence date of four million years ago, based on the molecular data, was wildly incorrect.

In 1973 I began a collaborative research project with the Geological Survey of Pakistan aimed at finding additional *Ramapithecus* fossils. The fossil-rich formations where we are still at work are the product of the erosion of the rising Himalayas; they are called the Siwalik series. They vary in age from one million to 17 million years and form a thick wedge that runs along the southern flank of the great mountain range from Afghanistan to Burma. In the section of the Siwalik formation in Pakistan we have been fortunate in finding many new hominoid fossils, including specimens of both *Ramapithecus* and its relative *Sivapithecus,* and thousands of other mammalian fossils.

Our large interdisciplinary team has now built up a dated sequence of faunal changes over the 16 million years of deposition in the region and has made considerable progress in reconstructing its animal communities of the past. It is clear from the new primate material that *Ramapithecus* and *Sivapithecus* are much more alike than I had thought; they may even belong to the same genus.

In any event the new *Sivapithecus* fossils, which include a partial skull and isolated limb bones, show a number of features, particularly of the face and palate, that resemble those of the sole surviving Asian great ape, the orangutan (*Pongo pygmaeus*). They do not, as expected, resemble features of *Australopithecus.* The jaw and tooth resemblances to *Australopithecus* remain but their phylogenetic significance can be dismissed; they are probably either instances of parallel evolution or features retained from the last common ancestor of all living hominoids. At least this is my opinion and that of my colleague Steven Ward of Kent State University. Some of the hominoid fossils from Pakistan that are at least 12 million years old show similar features. If our interpretation is correct, the split between African and Asian hominoids is certainly that old and probably even older. Perhaps the split coincided approximately with the continental-drift linkup of Africa

MILLIONS OF YEARS AGO	EPOCH	1978 CONSENSUS	1984 CONSENSUS
	LATEST PLEISTOCENE	Anatomically modern *H. sapiens* appears.	Anatomically modern *H. sapiens* appears (markedly different from archaic *H. sapiens* both cranially and subcranially). Abundant stone and bone tools and other artifacts.
.045		Little change from preceding phases	
	LATE PLEISTOCENE	*H. sapiens* appears.	
.125			New tool types and techniques; significant elaboration of behavior.
	MIDDLE PLEISTOCENE	Gradual hominid evolutionary progression.	Stone tools somewhat more complex. Archaic *H. sapiens* appears (little different from *H. erectus*).
.7			Use of fire?
	EARLY PLEISTOCENE	*H. erectus* appears: robust *Australopithecus* extinct.	Stone tools somewhat more complex.
1.6			*H. erectus* present. First *H. habilis* and then *A. boisei* and *A. robustus* extinct.
	PLIOCENE	Gracile, bipedal hominids in Tanzania and Ethiopia (including *A. afarensis*).	Earliest stone tools; *A. africanus, A. robustus* present in South Africa. Hominids diversify; *A. boisei, H. habilis* coexist. Laetoli footprints. Bipedal *Australopithecus* present in East Africa.
5			
	LATE MIOCENE	African apes and African protohominids diverge.	7.5-4.5 million years ago: major climatic/faunal changes; African hominoids diversify (inferred from molecular data).
11			
	MIDDLE MIOCENE	*Ramapithecus* established as protohominoid.	African and Asian hominoid split. *Sivapithecus (Ramapithecus)* related to *Pongo?*
17			
	EARLY MIOCENE		Hominoid radiation begins.
24	OLIGOCENE	Hominoids diverge from Old World monkeys.	Catarrhine radiation begins.

CHANGING VIEWS on hominoid and hominid evolution appear on this chart, which extends from Oligocene times at the bottom to the most recent Pleistocene at the top. Numerals between the geologic subdivisions give the terminal dates of each subdivision in millions of years or fractions thereof before the present. The 1978 consensus of scholarly opinion on the subject is given at the left in the chart, together with items of fossil or archaeological evidence in support of such opinion. Today's consensus of opinion is given at the right.

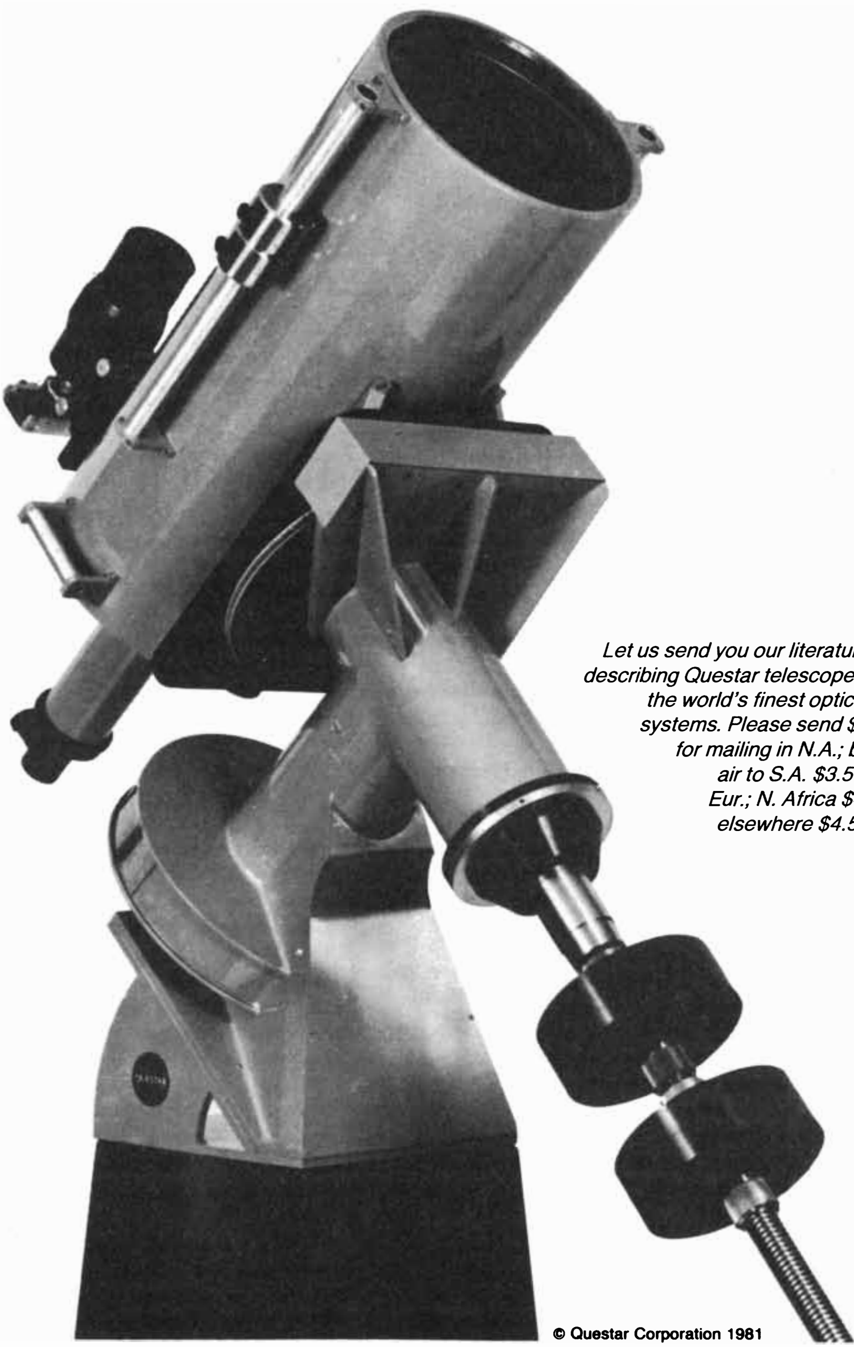
© 1984 SCIENTIFIC AMERICAN, INC

QUESTAR® 12 on the QUESTAR® MOUNT

The Questar 12, latest addition to the Questar family of fine optical instruments, now has its own Questar-designed mount. A German equatorial type, it is notable for its 360° continuous tracking in R.A. with precision tracking to better than 4 arc seconds. The Questar Mount is designed with over-sized components so that it can accommodate any Questar up to 18 inches. The standard mount shown is straightforward in design but can be modified so as to be compatible with more sophisticated tracking devices or other special equipment.

The Questar 12 is a superb instrument for the serious astronomer, for the university astronomy department or the engineer seeking sophisticated tracking and surveillance equipment for which Questar Corporation has a noted reputation.

Questar Corporation Box C, Dept. 20, New Hope, Pa. 18938 (215) 862-5277



Let us send you our literature describing Questar telescopes, the world's finest optical systems. Please send $2 for mailing in N.A.; by air to S.A. $3.50; Eur.; N. Africa $4; elsewhere $4.50

© Questar Corporation 1981

and Asia and the consequent mixing of faunas 17 or 16 million years ago. Tentatively I now place the split at about 16 million years ago.

Early Hominids

The first undoubted hominids appear from perhaps four million to 3.75 million years ago in Tanzania and Ethiopia. Hominid evolution before then is obscure. Changes in fauna and habitat between 7.5 and 4.5 million years ago reflect a number of climatic and geographic events, the most spectacular one being the drying up of the Mediterranean on several occasions about 5.5 million years ago. The connection between these events and the evolution of the hominoids of Africa is not clear. Nevertheless, the footprint evidence of hominids that walked on two legs, uncovered by Mary Leakey and her colleagues at Laetoli in Tanzania, is firmly dated at 3.75 million years ago. Moreover, the almost half-complete skeleton of "Lucy" and other specimens of *Australopithecus afarensis* found by Donald Johanson and his colleagues at Hadar in Ethiopia are dated, somewhat less certainly, at between four and three million years ago. There is some taxonomic dispute about Lucy and the associated hominid remains from Hadar. Do they represent a new single species, *Australopithecus afarensis?* A northern subspecies of *A. africanus?* Two or more species? Tiptoeing past these questions, I shall give only one version, acknowledging that the situation is not clear-cut.

A. afarensis is primitive in most of its features. The adult males were probably from 50 to 100 percent larger than the females. Individuals weighed from 25 to 50 kilograms or more. Their brain was the size of an African great ape's, but with such a wide range of adult weights it is not clear whether the brain size with respect to body weight was larger or smaller. In some features of the face and palate these early hominids resembled chimpanzees. Overall, however, the skull as it has been reconstructed looks more like a female gorilla's: *A. afarensis* has larger teeth and is more robust than a chimpanzee.

The dentition includes a few plausibly primitive features, although the canine teeth of both sexes have a low crown and have lost their apelike forward projection. The cheek teeth are relatively large and are capped with a thick enamel, probably an adaptation to chewing large quantities of fruits, seeds, pods, roots and tubers, some of which may have been quite tough.

It is impossible to prove that the footprints in Tanzania were made by hominids like those represented by the Hadar fossils, but the morphology of the hip, knee and ankle joints of Lucy and her

© 1984 SCIENTIFIC AMERICAN, INC

companions shows that the Hadar population was clearly bipedal. The foot bones show the same adaptation to bipedalism and an intriguing additional feature: the toe bones and metatarsals are long and curved. Compared with members of the genus *Homo* the arms of *A. afarensis* were long and the legs were short. Its hands were capable of powerful grasping. Judging by the proportions of the hand bones and the morphology of their joints, its manipulations were probably more precise than those of living chimpanzees.

The Hadar hominids lived in an area of woodland and savanna, away from the Pliocene forests, as part of a community of mammals that was structurally quite like the later communities of the Pleistocene. Such communities began to develop in eastern Africa in the later Miocene, and it is possible that they included hominids still unknown to us. By analogy with living mammals it is unlikely that the Hadar hominids were monogamous.

No stones that have been altered as tools have been found associated with the Hadar hominids. This may not, however, be entirely relevant. Suppose *A. afarensis* used the same kind of tools living chimpanzees do: stems, leaves, wood and stones casually adapted to food gathering, food processing and display behavior. Suppose, moreover, they did so more often than chimpanzees do today. Such tools would be impossible to recognize four million years later even if some near-miracle had preserved them.

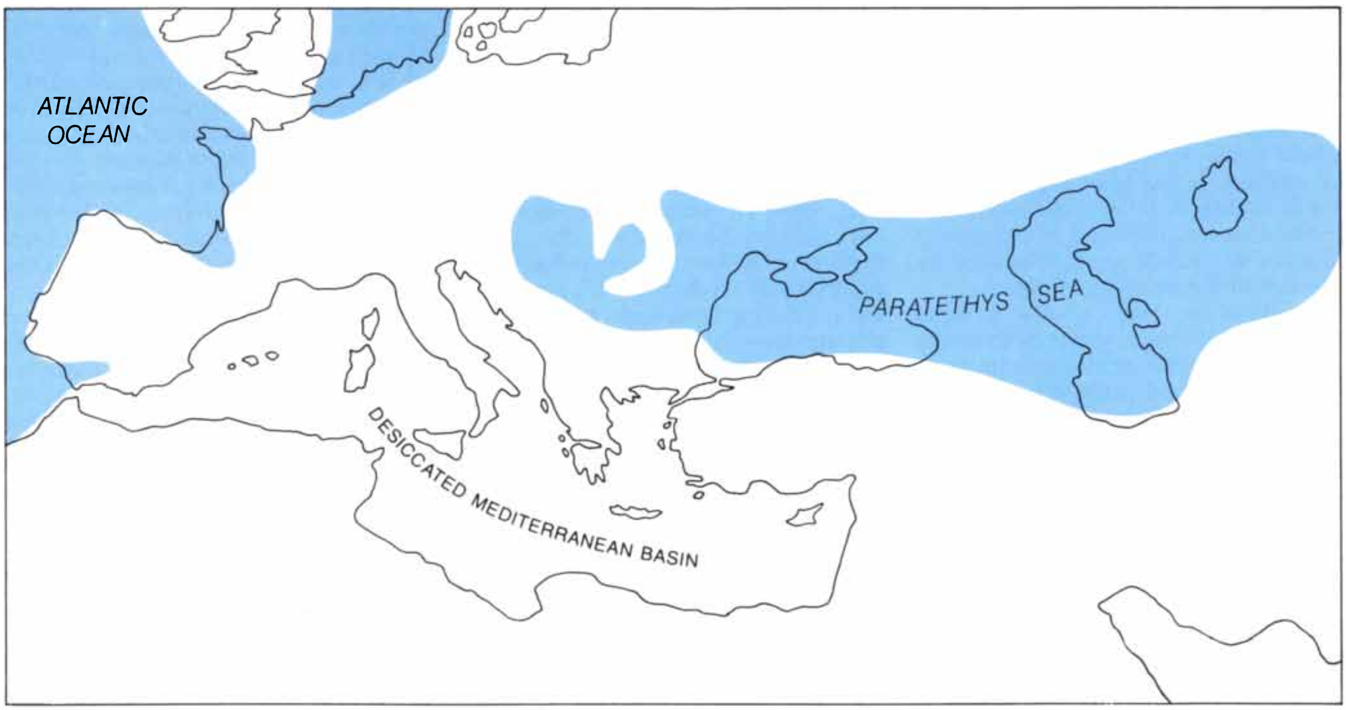
Tool use of this kind, if it may be postulated, might have been an important component of those behaviors that stimulated walking on two legs and the reduction of the canine teeth. Concepts such as these are much debated. Were the Hadar hominids protohunters and killers, opportunistic scavengers or gentle vegetarians? Did monogamy evolve along with bipedalism? Did pair-bonded males bring food to their female partner at a home base? Was gathered food, animal or plant, shared or did the Hadar and Laetoli hominids forage on an individual basis? Each of these viewpoints, to say nothing of others, has its adherents. For the present, however, any picture of what these early hominids were like as living, breathing animals must be deferred until there is more agreement on the nature of their basic adaptations. Above all, we must resist the temptation to see them as modern humans merely some distance removed from us in time.

Closing in on *Homo*

At some point between 2.5 and two million years ago African hominids underwent a modest adaptive radiation. By the later date at least two, probably three and perhaps even more species of hominids were present, although in any one part of Africa there is no evidence for more than three. In eastern Africa a species of *Australopithecus, A. boisei,* robust and with very large teeth, lived at the same time as a larger-brained hominid whose type-fossil cranium is catalogued in Kenya as E.R.-1470. The species is now most commonly known as *Homo habilis,* a name that was coined in 1964 by L. S. B. Leakey, Philip V. Tobias and John R. Napier, and was applied to E.R.-1470 by its discoverers, Richard Leakey and his colleagues. There was probably a third hominid species, small-toothed and small-brained, living in eastern Africa at the same time. Moreover, in southern Africa there was another species of *Australopithecus, A. robustus,* and possibly *H. habilis* as well. By about 1.75 million years ago, however, *H. habilis* disappeared from Africa and was replaced by an even larger-brained hominid: *Homo erectus.* A clearer picture of what kinds of animals these were is slowly emerging. The picture cannot ever, of course, be genuinely sharp, and one must avoid peering at it too closely lest, like a newspaper halftone, it dissolve into meaningless dots.

Consider *A. boisei.* A larger species than either *A. afarensis* or the *Australopithecus* species of southern Africa, it was small-brained and the male was markedly larger than the female. The species appeared more than two million years ago and survived, apparently with little morphological change, for a million years. *A. boisei* may have dug with sticks and hammered with stones, but there is no evidence that it used tools intensively.

Although *A. boisei* was larger than



LATE MIOCENE EPISODE, the drying up of the Mediterranean basin after it was sealed off from the Atlantic by mountain building, must have had catastrophic effects on the climate and ecology of Africa. The episode took place about 5.5 million years ago. Its effects on the mammalian fauna, hominoids included, are not clear, but it was perhaps a million and a half years later that hominids appeared.

© 1984 SCIENTIFIC AMERICAN, INC

other species of *Australopithecus*, it may have behaved much like them. Studies of the chewing surfaces of the teeth of various *Australopithecus* species by Alan Walker and by Frederick Grine of the State University of New York at Stony Brook suggest that all of them were basically rather apelike vegetarians, eating broadly similar diets that called for much repetitive chewing. Possibly their diet included more roots and tubers and less fruit than the diet of forest-dwelling apes.

Why did *A. boisei* evolve, and from what did it evolve? Was its evolution the result of a change in climate and habitat, perhaps a change in the availability of certain kinds of plant foods? Why did *A. boisei* become extinct after a million years of apparent stability? None of these questions can yet be answered. The fossil hunt in eastern Africa is intense, however, and one may hope *A. boisei*, its ancestor and the transition between them will eventually be better understood.

Understanding the transition to *Homo habilis* from its ancestor (which might also be the ancestor of *A. boisei*) is another matter. The various species of *Australopithecus* are difficult animals to "imagine"—to reconstruct as though they still lived—because nothing like them lives today. Still, they can be perceived dimly as an odd kind of ape. *H. habilis* is even more difficult to imagine. It is rather like *Australopithecus* in a number of ways, for example in its face and teeth, but it has a significantly larger brain, averaging about 700 cubic centimeters. There are no clear associations between the skulls of *H. habilis* and the other bones of the species, but the limb bones that are assumed to represent it, unlike those of *Australopithecus*, resemble those of later species of the genus *Homo* (with the exception of modern *H. sapiens*). It has been suggested that the resemblance reflects changes in the mode of locomotion and in the dimensions of the female pelvis demanded by the size of the newborn infant.

At about the same time that *A. boisei* and *H. habilis* appear in Africa so do the first archaeological sites: concentrations of used or altered stone, often brought from some distance away, together with animal remains. The consensus is that the sites document a shift in diet to include more animal food; *H. habilis* rather than *A. boisei* is generally implicated. The dietary change and the increase in brain size are also linked causally. Thereafter the consensus breaks down. Was *H. habilis* a hunter-gatherer who transported food to home bases to be shared and who practiced a division of labor that left hunting to the males and gathering to the females? This is to ask: Was *H. habilis* behaviorally "modern" or was it still basically vegetarian, perhaps adding some meat to its diet mainly by scavenging?

The archaeological record is itself ambiguous. The same scattering of tools and bones could be produced by opportunistic scavenging for individual consumption, with little or no food transport and sharing, or could represent home bases as sites of intense social interaction. In short, the picture of *H. habilis* remains blurred. More fossil material and more archaeological sites are needed, and even then the transition from *Australopithecus* to *Homo* is likely to be obscure. Both animals are too different from any known today. What happened to *H. habilis?* It survived for a few hundred thousand years only to be replaced by a much more durable species, *H. erectus*. Was *H. habilis* the ancestor of *H. erectus?* Although it is widely assumed to be the case, it may not be so.

Getting Around

Homo erectus is the first widely distributed hominid species. It appeared earliest in Africa, as far as is known, and it may indeed have originated there some 1.6 million years ago. Whether or not that is so, by one million years ago the species was present in southeastern and eastern Asia and survived in that area at least until 300,000 years ago. In that span of time, well over a million years, the physical record of *H. erectus* is one of prolonged morphological stability.

H. erectus resembled later species of *Homo* (modern *H. sapiens* excepted) in both body size and robustness. Larger-brained than *H. habilis* (more than 800 cubic centimeters), it had front teeth as big as those of earlier hominids, but its cheek teeth and its face were smaller. The archaeological record suggests that some *H. erectus* populations were makers of larger, symmetrically flaked stone tools: bifaces, or "hand axes." Moreover, some *H. erectus* populations may have used fire. Some later *H. erectus* populations may have commanded more sophisticated techiques of producing and modifying stone tools, but the overall impression is one of prolonged stability, in behavior as well as in morphology. Arthur Jelinek of the University of Arizona has proposed that "paleocultural" is the appropriate term for such behavior, so different is it from the rapid changes we associate with "cultural" behavior in modern humans.

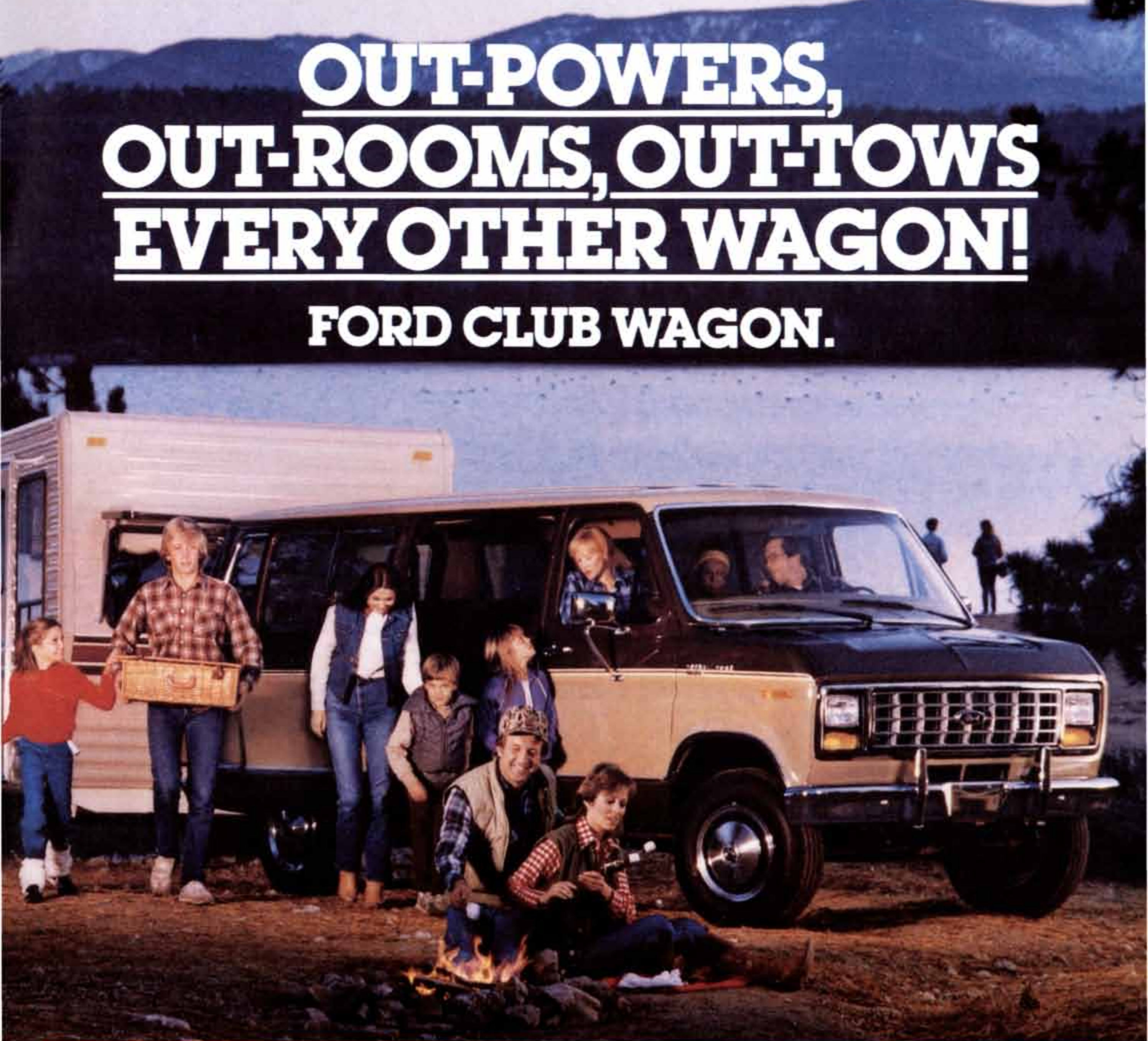
Paleoanthropologists shift gears, often without realizing it, when they talk about human evolution over the past 1.5 million years. The problems in hominoid and hominid evolution I have presented up to this point concern macroevolutionary questions: issues of broad adaptations, evolutionary trends and speciation. The hominids labeled *H. erectus*, "archaic" *H. sapiens* (including the Neanderthals) and "modern" *H. sapiens* probably represent a continuum, and the patterns of change within these lineages represent in contrast microevolution. Nevertheless, some of the most interesting paleoanthropological work concerns the last great step in human evolution: the shift, about 45,000 or 40,000 years ago, from archaic *H. sapiens* to modern *H. sapiens*.

Erik Trinkaus of the University of New Mexico has shown that the Neanderthals (and probably their contemporaries in other parts of the Old World) were probably as unlike us behaviorally as they were physically. The Neanderthals' skeletons were much more robust and the muscle attachments on the bones indicate that they were much stronger than we are. Their teeth were larger and are heavily worn, probably (by analogy with living Eskimos) from being used for a variety of nonfeeding activities (such as the chewing of animal hides to soften them). The transition to modern *H. sapiens* was marked by a loss of the Neanderthal robustness in skeleton, face and dentition. There were also changes in the morphology of the female pelvis. These perhaps suggest that formerly easy births had become harder or even that the time of gestation was being reduced to our present nine months from perhaps 11 months (a period in line with predictions based on general mammalian relations among maternal body size, fetal size and length of gestation).

Important behavioral changes are also evident in the archaeological record. They include a proliferation of superior stone and bone tools, shifts in hunting patterns, in the use and control of fire, in the use of clothing, in settlement patterns, in population size, in ecological range, in art and other evidences of ritual activity. All of this points to the emergence of a species possessing modern behavioral capabilities (and potential) from an ancestral species lacking, at least by modern standards, in some significantly human characteristics.

No brief summary such as this one can do justice to the richness and complexity of the information available regarding this most recent, even if microevolutionary, transition. Yet it remains the best-documented and potentially the most understandable of all the transitions of the past 30 million years. As an outsider, working in a time span that may contain no relevant fossils at all in any particular 100,000-year period, I am ruefully envious of those who work with this most recent period. Viewed at long range it is, in a very real sense, the icing on the cake. Yet one must remember that it is only the icing. The cake itself—the many earlier stages of human evolution—remains for now much harder to digest.

© 1984 SCIENTIFIC AMERICAN, INC



BEST-BUILT AMERICAN TRUCKS

POWERFUL CHOICE: SIX, DIESEL AND NEW MID-SIZE V-8!

For great outings, go with the wagon that outdoes all the rest—Ford's Club Wagon.

For starters, no other wagon, big or small, gives you such a power choice. V-8s headed by a new High Output 5.8L job* with 4-barrel carburetor — and 27% more power than Chevy's mid-size V-8. Ford's 4.9L I-6, the biggest Six in any wagon. The 7.5L,† most powerful V-8. And husky 6.9L diesel† that packs 35 more horsepower than Chevy diesel. (Dodge doesn't even offer one.)

And no other wagon gives you so much usable space—for people and/or cargo. You can seat up to 12 in roomy comfort (15 in Ford's Super Wagon).

Ford Clubs are built tough, with strong body-on-frame construction. The bigger models are rated to tow as much as 10,000 lb.: 2,500 more than Dodge, 3,000 more than Chevy.

*Optional: not available in Calif. or with manual transmission.
†Optional: 4.9L I-6 standard.

Get it together — Buckle up!

Quality is Job 1.

Ford's commitment to quality results in the best-built American trucks, based on a survey of owner-reported problems during the first three months of ownership of 1983 trucks. And the commitment continues in 1984.

BUILT FORD TOUGH

Ford

© 1984 SCIENTIFIC AMERICAN, INC

Excitonic Matter

A conduction electron can combine with a positively charged "hole" in a semiconductor to create an exciton. Excitons in turn can form molecules and liquids, and a new phase of matter may be attainable

by James P. Wolfe and André Mysyrowicz

The interaction of light with solid matter is a phenomenon of fundamental importance for exploring the quantum-mechanical underpinnings of nature. Albert Einstein's recognition that light energy is carried by quantized packets of radiation, now called photons, was based on the observation of one such interaction: the photoemission of electrons from matter. More recently it has been shown that the distribution of the electrons among their quantized energy states in a solid can be altered in a variety of ways in response to the energy carried by incident photons. A detailed understanding of these effects has significant implications for technology as well as for science. In particular the photoelectric properties of semiconductor crystals, which are now grown synthetically to ultrahigh purities in the manufacture of electronic components, have already been exploited to fabricate devices such as solid-state light detectors, light-emitting diodes and solar cells. As the interactions of electrons with photons in a solid become better understood, new and more efficient designs for such devices may become possible.

One of the most fascinating questions in the study of solids is presented by the beautiful effects of light absorption in a semiconductor crystal. For example, if green laser light is shined on the surface of a crystal of the semiconductor cuprous oxide (Cu_2O), the crystal will give off a red glow, which has a lower frequency than green light does. The energy of a photon is proportional to its frequency, and so the energy of each incident green photon is greater than the energy of each red photon that is emitted. A problem therefore arises: How is the energy of the green photons converted into the energy of the red photons?

Almost 30 years ago it was discovered that the energy of incident photons can be converted inside the crystal into short-lived neutral entities called excitons. Simply stated, the exciton is a quantum of electronic excitation in the crystal that is generated by an absorbed photon; the red light emitted by the crystal is the energy released when the electronic excitation of the crystal is relaxed. What makes the excited state interesting is that it can be understood as a particle. Indeed, the exciton resembles the hydrogen atom: it is made up of two oppositely charged carriers bound together by electrostatic attraction. In the hydrogen atom the positive charge is a proton, which is surrounded by the negatively charged electron. In the exciton the positive charge has a mass 1,000 times less than the mass of the proton.

In order to appreciate the implications of the discovery of the exciton imagine the world inside a semiconductor crystal in which the extremely light exciton has taken over the role of the hydrogen atom. A number of questions arise from the analogy: Can the exciton propagate freely through the crystal like a free hydrogen atom in a gas? Can two or more excitons combine to form a molecule? Can the excitonic "atoms" or the molecules made up of them form liquid or solid phases? Can more exotic phases of condensed excitonic matter come into being? The study of excitons inside crystals yields answers to these questions.

How are excitons created by light in a crystal? Even more fundamentally, why does the crystal absorb light at all? To answer these questions let us briefly examine the electrons in the crystal, because they are the particles that respond to the incident light. Instead of having well-separated values of energy as the electrons do in an isolated atom, the electrons in a crystal occupy broad bands of allowed energies. The bands are separated from each other by intervals of energy that are not accessible to an electron. Each band is made up of discrete states whose associated energy levels are packed so close together that they cannot be resolved. The number of states in a band is roughly equal to the number of atoms in the crystal; for a crystal of macroscopic size the number is on the order of 10^{23}. Hence within a band it is reasonable to think of the electronic energies as being spread over a continuum of states.

The electrons fill the allowed states in an energy band from the bottom upward, just as water fills a bottle. The distribution of the electrons among the energy bands in a solid determines whether the solid is a metal, a semiconductor or an insulator. In a metal the highest occupied band is only partially filled with electrons. Electric current flows easily in a metal because the change in the kinetic energies of the electrons that is required for a current can readily come about through a shift in the distribution of the electrons in the partially filled band. On the other hand, in a semiconductor or an insulator at low temperatures the highest occupied band, which is called the valence band, is completely filled with electrons just as a bottle is filled to the cork. The next energy band is called the conduction band, and it is empty at low temperatures; it is separated from the valence band by a "forbidden" energy gap, or band gap. There can be no net electric current associated with a completely filled band of electronic energy states. The band gap in a semiconductor is smaller than it is in an insulator, and so electrical conductivity does arise in a pure semiconductor as some valence electrons are thermally excited across the band gap. At low temperatures, however, a semiconductor is highly insulating.

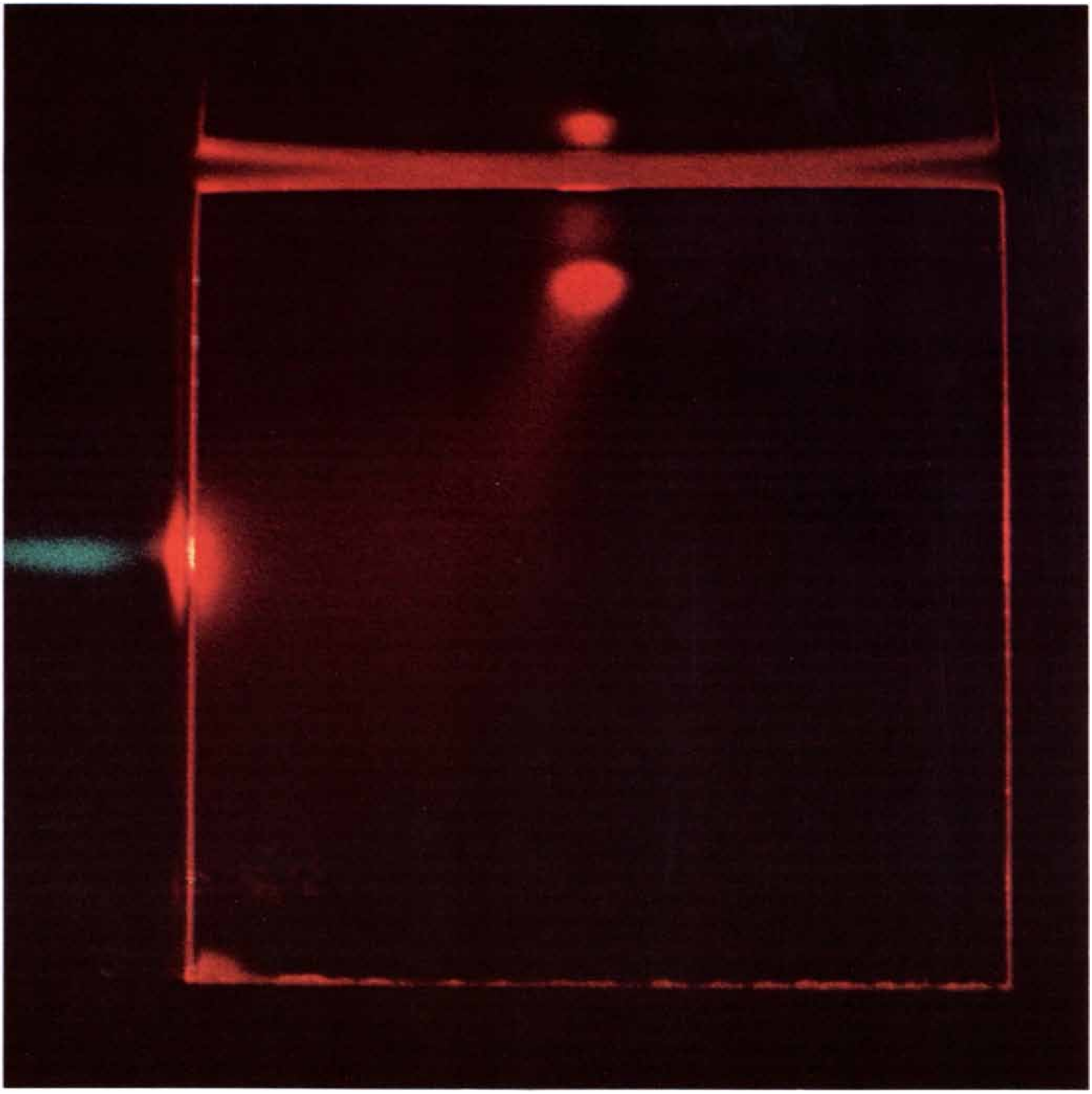
When a photon is absorbed by a semiconductor crystal, it promotes an electron from a filled valence band to an empty conduction band. In this process energy must be conserved, and energy conservation implies that in order to be absorbed the energy of the photon must be equal to or greater than the energy of the forbidden band gap. If the energy of the photon is lower than this threshold, no valence electron can absorb the light, because to do so would be to occu-

© 1984 SCIENTIFIC AMERICAN, INC

py the forbidden gap. In the absence of other effects a photon whose energy is less than the band-gap energy propagates unattenuated through the crystal.

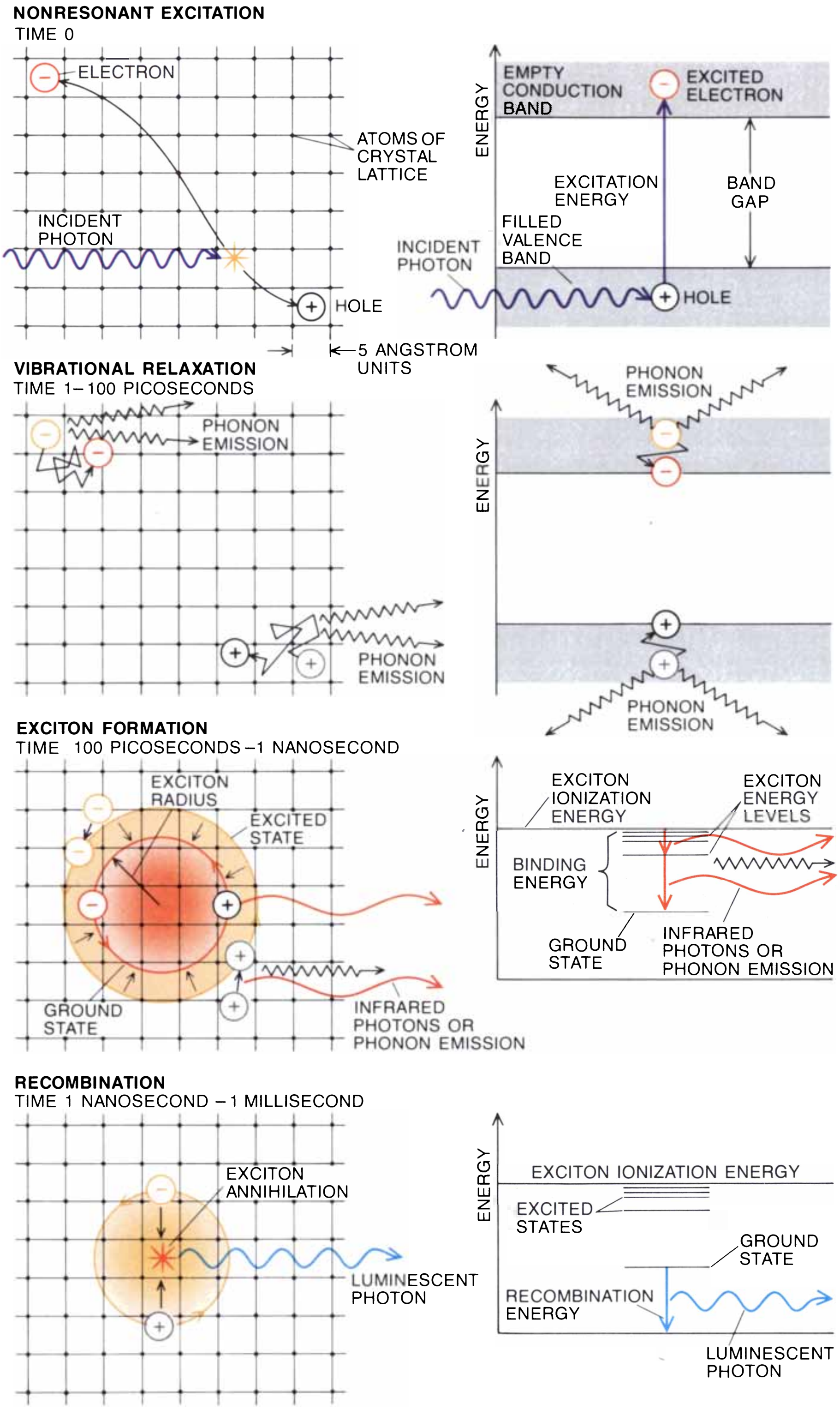
When an electron is removed from the valence band by a photon, it leaves an empty state behind, just as a bubble is formed when a drop is removed from a filled bottle of liquid. This empty state, or absence of an electron in the valence band, is called a hole. The hole can be thought of as a region of net positive charge in the crystal lattice, and in many respects its properties are like those of an ordinary particle. For example, an electron from a neighboring region of the lattice can move into the first region and make it electrically neutral, but in so doing the second region acquires the surplus positive charge. The region of positive charge can therefore propagate through the lattice much as if it were a real particle.

Once they are created in their respec-



EXCITON GAS is generated in a solid crystal of cuprous oxide (Cu_2O) by an intense beam of green light emitted by an argon-ion laser. Electrons that are associated with the atoms in the crystal are excited by the laser energy and move freely about in the crystal. The holes, or regions of net positive charge left behind by the excited electrons, also propagate freely like positively charged electrons. The crystal, which measures two millimeters on a side and is 1.5 millimeters thick, is immersed in a bath of liquid helium maintained at two degrees above absolute zero. At this temperature each free electron and hole can form a short-lived bound state called the exciton. The red light emitted by the crystal signals the presence of excitons; it is caused by the recombination of the electrons and holes after a few microseconds in the bound state. At the top of the photograph a round rod is pressed against the surface of the crystal, which reduces the energy for excitons just under the stress point. The excitons generated on the left side of the crystal stream rapidly to the low-energy region, which appears as a bright red spot because the excitons collect there. The spot is reflected by the stress rod, and the reflection is seen just above the crystal. The photograph was made by David P. Trauernicht and the authors at the University of Illinois at Urbana-Champaign.

© 1984 SCIENTIFIC AMERICAN, INC



NONRESONANT EXCITATION can give rise to excitons in a semiconductor when the excitation energy is greater than the energy of the band gap, the difference in energy between the most energetic electrons in the unexcited semiconductor crystal and the least energetic quantum-mechanical state to which they can be promoted. In the diagram the motion of the particles is shown schematically at the left and the energy transitions associated with each motion are shown at the right. The initial state of each stage in the lifetime of the exciton is shown in orange and gray and the final state is shown in red and black. After the initial excitation the free electron and the hole quickly relax to an energy equal to the band-gap energy by emitting phonons, or quanta of vibrational energy, to the surrounding crystal. They can then become bound and form an exciton in one of the discrete energy levels allowed by quantum mechanics. An exciton formed in an excited energy state can fall to a lower state before the electron and the hole recombine. The relative energy of each transition is indicated by a color: low-energy transitions are accompanied by photons of low frequency, shown schematically in red, whereas higher-energy transitions are accompanied by photons of higher frequency, shown in blue or violet. The spacing between the excitonic energy levels has been exaggerated for clarity.

tive bands, both the electron and the hole quickly proceed to their state of minimum energy. The excited electron falls to the bottom of the conduction band, and the hole rises like a bubble in a bottle of water to the top of the valence band. The hole rises because the electrons remaining in the valence band displace the hole as they fall into the lowest of the available states. Then within a nanosecond (10^{-9} second) the electron and the hole can further reduce their total energy by binding together to form an exciton.

It is well known that the electron and the proton in the hydrogen atom can assume one of a series of discrete energy values. Similarly, there is a set of allowed energy values for the exciton, corresponding to different excited states of the bound electron-hole pair. The excitonic energy levels converge toward a high-energy limit, which is equal to the energy of the band gap of the semiconductor. The binding energy of the exciton, however, is typically 100 to 1,000 times lower than the binding energy of the hydrogen atom.

The weak bond implies that the mean distance between the electron and the hole can be one or two orders of magnitude greater than the diameter of the hydrogen atom. Indeed, the exciton orbit can range over many spacings of the crystal lattice. For example, the binding energy of the exciton formed in the semiconductor germanium is only .0041 electron volt and its radius is 130 angstrom units (10^{-8} centimeter). (For comparison the binding energy of the hydrogen atom is 13.6 electron volts and its radius is .5 angstrom; the typical spacing between atoms in a crystal is from three to five angstroms.) At ordinary temperatures the thermal vibrations of the lattice are violent enough to overwhelm the weak excitonic bond. To produce excitons in a significant quantity it is usually necessary to cool the semiconductor to within a few tens of degrees above absolute zero.

How can one detect the excitonic energy levels? One way is to measure the amount of light that is transmitted through a semiconductor crystal for each color of incident light. A number of dark lines are observed in the spectrum of light transmitted by the crystal. Each line is caused by the conversion of light into excitons in a distinct energy state. The photons that cause the transitions are absorbed by the crystal and so the position of the dark lines in the spectrum of the transmitted light gives the color and hence the energy of the absorbed light.

A classic example of an absorption spectrum that gives the energy structures of excitons is the spectrum of cuprous oxide. As many as eight different exciton absorption lines are observed,

© 1984 SCIENTIFIC AMERICAN, INC

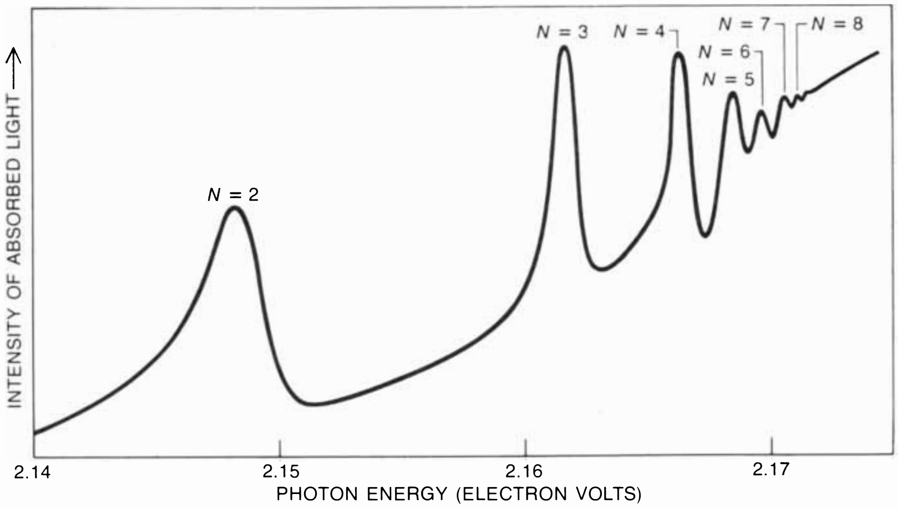
which form a converging series similar to the lines in the spectrum of monatomic hydrogen. The absorption spectrum gives the binding energy of the exciton in cuprous oxide: it is .14 electron volt, which corresponds to an exciton radius of 10 angstroms.

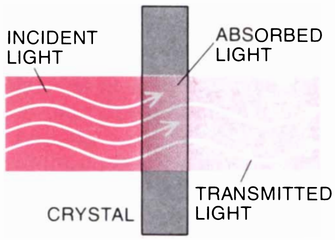
Another way to extract information about excitons is to detect the light they emit. Once an exciton is formed it does not live forever. The electron eventually drops back into the hole, a photon is emitted and the exciton vanishes. The process is called recombination. The probability of recombination varies a great deal from crystal to crystal, depending on detailed characteristics of the valence band and the conduction band. In pure crystals the exciton lifetime ranges from nanoseconds to milliseconds.

By analyzing the spectrum of the luminescent photons emitted in recombination one can reconstruct the energy distribution of a collection of excitons. Suppose, for example, a crystal of cuprous oxide is illuminated with green light, whose energy exceeds the energy of the band gap. Remember that in these circumstances the crystal emits a characteristic red light. The emission is caused by the recombination of excitons in the ground state, the exciton state of lowest energy. The light is not, however, emitted at a single frequency; it includes a range of frequencies whose width increases as the temperature of the crystal is raised. The range of emitted frequencies corresponds to the range of exciton energies present in the crystal at the time of recombination. Since the total energy of the exciton is determined by the band-gap energy, the binding energy and the kinetic energy associated with the motion of the center of the exciton mass, the range of emitted frequencies gives the distribution of kinetic energy, and thus the distribution of velocity, for the excitons.

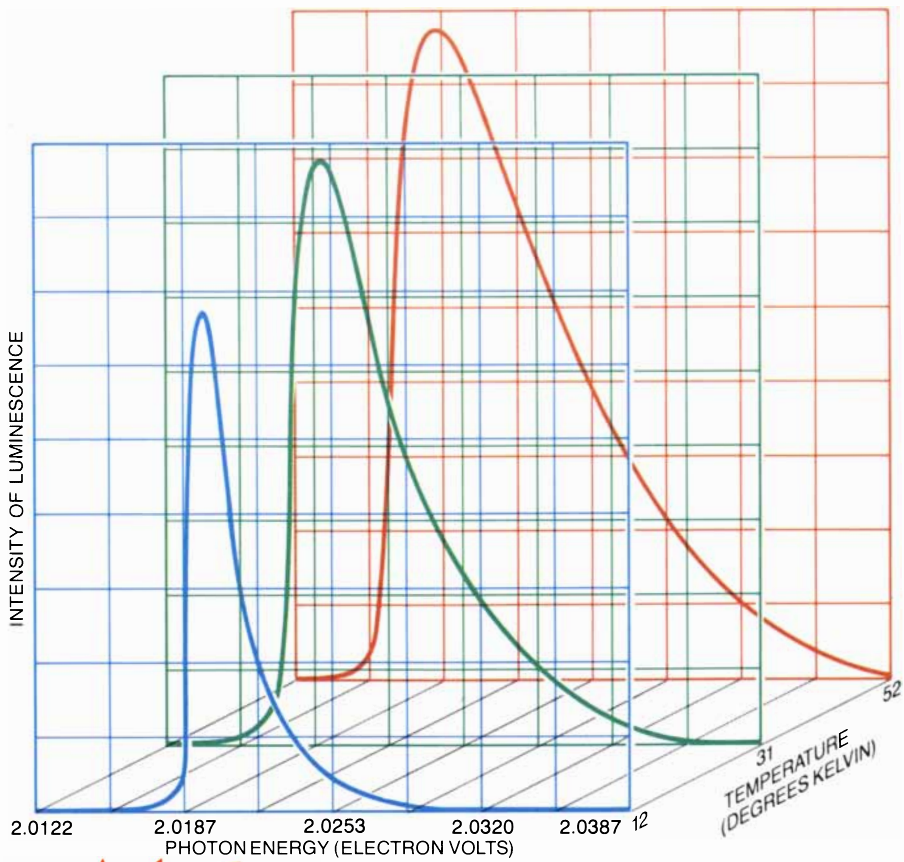
The luminescence spectrum of cuprous oxide shows unequivocally that the excitons must be in motion within the crystal. The distribution of the exciton velocities agrees well with the distribution of the velocities of ordinary atoms in a low-density gas. In other words, at low densities a group of excitons can be viewed as a gas of particles confined inside a solid medium!

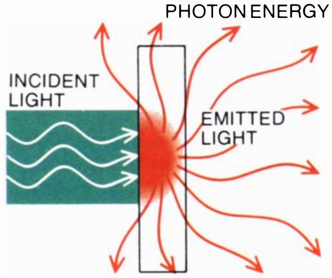
In order to understand the motion of a gas of excitons think first about a gas that can be seen, such as the steam emanating from the spout of a kettle about to boil. The steam diffuses from the spout into the surrounding air and eventually is dissipated. Similarly, the excitons created at a crystal surface by photons migrate into the crystal and eventually recombine. The migrating excitons carry no electric charge because they are neutral, but they do transport energy,





RESONANT EXCITATION of the exciton is caused by the absorption of a photon whose energy corresponds precisely to one of the excitonic energy levels. The absorption spectrum of crystalline cuprous oxide is plotted here. The peaks in the spectrum, designated by the numbers *N*, correspond to the energy levels of excitons generated in the crystal by the energy of the incident light photons. Experiments of this kind were originally carried out by Evgenii F. Gross of the University of Leningrad and by Serge Nikitine of the University of Strasbourg.





LUMINESCENCE SPECTRUM emitted as excitons recombine from their ground state in cuprous oxide makes it possible to infer the distribution of the kinetic energies of the excitons just before recombination. The energy of recombination is the energy of the ground state and the kinetic energy of the exciton. The width of the spectrum narrows as the crystal temperature is lowered, which shows the distribution of kinetic energies also narrows. Data are from Danièle Hulin, André Antonetti and one of the authors (Mysyrowicz) at the University of Paris.

© 1984 SCIENTIFIC AMERICAN, INC

and the energy is released when the electron and the hole recombine. It is worth noting that excitons are also observed in living matter. The migration of excitons in living matter may be an important mechanism of energy transport for photochemical reactions such as photosynthesis.

What does the motion of an exciton gas look like on a microscopic scale? The exciton luminescence spectrum indicates that the excitons are in thermal equilibrium with the atoms in the crystal. To achieve the equilibrium the excitons must be able to exchange their kinetic energy with the vibrational energy of the lattice. In other words, each exciton must scatter from the thermally vibrating atoms, and so one might expect it undergoes random, Brownian motion. It can be shown that the average radial distance traversed by such a particle in time t is proportional to the square root of t; hence the average cross-sectional area of a cloud of gas expanding from a central point is proportional to the elapsed diffusion time t. The constant of proportionality is called the diffusion constant.

How is the diffusion constant measured for a gas of excitons? The basic experimental idea is to create the excitons at a point on the surface of a crystal with a pulse of laser light say 100 nanoseconds long. A "snapshot" of the exciton luminescence is then made at various times during the expansion of the gas. The effective shutter speed required for such a snapshot is about 10,000 times faster than the fastest shutter speed of an ordinary camera, and the amount of light emitted by the excitons in that time is minuscule. Thus in order to make the snapshot it is necessary to count individual luminescent photons with a photomultiplier tube.

The photon pulses are electronically selected from 100-nanosecond intervals centered on various times after each laser pulse. The procedure is repeated about a million times; each laser pulse creates a new cloud of excitons that diffuses and decays before the next pulse is generated. For each pulse the number of photons emitted from a small region of the crystal is recorded by a computer, and so a picture of the luminescence from different regions of the crystal is built up region by region.

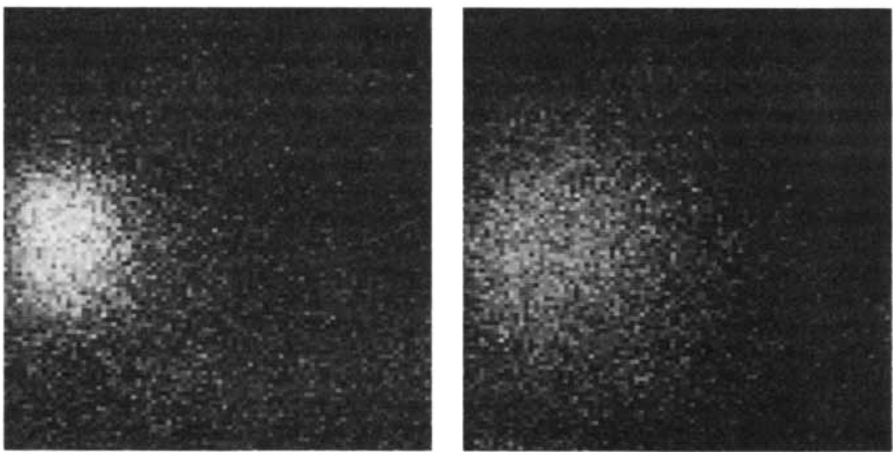
The resulting image of the exciton luminescence for a given time interval after the laser pulse can be displayed on a television monitor. When the cross-sectional area of the bright region of luminescence is plotted against time, the area is found to increase in direct proportion to the time elapsed after the pulse. This relation shows that the excitons diffuse like a dilute gas of atoms. The diffusion constant for an exciton gas, however, is much larger than it is for an ordinary atomic gas. In cuprous oxide, for example, the diffusion constant is 1,000 square centimeters per second at a temperature of 1.2 degrees Kelvin (degrees Celsius above absolute zero), whereas for nitrogen molecules in the air at room temperature it is about .2 square centimeter per second. The extremely fast diffusion of the exciton is a result of its small mass and its relatively infrequent scattering by other particles at low temperatures.

Another informative means of characterizing the motion of the exciton is to measure its mobility. When a force is applied to an exciton, Newton's second law of motion requires that the exciton be accelerated through the crystal in inverse proportion to its effective inertial mass. Collisions between the exciton and the thermally vibrating atoms of the lattice, however, constantly change the direction of motion of the exciton. The exciton must therefore begin accelerating again each time it is scattered. For a given field of force it can be shown that the exciton must attain some average drift velocity along the field whose magnitude is proportional to the strength of the field. The mobility of a continually scattering particle is therefore defined as the ratio of the drift velocity of the particle to the force applied. Mobility is commonly measured for a charged particle such as the electron in an electric field, and the force on such a particle is measured in volts per centimeter. Hence mobility is usually expressed in centimeters squared per volt per second.

In order to measure the mobility of the exciton one must determine its velocity under an applied force. Obviously the force cannot be derived from an electric field because the exciton is electrically neutral. The trick is to apply an external stress to the semiconductor crystal. The stress lowers the band-gap energy of the crystal and so reduces the energy of the exciton. If a gradually increasing stress is imposed across the crystal, the exciton can lower its energy by moving from a region of low stress to one of higher stress. The energy gradient, or change in the energy of the exciton per unit distance, is the force on the exciton.

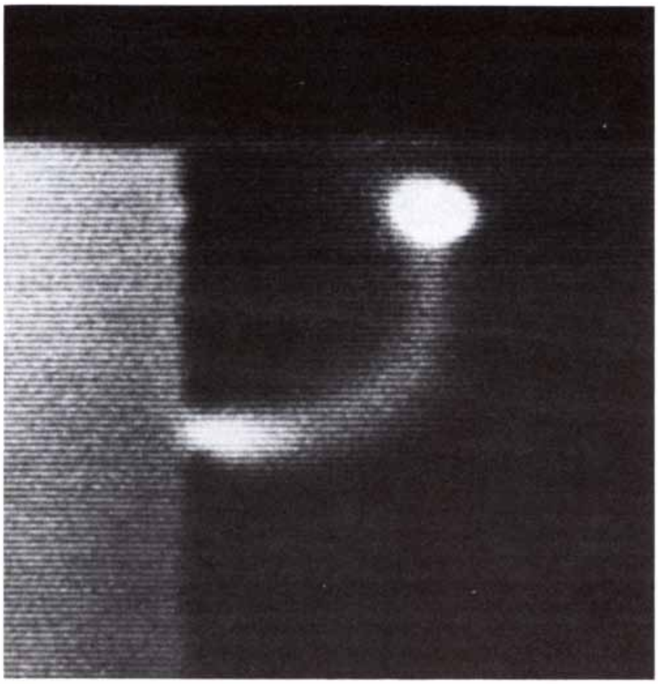
Stress-induced energy gradients have been applied to measure the mobility of the exciton in silicon, germanium and cuprous oxide. The crystal is pressed at the center of one of its polished surfaces by the rounded end of a rod. The stress in the material is maximized inside the crystal a few tenths of a millimeter under the contact point. Excitons generated by light shining on an adjacent surface of the crystal stream toward this high-stress region, and a few of them recombine along the way. The force on the exciton can be determined by plotting the shift in the frequency of the light emitted by excitons against their position in the crystal. The spectral shift for a unit displacement is equivalent to a change in the band-gap energy for that displacement, and this ratio is the force on the exciton.

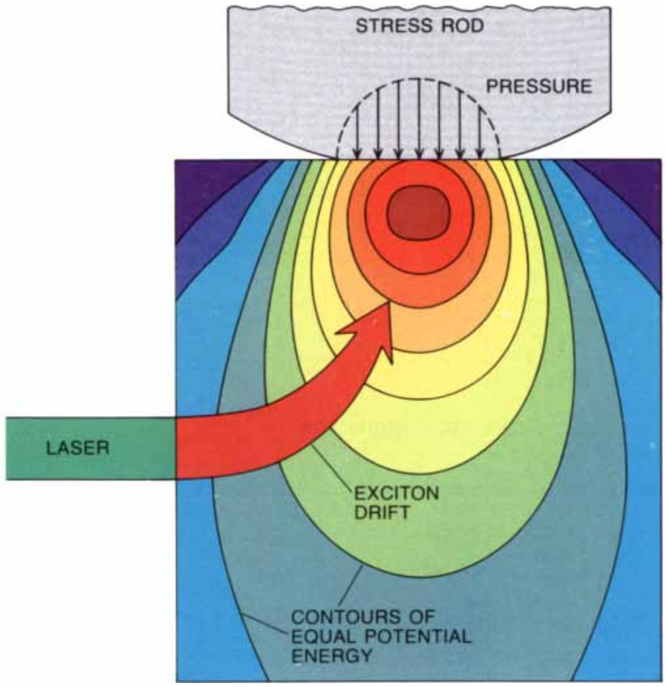
A striking example of exciton drift is shown in a photograph of a silicon crystal [*see illustration on opposite page*]. The crystal was cooled to about 10 degrees K. and then excited by a continuous laser beam. A vidicon television camera sensitive to infrared radiation



DIFFUSION OF EXCITONS created near the surface of a crystal of cuprous oxide is shown in two stages. As the exciton gas expands into the crystal some excitons recombine and give off a characteristic luminescence. The intensity of the luminescence was measured by a photomultiplier for each small picture element in the two images above. At the left the expansion is shown 200 nanoseconds (billionths of a second) after excitation by a laser beam at the left edge of the crystal; at the right the expansion is shown after 600 nanoseconds. The diffusion constant, which measures the rate at which the exciton gas expands into the crystal, can be determined from the images; because of the low exciton mass, the exciton diffusion constant is higher than that of any known atomic or molecular gas. The images were obtained by Trauernicht.

© 1984 SCIENTIFIC AMERICAN, INC





ENERGY OF THE EXCITON can be changed from region to region across a crystal by putting pressure on the surface of the crystal with the rounded end of a stress rod. The resulting energy gradient, or change in the energy of the exciton per unit distance in the crystal, is a force that tends to drive the exciton to the regions of lower energy. In the photograph at the left, which was made by Paul L. Gourley and one of the authors (Wolfe) at Illinois, the stress rod presses down on the center of the top surface of a crystal of silicon. The potential well, or region of lowest potential energy for the exciton, lies just under the stress point, as it does in cuprous oxide; the luminescence that signals the flow and pooling of the excitons, however, is in the infrared region of the spectrum, and so it was recorded by a television camera sensitive to infrared radiation and then converted to a visible image on the screen of a cathode-ray tube. The recombination frequency of the excitons is reduced as they move to regions of lower energy. The diagram at the right shows contours of constant potential energy that were calculated by Robert S. Markiewicz at the University of California at Berkeley. Regions of decreasing exciton potential energy for the crystal in the photograph are shown from violet through blue, green, orange and progressively darker shades of red, which represent the decreasing frequency of the infrared radiation emitted when pairs of electrons and holes recombine in these regions.

recorded the exciton luminescence at a wavelength of 1.2 micrometers, which corresponds to a photon energy about half that of yellow light. The path of the exciton drift is clearly observable because the strain gradient, or change in the stress per unit distance, can be quite steep in silicon. In order to make the photograph a force corresponding to about 100 kilograms was applied to the silicon crystal over an area of a square millimeter; such a pressure would crush many other semiconductor crystals in which excitons can be generated, such as cuprous oxide.

The drift velocity of the excitons is determined by pulsing the laser excitation light and then observing the peak of the exciton distribution at various times as it drifts into the crystal. The drift velocities of excitons in highly purified crystals maintained at 1.5 degrees K. vary from 100 to 1,000 meters per second, and so the calculated mobilities are on the order of 10^6 to 10^7 centimeters squared per volt per second. For comparison the drift velocity of the electrons that make up an electric current of one ampere in a copper wire one millimeter in diameter is less than .1 millimeter per second. In other words, if the electrons in the wire could drift at the same velocity as the excitons drift in a crystal, the current in the wire would be nearly 10 million amperes. At room temperature the mobility of the electrons in the wire is about 40 centimeters squared per volt per second, which is about five orders of magnitude less than the mobility of the exciton.

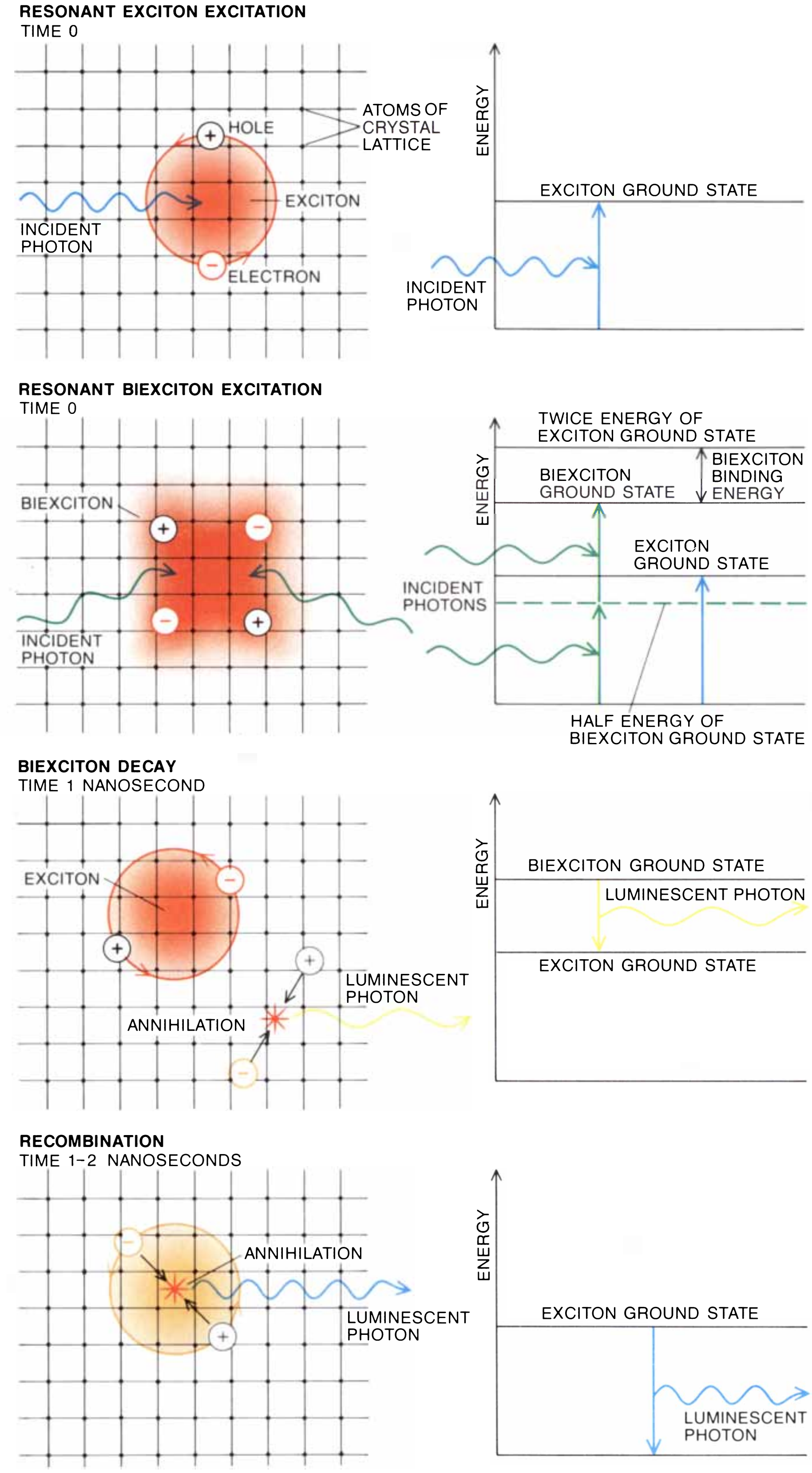
In all semiconductors studied so far the mobility of the excitons in a crystal increases dramatically as the temperature of the crystal is lowered. Such a relation between mobility and temperature is additional confirmation that the excitons are scattering from the crystal lattice: as the crystal temperature is lowered the amplitude of the lattice vibrations is reduced, and so the time interval between scattering events should increase. Drift measurements can therefore yield microscopic information about the motion of excitons. For example, the exciton in cuprous oxide at 1.5 degrees K. must travel about 30 micrometers between each scattering event, an enormous distance on the atomic scale. The exciton is clearly an extremely mobile particle.

How do excitons react to the presence of other excitons? The probability of interactions among excitons can be increased by turning up the power of the incident laser, thereby increasing the density of the exciton gas. In this context it is again fruitful to pursue the analogy between the exciton and the hydrogen atom. Even though the hydrogen atom is electrically neutral, the atoms in a gas have a strong tendency to combine and form molecular hydrogen (H_2). Each atom acts like an electric dipole, and two electric dipoles attract each other just as two magnetic dipoles do. Because of this attraction, the energy of the hydrogen molecule is lower than that of two separated atoms.

The energy of two excitons can similarly be reduced if they form an excitonic molecule, or biexciton. Because the biexciton is made up of four particles of similar mass, the molecule was originally predicted to be only marginally stable. Nevertheless, biexcitons have been detected in such crystals as silicon, germanium, copper chloride, silver bromide and cadmium sulfide.

A common method of creating biexcitons is to generate excitons in such copious quantities that they bind together. The presence of biexcitons is often signaled by a new luminescence line. The energy of the photons that give rise to the new line is lower than that associated

© 1984 SCIENTIFIC AMERICAN, INC



BIEXCITON, or excitonic molecule, is created when two laser beams that have the same frequency converge in a thin crystalline film. Each beam by itself would pass relatively unattenuated through the crystal, but the interaction of the two beams with the crystal can lead to absorption at a frequency equal to the sum of the frequencies of the beams. The absorption takes place when the frequency of each beam is adjusted to correspond to half of the total energy of the biexciton. The biexciton is a bound state of two electrons and two holes that forms because its energy is lower than that of two free excitons. It decays when one electron-hole pair recombine and the second pair form an ordinary exciton. The second pair recombine shortly thereafter. The diagrams follow the same conventions as the ones in the illustration on page 100.

with the recombination of a single exciton. Hence the line could arise when an electron-hole pair in a biexciton recombine and leave a free exciton behind. Moreover, in silicon and germanium the intensity of the new luminescence line increases as the square of the intensity of the exciton luminescence. This increase is precisely what is expected for biexcitons, because the probability of creating a biexciton increases as the square of the exciton density.

Although biexcitons are often generated simply by increasing the density of excitons, they can also be produced by direct optical excitation. One method requires the simultaneous absorption of two photons, each one adjusted to precisely half the energy of the biexciton. Since the biexciton is made up of two bound electron-hole pairs, the total energy of the two photons must be equal to the energy of two free excitons, less the binding energy of the biexciton [*see illustration at left*]. When this resonant condition is met by adjusting the laser frequency, a strong absorption line is observed. The line is quite special: it appears only when the intensity of the incident light is high. In a crystal of copper chloride one micrometer thick a laser beam with a power of one megawatt is needed to cause 10 percent of the photons to be absorbed. Compared with other processes in solids that involve the absorption of two photons this optical effect is quite strong.

What about other phases of excitonic matter, analogous to the liquid and solid forms of molecular hydrogen found at low temperatures? So far neither a molecular liquid nor any solid phase of excitonic matter has been observed. On the other hand, a new phase, not observed in hydrogen, has been found in a number of crystals. As the temperature of such a crystal is reduced to a few degrees above absolute zero a broad range of luminescent frequencies is emitted; neither exciton nor biexciton recombination can account for the range of frequencies or for the frequency of the most intense light within the range. Moreover, it has been found that the density of the particles that give rise to the emission is much higher than it is in an exciton gas.

An elegant way to demonstrate the increase in density experimentally is based on a method similar to the one applied to measure exciton drift. Remember that as a crystal is stressed by the rounded end of a rod a region of maximum shear stress is set up inside the crystal. A similar effect is well known to those who design ball bearings: a ball bearing under extreme stress fractures from the inside because the point of highest stress is inside the ball. Since the band-gap energy decreases with increasing stress, a

© 1984 SCIENTIFIC AMERICAN, INC

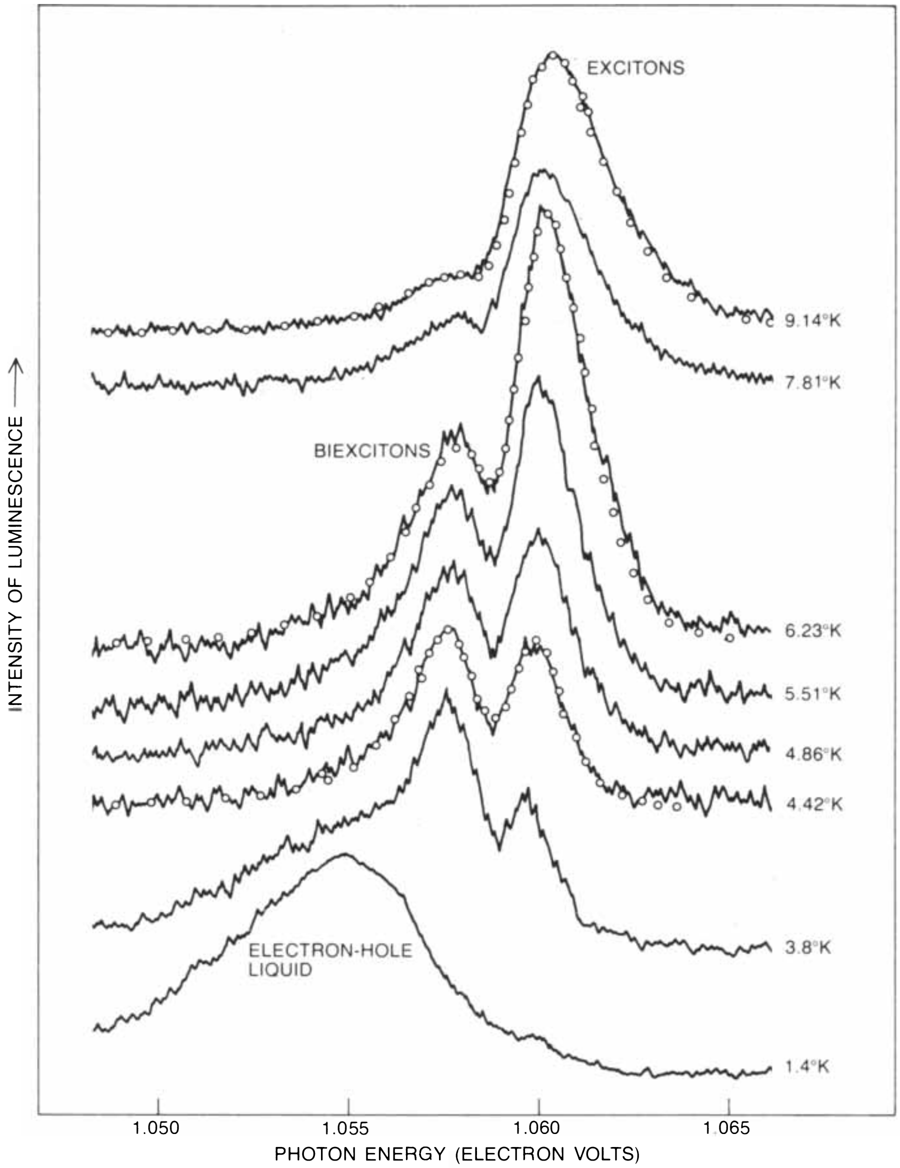
"potential well," or region of minimum potential energy for the exciton, is formed in the crystal. The potential well traps the excitonic matter and confines it. In experiments with silicon it has been found that the appearance of the new luminescence peak emitted from the potential well is accompanied by an abrupt decline in the volume of the emitting region of the well. In other words, the density of the particles confined by the potential well is increased.

Why are the energies of particles in the condensed phase distributed over such a large range? To understand this phenomenon the statistical properties of excitons must be examined more closely. At a fundamental level all particles in nature can be classified into two major groups according to their spin. The spin of a particle is like the spin of a ball, except that in quantum mechanics spin is defined only for discrete values. A particle whose spin is an integer is called a boson, and a particle whose spin is an integer plus 1/2 is called a fermion.

Both the electron and the hole have spin 1/2, and so they are fermions. An exciton, which is made up of two spin-1/2 particles, has integral spin, and so it is a boson. At high densities a collection of fermions has quantum-mechanical properties quite different from those of a collection of bosons. According to the exclusion principle, first formulated by Wolfgang Pauli, only one fermion can occupy a quantum-mechanical state at a time. On the other hand, there is no limit to the number of bosons that can occupy a given state.

The distinction between fermions and bosons is particularly important at low temperatures, where the particles tend to occupy the lowest energy states available. If the temperature is reduced for a collection of fermions, the states of lowest energy fill up without multiple occupancy. Hence even at absolute zero a collection of fermions must have a broad range of energies; the occupied state with the highest energy depends on the density of the fermions.

It is this fermionic property of electrons and holes that accounts for the broad range of low-temperature luminescence in silicon. The width of the spectral range narrows only slightly as the temperature is reduced, which indicates that the excitons have dissociated into electrons and holes that fill all the available states according to the exclusion principle. The measured width of the spectral range gives the density of the particles in the condensed phase; for electrons and holes confined at the bottom of a potential well in silicon the density is 3×10^{17} particles per cubic centimeter, or one-millionth of the atomic density of the crystal. This fluid of fermions is called the electron-hole liquid.



BIEXCITON GAS AND A LIQUID PHASE made up of electrons and holes can be detected from the luminescence spectrum of excitonic matter confined to a potential well in a crystal of silicon. The potential well is the stress-induced region of lowest energy in the crystal shown in the illustration on page 103. At the highest temperature for which the spectrum is plotted excitons predominate in the potential well. As the temperature is lowered a distinct peak of luminescence from biexcitons is observed, shifted from the energy of the excitons by an amount equal to the binding energy of the biexciton. At the lowest temperatures a broad peak appears, which signals the condensation of the excitons and biexcitons into the liquid phase. The open circles indicate theoretical values of the intensity. Data were collected by Gourley and Wolfe.

In germanium a single large droplet of pure electron-hole liquid up to a millimeter in diameter has been observed and studied by various means. Because the liquid is made up of free charge carriers, it has an electrical conductivity comparable to that of copper wire.

In crystals without external stress the electron-hole liquid collects into much smaller droplets. The size of the droplets has been estimated in germanium by measuring their scattering of infrared laser radiation. They are about five micrometers across, which corresponds to 100 million electron-hole pairs per droplet. Typically the droplets form a cloud surrounding the excitation point on the surface of the crystal and inside the exciton gas.

The cloud of droplets is like a fog of liquid particles, but it does not remain at rest. The droplets nucleate like water droplets in the atmosphere; they have a weak but finite surface tension equal to about a millionth of the surface tension of water. Moreover, they can be pushed through the crystal by a "wind" of phonons: the particlelike quanta of vibrational energy in the atomic lattice. For example, a nonuniform wind of phonons propagates through the crystal from the laser-excitation point, and the

© 1984 SCIENTIFIC AMERICAN, INC

wind blows the cloud of electron-hole droplets farther in some directions than in others. The study of interactions of the phonons with the droplets has led to a better understanding of phonon propagation at low temperatures.

Not all excitons in semiconductors can condense into electron-hole liquids. The excitons in cuprous oxide, for example, are predicted to remain in the gas phase even at high densities, and neither biexcitons nor an electron-hole liquid has been observed. In copper chloride biexcitons can form, but no liquid has been observed. If the excitons in these crystals remain in the gas phase, the width of the spectral range that signals exciton recombination should narrow as the temperature is decreased, because the excitons are bosons and bosons tend to fill up the states of lowest energy. Indeed, if the concentration of a gas of bosons exceeds a critical value, which depends on the temperature and the mass of the boson, all particles added to the gas are predicted to condense into a nearly motionless state of minimum energy. The effect, which is based on a theory formulated for massless photons in 1924 by the Indian physicist S. N. Bose and subsequently generalized by Einstein for particles with mass, is called Bose-Einstein condensation.

It is not yet possible to state without reservation that Bose-Einstein condensation has been observed. Many people believe the condensation can account for the sudden drop in viscosity and for the other properties of a superfluid: the state entered into by liquid helium at a temperature near absolute zero. Because of the strong interactions among the particles in a liquid, however, it is difficult to describe theoretically how Bose-Einstein condensation takes place in helium 4, the isotope of helium that becomes superfluid at 2.17 degrees K. Many investigators have sought instead to observe the effect in a dilute gas, for which the theoretical model is much more straightforward.

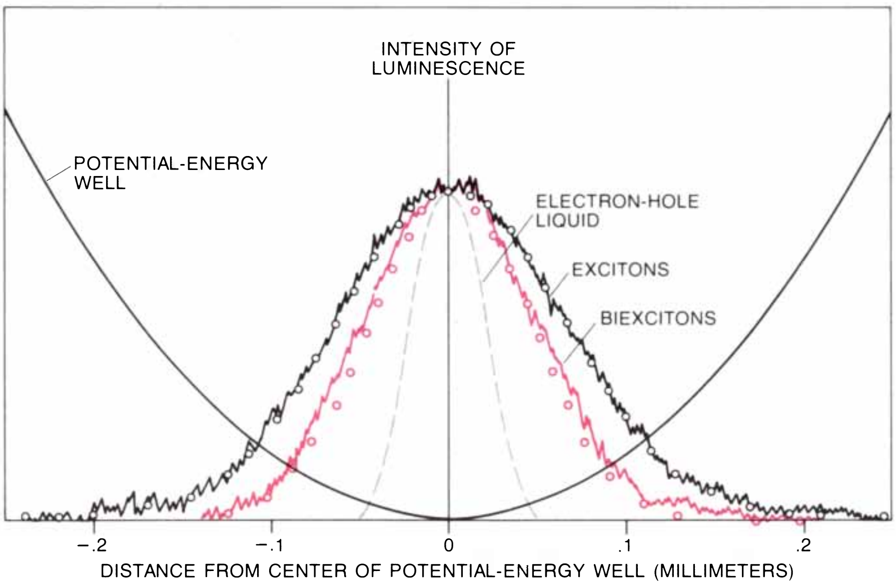
A Bose-Einstein condensate, if it exists, represents an entirely new state of matter, and so its properties may be exotic. For example, it is quite different from the spatially condensed states of matter, the solid and the liquid phase; indeed, the term condensation refers only to the aggregation of the bosons in low-energy states. The behavior of the particles is coherent over macroscopic distances, much like the coherent motions of the pairs of electrons that carry electric current without resistance in a superconducting metal.

For a given density in a gas the smaller the mass of the particle, the higher the critical temperature for Bose-Einstein condensation. The critical temperature for atomic hydrogen, the lightest element in the periodic table, is slightly less than .1 degree K. for a gas density of 10^{19} atoms per cubic centimeter. At that density Bose-Einstein condensation is generally precluded for hydrogen because the hydrogen atoms readily combine to form diatomic molecules. Nevertheless, promising efforts are being made to generate a high-density gas of atomic hydrogen by aligning the spins of the atoms magnetically. Nearly all other elements solidify before reaching the extreme conditions of temperature and density necessary for Bose-Einstein condensation. Because of the small mass of the exciton, however, Bose-Einstein condensation of an exciton gas whose density is 10^{19} particles per cubic centimeter is expected to take place at approximately 100 degrees K., a much higher temperature than that expected for any atomic system.

There are now several sources of evidence for the Bose-Einstein condensation of an exciton or biexciton gas. If the density of excitons in cuprous oxide at two degrees K. is gradually increased, the luminescence spectrum begins to depart from the spectrum predicted by the classical distribution of the particle velocities. In contrast the velocity distribution obtained by assuming the excitons are on the threshold of Bose-Einstein condensation fits the data quite well. Workers in the U.S.S.R. also report that the range of emitted photon energies signaling exciton recombination in germanium narrows as the exciton density is increased. The finding is consistent with predictions for the behavior of a gas of bosons near the threshold of Bose-Einstein condensation.

Finally, at high densities a gas of biexcitons in copper chloride shows anomalies in its luminescence spectrum. Biexcitons are created directly in a low-energy state by resonant-light excitation. At temperatures below 30 degrees K. and densities above 10^{18} particles per cubic centimeter a sharp recombination line appears in the spectrum, which indicates that most of the biexcitons are decaying from a low-energy state. If this system is probed by adding biexcitons to it through nonresonant excitation, the biexcitons are drawn into the occupied low-energy state. On the other hand, if there are no resonant biexcitons already present in the system, the probe biexcitons acquire a non-quantum-mechanical velocity distribution.

Although this result strongly suggests a Bose-Einstein condensate is present in the crystal, it can still be argued that the effects are observed only because the initial resonant excitation forces all the biexcitons into the same low-energy state. What one would like to see, and what has not yet been observed, is the Bose-Einstein condensation of a gas that is initially in thermal equilibrium. It would not be surprising, however, if an exciton gas became the first physi-

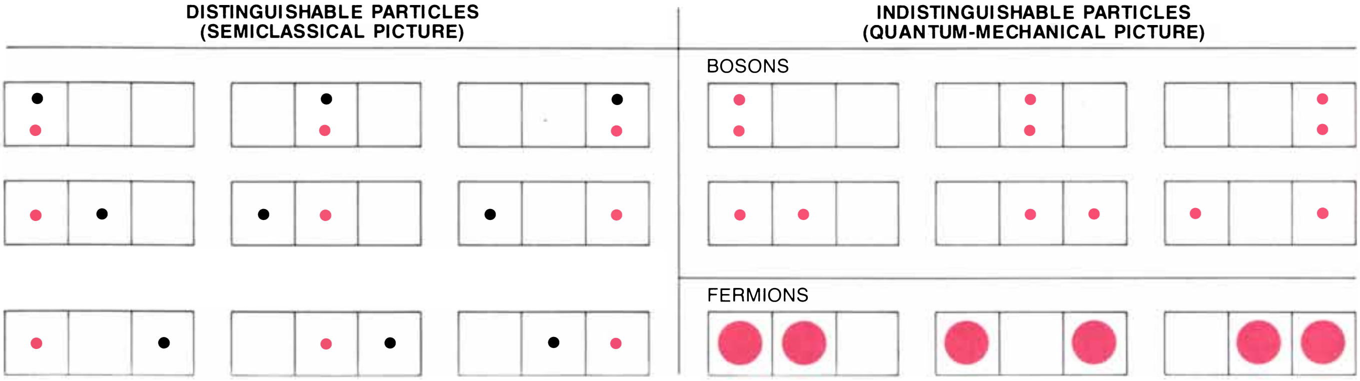


VOLUME OF THE LIGHT-EMITTING REGION of the potential well associated with each peak in the luminescence spectra shown in the illustration on the preceding page confirms the presence of the biexciton gas and the electron-hole liquid. The graphs show that exciton luminescence is emitted from the greatest volume. Because the mass of the biexciton is twice the mass of the exciton, it is predicted that the luminescence from biexciton decay (*open circles in color*) should be emitted from a region whose diameter is $\sqrt{2}/2$, or about .7, times the diameter of the region associated with exciton decay (*open circles in black*). The graphs show the data are in accord with the prediction. The volume of the region that emits luminescence characteristic of the electron-hole liquid is even smaller (*broken curve*), which indicates condensation from the gas to the dense, liquid phase. The data were collected by Gourley and Wolfe.

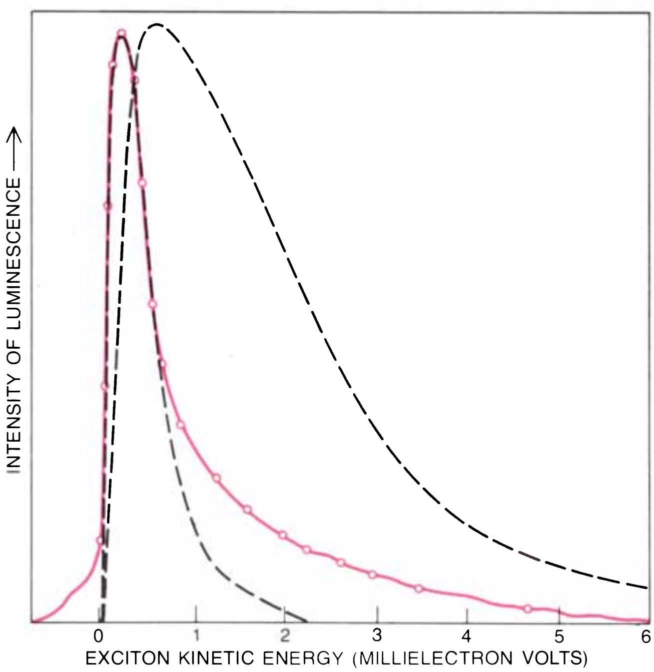
© 1984 SCIENTIFIC AMERICAN, INC

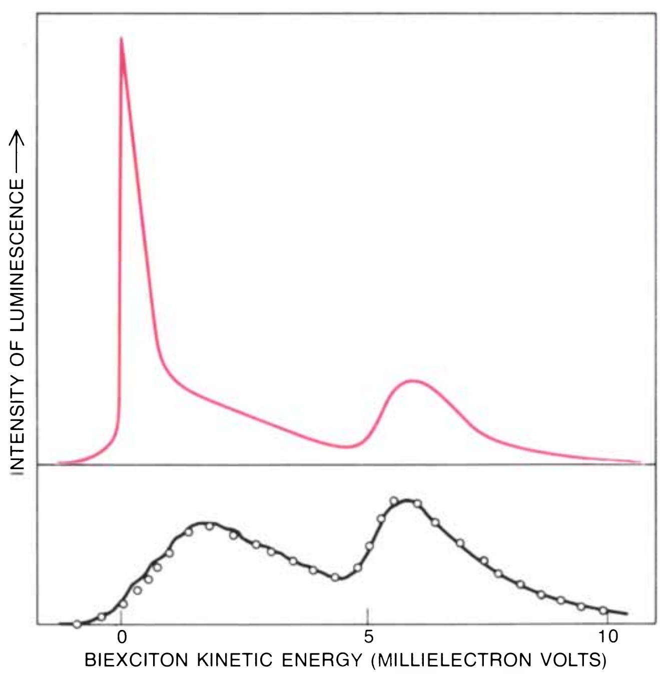
cal system to exhibit the condensation unequivocally.

The study of exciton interactions explores domains of temperature, pressure, light intensity and time scale that are currently at the limits of experiment. Moreover, excitons are observed in a wide variety of materials: in addition to the semiconductors and the biological systems we have mentioned they are found in insulators, in organic crystals and in the manmade periodic structures called superlattices. Newly developing tools such as lasers capable of generating pulses as brief as a fraction of a picosecond will enable workers to probe the dynamics of excitons with even better resolution in time. In such domains new phenomena are almost sure to be found.



INDISTINGUISHABILITY OF ELEMENTARY PARTICLES accounts for the difference between the quantum-mechanical and the semiclassical, or non-quantum-mechanical, distributions of particles among available energy states. In the semiclassical distribution each particle must occupy one of several quantized energy states, but because the particles are distinguishable, they can be distributed among the energy states in more ways than they can in the quantum distributions. For example, if two distinguishable particles can both fit in any one of three states, there are nine ways they can be arranged. If only one particle can fit in a state, there are six ways to arrange the particles (*left*). In quantum distributions all particles of a given kind are indistinguishable. For bosons, which can occupy the same energy state, three out of the six possible arrangements of the particles are made up of multiply occupied states, a higher proportion than three out of nine (*top right*). At low temperatures, for which the number of particles is comparable to the number of available energy states, this property of the quantum distribution of bosons effectively confines a higher proportion of the particles in the states of lowest energy than the proportion predicted by the semiclassical distribution. For fermions, no two of which can occupy the same energy state, there are only three ways to arrange particles among the three energy states, instead of the six predicted by the semiclassical distribution (*bottom right*).





BOSE-EINSTEIN CONDENSATION may account for the anomalous shape of the two luminescence spectra shown here. The condensation is predicted from the quantum-mechanical distribution of excitons and biexcitons, which are both classified as bosons, among the energy states available to them at low temperatures. At the left is the luminescence spectrum for a dense gas of excitons in cuprous oxide. The distribution of the particle velocities that can be inferred from the spectrum (*color*) closely matches the distribution predicted by Bose-Einstein statistics (*open circles*). The semiclassical distributions shown for two possible temperatures (*black*) can fit portions of the observed spectrum, but the overall match is not as precise as it is if Bose-Einstein statistics is assumed. The data were collected by Hulin, Claude Benoît à la Guillaume and Mysyrowicz at Paris. At the right is the luminescence spectrum for a biexciton gas in copper chloride (CuCl). The black curve is the spectrum obtained from a relatively rarefied gas of biexcitons generated by a laser at low power, and it agrees well with the velocity distribution of the biexcitons that is calculated from a semiclassical distribution (*open circles*). The colored curve is the spectrum obtained at higher laser power; it shows a large increase in the biexciton luminescence at low energy. The spectral peak is sharp, which indicates all the biexcitons have nearly the same kinetic energy. Such a spectrum is expected from a Bose-Einstein condensate. The data were collected at Indiana University by Lloyd L. Chase, Nasser Peyghambarian, Gilbert Grynberg and Mysyrowicz.

© 1984 SCIENTIFIC AMERICAN, INC

Muscle Sounds

Contracting muscle generates distinct sounds that are not heard only because the human ear is insensitive to low frequencies. Such sounds are now studied for their possible usefulness in science and medicine

by Gerald Oster

In an awake human being the muscles of the body are continually contracting and relaxing. Strange as it may seem, as they do so they generate sound. Since this sound is not heard under ordinary conditions, you may be skeptical. If you are, put both thumbs gently in your ears and make a fist. You will hear a low rumble. The tighter you make the fist, the louder the sound will be. What you are hearing is the sound made by the muscles of the forearm as they contract.

Muscle sounds are normally inaudible because the human ear is insensitive to low-frequency sound. The main frequency of muscle sounds is 25 hertz (cycles per second), which is near the lower limit of hearing. Although the existence of such sounds has been known for centuries, physiologists and physicians have for the most part considered them a nuisance. Recent developments in computer science and medical instrumentation, however, have made it possible to investigate muscle sounds in detail. The results are most interesting. Work I have been doing with the new methods suggests the sound is generated by the muscle fibers as the muscle contracts. It could well be that such effects will yield a better understanding of muscle physiology, of muscle pathologies (including those of the heart) and even of certain kinds of animal communication.

The first account of hearing muscle sounds through the thumbs in the ears was written by Francesco Maria Grimaldi, an Italian Jesuit, in his book *Physicomatheis de lumine.* That work, published two years after Grimaldi's death in 1663, is a treatise on light. Grimaldi is best known for having discovered the diffraction of light, but he was also interested in acoustics. He attributed the rumbling muscle sound to "the hurried motions of animal spirits." Ever since the time of Galen, the second-century Greek physician, it had been generally agreed that voluntary movements were governed by a fluid—animal spirits—emanating from the brain. After Grimaldi's death the subject of muscle sounds was largely ignored for a century and a half.

It was taken up again by the British physicist, chemist and physician William Hyde Wollaston. In the Croonian Lecture to the Royal Society of London in 1810 Wollaston compared the sound in Grimaldi's demonstration to the distant rumble of carriages moving over the cobblestone streets of London early in the morning. Wollaston exploited the comparison in a remarkably literal manner in order to determine the frequency of the muscle sound. He had his carriage driven through the streets at various speeds until the noise matched the rumbling he heard through his thumbs. When the carriage was moving at eight miles per hour, it gave off a sound that closely resembled the one coming from the forearm muscles.

The regular design of the London streets helped Wollaston considerably. In 1810 the cobblestones of London were a uniform six inches across. Eight miles per hour is 11.7 feet per second, and so the carriage wheels were hitting a little more than 23 cobblestones per second. Wollaston concluded that the frequency of the muscle sound is about 23 hertz.

With a simple stethoscope consisting of a wood stick attached to a small pillow against which he put his ear Wollaston made two additional observations. He noted that all the muscles of the body emit the same rumbling sound regardless of their size. In addition he found the sound did not change with the force exerted unless the force was very large. With such a force the sound emitted by the muscle seemed to be of a slightly higher pitch.

After 1810 muscle sounds were again ignored for more than 150 years, except for occasional references in scientific publications to their presence as noise that interfered with listening to other sounds in the body. My interest in muscle sounds was aroused in the course of investigating congenital heart defects in infant rats whose mothers had been exposed to certain chemicals while they were pregnant. To listen to the sounds of the rat heart it was convenient to use an electronic stethoscope.

The electronic stethoscope amplifies low-frequency sounds. The rats were lightly tranquilized to make them manageable, but their legs occasionally twitched. When they did, the muscle contractions were accompanied by a loud rumble in the earpieces of the instrument.

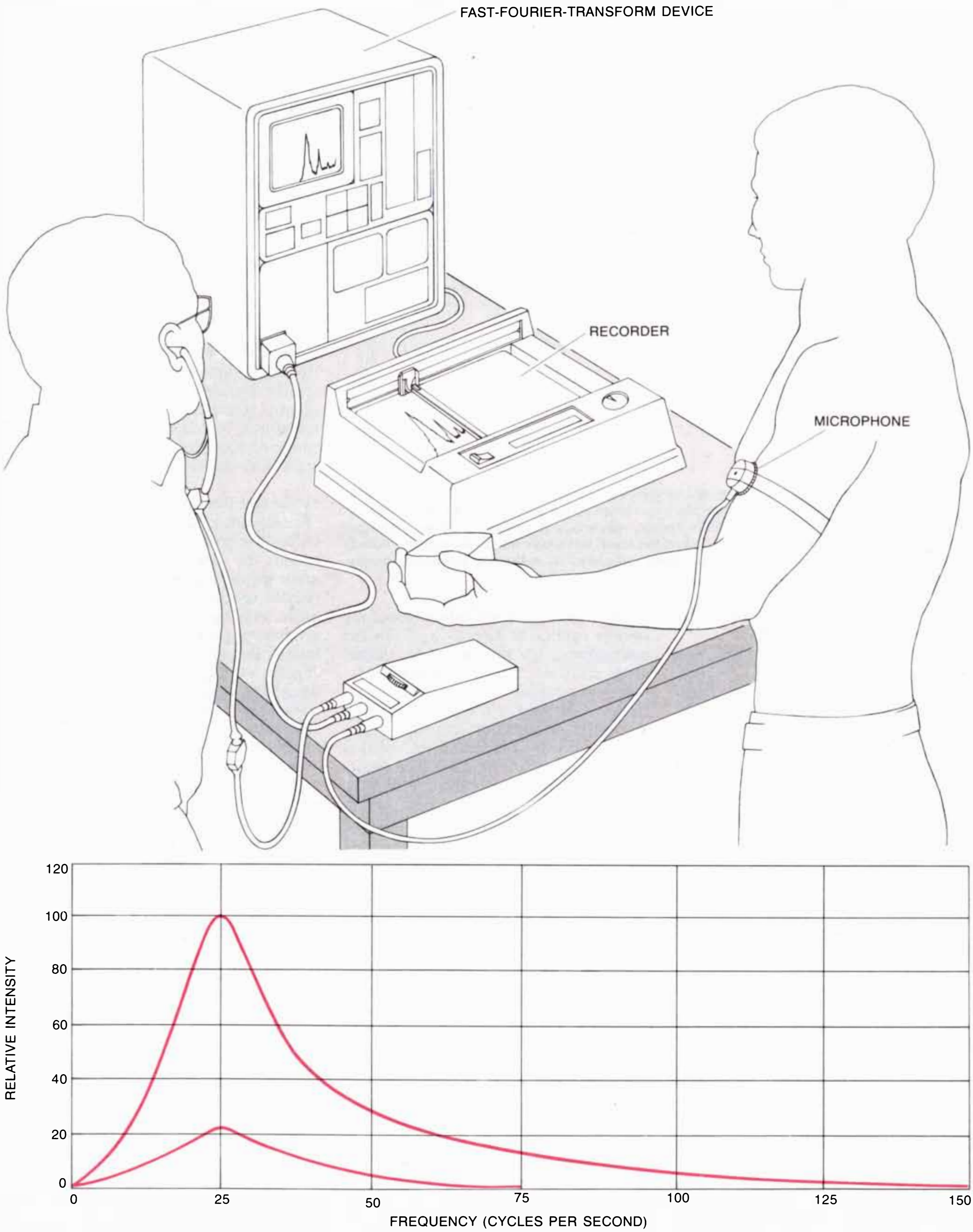
At first I supposed muscle sounds must be widely known. After studying the scientific and medical literature and finding they had been largely overlooked, I decided to study them with modern instrumentation. I enlisted as a collaborator Joshua S. Jaffe, then a student trainee at the Mount Sinai School of Medicine in New York.

It is not surprising that physicians have paid little attention to muscle sounds. The mechanical stethoscope, the device most widely employed to listen to sounds in the body, is useless for working on muscle sounds. It is valuable for detecting sounds with a frequency of from 75 to 200 hertz, such as heart murmurs or lung rales. Because of its design, however, it filters out most of the sound with a frequency of less than 50 hertz.

Moreover, the small amount of low-frequency sound that does pass through the mechanical stethoscope is almost inaudible because human hearing is not sensitive to low-frequency tones. For example, human beings are only about a ten-thousandth as sensitive to a 25-hertz tone as they are to a 250-hertz one. Most audiologists consider 20 hertz to be the lower limit of human hearing; below about 30 hertz sound loses its tonal quality and can be described by the listener only as a deep rumble.

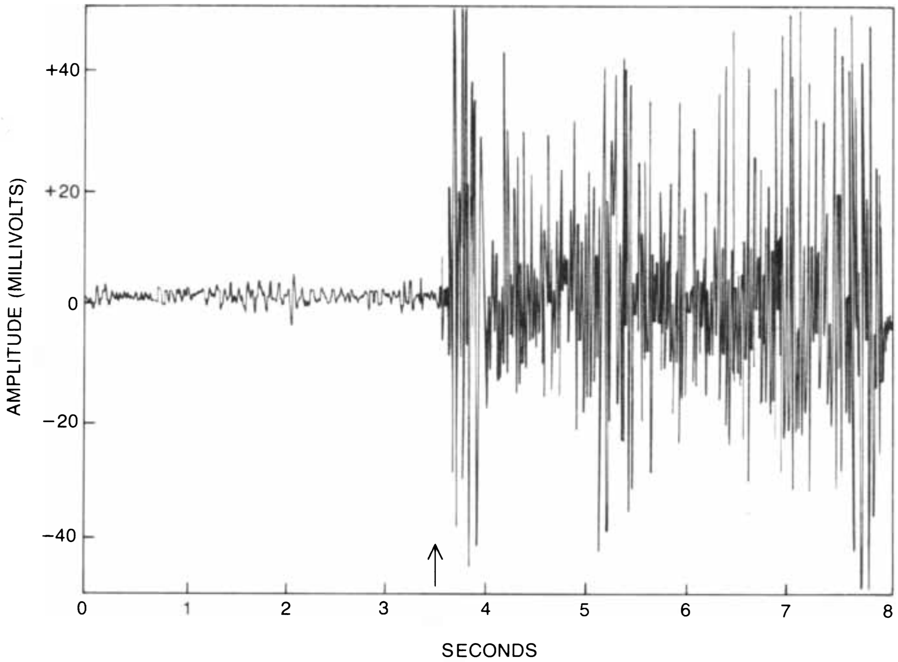
The electronic stethoscope is an excellent means of overcoming such difficulties. Its key component is a microphone that acts as a transducer to convert pressure waves, such as the waves that carry sounds to the ear, into electric signals. The microphones in most electronic stethoscopes rely on a piezoelectric

© 1984 SCIENTIFIC AMERICAN, INC



ELECTRONIC STETHOSCOPE was employed by the author to detect low-frequency sounds from contracting muscles. The upper panel in this illustration shows the setup in a typical muscle-sound experiment. The subject (*right*) holds a weight in his hand and the microphone of the electronic stethoscope is taped to his biceps. The observer (*left*) listens to the amplified muscle sounds in the earpieces of the stethoscope. The electric signal from the stethoscope is processed by the computer program known as the fast Fourier transform. The transform yields the spectrum of the muscle sound: the power density as a function of frequency. The spectrum is recorded on paper. The lower panel in the illustration shows the spectrum of two muscle sounds, one made with a five-kilogram weight held in the the hand (*top*) and the other with a two-kilogram weight (*bottom*). The sound has a maximum intensity at 25 hertz (cycles per second).

© 1984 SCIENTIFIC AMERICAN, INC



ELECTRIC SIGNAL from the electronic stethoscope changes dramatically when the muscles of the arm contract. The illustration shows a record of the signal received from a microphone attached to the flexor carpi radialis muscle, which is in the forearm. At the beginning of the trace (*left*) the arm was relaxed. When the hand was made into a fist and the muscle contracted (*arrow*), the amplitude of the signal, measured in millivolts, increased sharply.

crystal, which can convert pressure impulses directly into electricity.

A well-designed electronic transducer can respond uniformly to all frequencies up to about 1,000 hertz. The signal from the crystal is modified by solid-state circuits to amplify tones with the desired frequencies. In the work on muscle sounds all tones from 15 hertz to 100 hertz were so amplified. The resulting signal can be converted back into sound in an earpiece or can be processed to yield a visual image like the one on an oscilloscope.

Although the electronic stethoscope clearly has advantages over the mechanical stethoscope, it does have one disadvantage. It is that the device amplifies not only the sounds from the muscles but also background noises made by air conditioners, automobile traffic and even footsteps. Such low-frequency background noises are not easy to filter out. The frequency of sound waves is inversely proportional to the wavelength, and the size of the filter necessary to block the sound increases with the wavelength. Sounds with a high frequency, such as speech, jet-airplane noise or a police siren, have a short wavelength and are readily absorbed by building materials. Sounds with a frequency of less than 100 hertz are much harder to eliminate. For example, tones with a frequency of 20 hertz have a wavelength of 50 feet, and preventing them from reaching the electronic stethoscope would require a filter as large as a building.

Since such extraneous low-frequency sounds cannot be kept away from the microphone, the output of the stethoscope includes a random component—noise—in addition to the muscle sound. Hence a method is needed to separate the useful information from the noise. I turned to the mathematical method known as autocorrelation. The autocorrelation of random vibrations is a method formulated by G. I. Taylor of the University of Cambridge for analyzing turbulence in fluids. The mathematical details of the method are not relevant here, but in essence it consists in comparing a signal with the same signal displaced in time. The useful information in the signal has a regular form that is repeated over time, as is the form of the pressure waves of the muscle sound. The background sounds, however, generally show no such regularity. As a result when the signal is displaced and compared with itself, the useful information is augmented and the noninformation is canceled.

It is the autocorrelation procedure that shows the dominant frequency of the muscle sound is 25 hertz. It also shows the muscle sound does not consist of a single pure tone: it includes a range of frequencies. The spectrum of a sound that includes several frequencies can be defined as the density of power as a function of frequency. The approximate spectrum of the muscle sound reaches a maximum at 25 hertz plus or minus 2.5 hertz. The muscle sound is equivalent to random noise passing through a filter that blocks all sound with a frequency of less than 22.5 hertz or more than 27.5 hertz.

The question naturally arises of why the spectrum of the muscle sound was determined by means of the autocorrelation procedure rather than being determined directly from the data gathered with the electronic stethoscope. The reason is that the calculation of the spectrum from vibrational data such as those arising from muscle contractions would until recently have been prohibitively time-consuming and expensive even with the fastest digital computers.

The calculation of such a spectrum has now been much simplified by a computer program written by James W. Cooley of the IBM Thomas J. Watson Research Center and John W. Tukey of Bell Laboratories and Princeton University. The algorithm on which the program is based is called the fast Fourier transform. With it a million calculations can be replaced by 5,000, a 200-fold saving in calculation time.

The fast Fourier transform has had a dramatic effect on work in many scientific disciplines. In the study of muscle sounds the availability of such a program means the exact spectrum of the muscle sound can be obtained, doing away with the need to rely on the approximate spectrum given by autocorrelation. In our experiments a machine capable of doing the fast Fourier transform was connected to the electronic stethoscope. The machine computed the spectrum of muscle sound almost simultaneously with the muscle contraction.

In a typical experiment the subject (usually Jaffe) supports a lead weight in his hand with his palm turned up. The microphone can be attached to the biceps. When the weight is held steady to maintain a constant contraction, the amplitude of the muscle sound is directly proportional to the weight.

The fact that the amplitude of the sound is proportional to the load implies that the measurement of muscle sounds can be utilized to determine how hard a particular muscle is working. With a weight supported in the hand the least sound comes from the biceps when the angle between the forearm and the upper arm is 115 degrees. Weight lifters have long known this is the angle at which the arm can support the greatest weight.

When the body is in the prone position, the gastrocnemius, the largest calf muscle, yields almost no sound. When the body is in the standing position, the sound from the gastrocnemius is fairly loud, and when the weight of the body is supported on the toes, the sound is intense. When Jaffe stood on tiptoe, the intensity of the sound from his gastrocnemius was equal to the intensity of the sound from his forearm supporting 15 pounds, which was the max-

© 1984 SCIENTIFIC AMERICAN, INC

imum weight he could hold steadily in his hand.

When we measured sound from the muscles of a professional ballerina, we found she could not hold as much weight in her hand as Jaffe could. When she was supporting her weight on her toes, however, the sound coming from the gastrocnemius was much intenser than it was when Jaffe was standing on his toes. The dancer's training in standing on point had apparently led her to adopt a stance different from the one Jaffe took up. This suggests that the measurement of muscle sound can reveal which muscles are operating in a particular physical maneuver.

The soleus is a lower-leg muscle near the gastrocnemius. I probed it with a miniature microphone that had a membrane diameter of only half a centimeter. When the body was in a normal standing position, the sound from the soleus was nearly 10 times as intense as that from the gastrocnemius. The intensity of sound from the soleus is related to its physiology and also to its role in locomotion. The soleus is responsible for maintaining the angle between the leg and the foot and therefore makes a major contribution to the upright human posture.

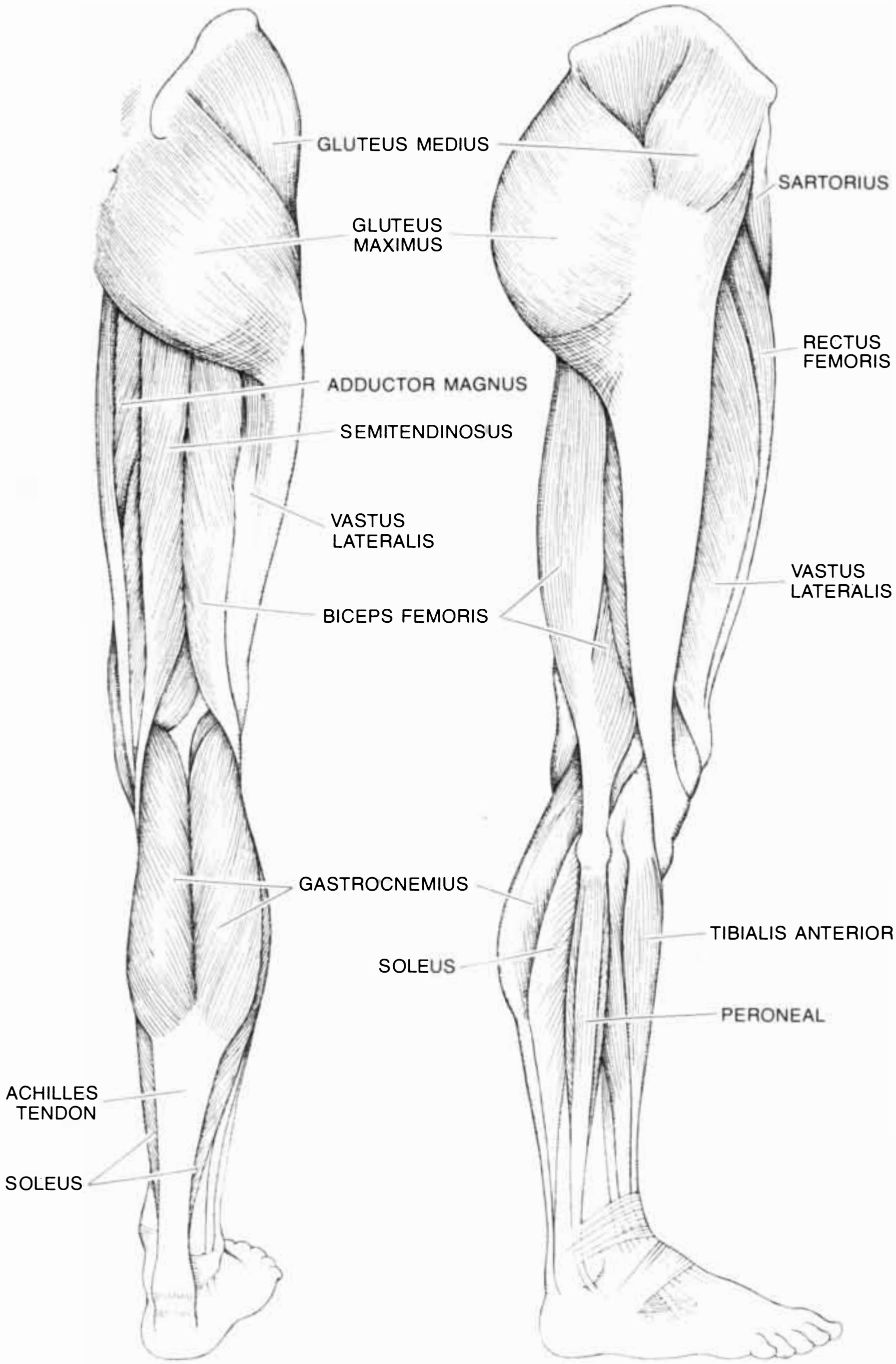
In standing, walking or running the soleus contracts many times over a long period without fatigue. The other skeletal muscles are more easily fatigued, and it is apparent that the rate of energy transformation in the soleus is considerably less than the rate in the other muscles. The endurance of the soleus results from the fact that it is richly supplied with mitochondria, the energy-transducing organelles of the cell. In the mitochondria oxidative enzymes produce adenosine triphosphate (ATP), the universal cellular fuel.

Muscle fibers that have many mitochondria are known as slow-twitch fibers; fibers with fewer mitochondria that can contract rapidly are known as fast-twitch fibers. It is obvious slow-twitch fibers are important in activities calling for endurance and fast-twitch fibers are important in activities calling for speed. The soleus, which has a preponderance of slow-twitch fibers, is more highly developed in champion marathon runners than it is in sprinters because sprinting does not call for the sustained contraction of the soleus. It is possible that a daily measurement of the intensity of sound from the soleus compared with the intensity from, say, the gastrocnemius could serve as a means of evaluating the training progress of a marathon runner.

Muscle sounds could be of interest to students of animal behavior. The sounds can readily be heard under water. If the head is under water and a fist is made near the ear, the muscle sound is faintly audible. The reason sound can be heard better in water than in air is that air is 1,000 times less dense than human tissue. Because of the difference in density, sound is transmitted poorly from one medium to the other. Since water is much denser than air, the muscle sound is transmitted more efficiently.

One of the few fishes that can be heard above the surface of the water is the purring gourami (*Trichopsis vittatus*). Several other species of fish generate sound by vibrating their swim bladder, the inflatable organ that helps to maintain buoyancy. Gouramis emit sound only when they are making violent and rapid motions associated with mating. The sounds do not appear to come from the swim bladder or from water turbulence, so that it is quite possible their source is muscular contraction.

Some sharks are attracted to low-frequency sounds. In 1963 Donald R. Nelson and Samuel H. Gruber of the Scripps Institution of Oceanography showed that lemon sharks move toward the source of a sound if the sound has a frequency of from 20 to 50 hertz. If



ANATOMY OF THE HUMAN LEG is shown in rear and side view. The two main calf muscles are the soleus and the gastrocnemius. In the normal standing position the sound emitted by the soleus is nearly 10 times as intense as the sound emitted by the gastrocnemius. The soleus is the muscle responsible for maintaining the angle between the leg and the foot. It consists mainly of what are called slow-twitch fibers, which can contract many times without fatigue.

© 1984 SCIENTIFIC AMERICAN, INC

the frequency is outside this range, the sharks are not attracted.

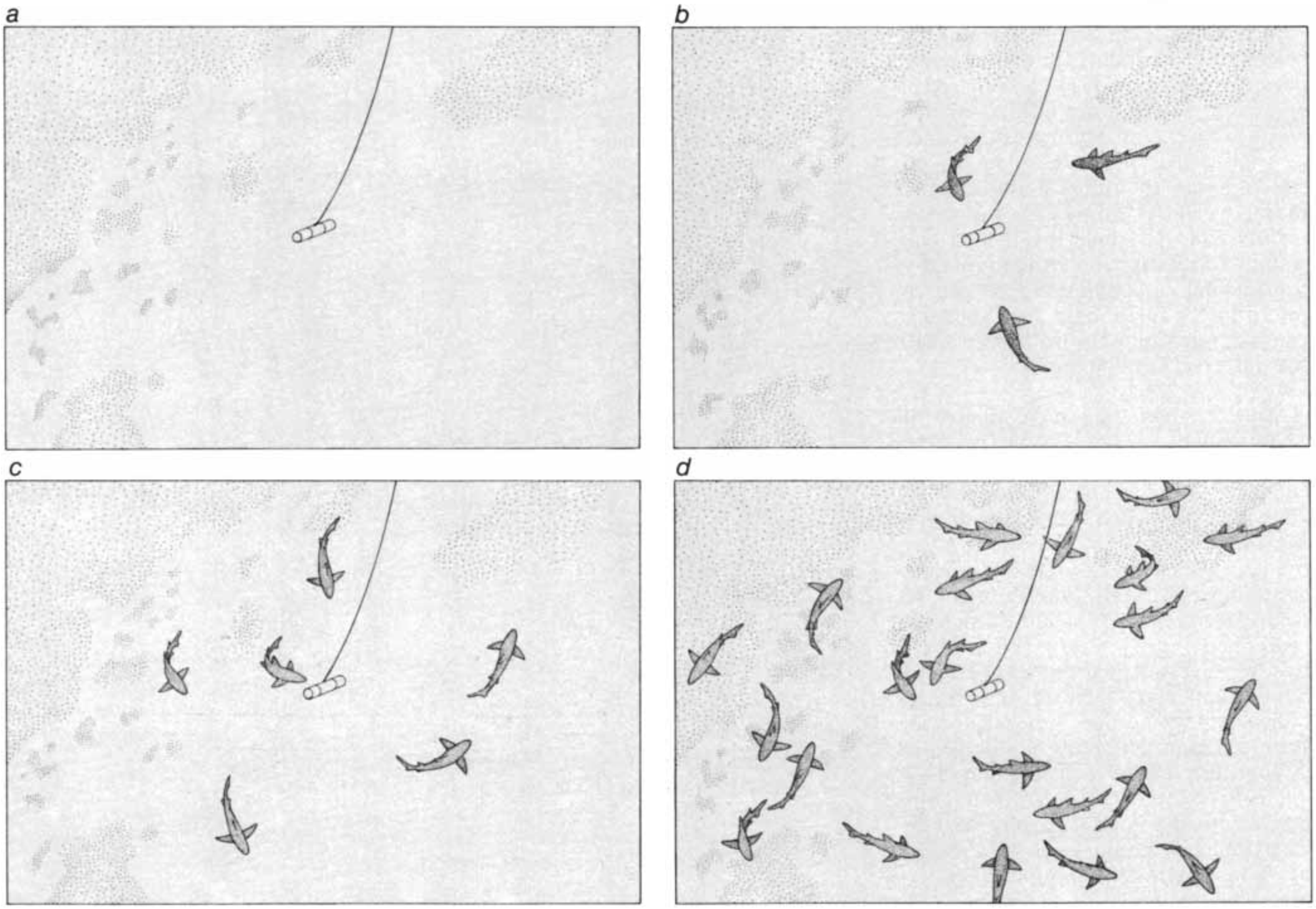
Nelson, who is now at California State University at Long Beach, has gone on to show that intermittent low-frequency sounds from an underwater loudspeaker have considerable attraction for the gray sharks that inhabit the reefs of the Pacific. The sharks also move toward the loudspeaker when it plays a recording of a fish on a hook or one of a fish in the mouth of a shark. It seems likely that the sharks are attracted to the intense muscle sounds emitted by the struggling fish.

Muscle sounds can serve as a channel of communication even among embryos. Margaret A. Vince of the Agricultural Research Council in Britain detected low-frequency sounds coming from the eggs of Japanese quail. Working with a transducer designed specifically for listening to egg sounds, Vince noted that the sound begins about four days before hatching. In order to see the embryo she made an opening in the shell and covered it with transparent plastic. Through the window she observed muscle movements when the rumbling sounds were emitted. Other evidence suggests the muscle sounds of one embryo can stimulate other embryos to start their effort to get out of the shell. Vince has proposed that muscle sounds explain the fact that eggs in contact with another egg usually hatch earlier than eggs that are not in contact.

What is the source of the muscle sounds? It appears that the rumbling sound originates in the action of single muscle fibers, in particular the fast-twitch fibers. For experimental purposes individual fast-twitch fibers can be mounted between mechanical supports. When the fibers are stretched over a fixed distance and then electrically stimulated, the individual fiber exerts a force on the supports. The fiber cannot become shorter, but a measurement of the force exerted on the supports provides a way of examining its response. Under such conditions fast-twitch fibers respond in about 40 milliseconds, or a 25th of a second. This is roughly the time required for the turnover—the utilization and regeneration—of ATP. The turnover of ATP is thought to control the chemical reactions underlying the contraction of muscle. It is notable that a 25th of a second is the period necessary for one complete cycle of the 25-hertz muscle tone.

Several workers have shown that muscle fibers vibrate when the load on the muscle is changed slightly. Clara Franzini-Armstrong, Andrew F. Huxley and F. J. Julian of University College London worked with individual fibers from frog muscle. The fibers were under tension and the load could be varied with considerable precision. The investigators showed that when the load is increased or decreased, the fibers vibrate along their long axis at a frequency of 25 hertz. The change in length is only about 1 percent of the length of the fiber and therefore can be detected only with sensitive instruments.

G. Goldspink, R. E. Larson and R. E. Davies of the University of Pennsylvania worked with the anterior latissimus dorsi, a back muscle, from 24-day-old chicks. The anterior latissimus dorsi is a fast-twitch muscle, as are all the muscles in the white meat of the chicken. (The dark meat consists of slow-twitch muscles.) Employing a technique quite different from that of the London workers, these investigators found there is a vi-



PACIFIC REEF SHARKS are attracted to recordings of low-frequency sounds played over an underwater loudspeaker. The panels of the illustration are based on frames of motion-picture film made by Donald R. Nelson and Richard H. Johnson of California State University at Long Beach. The film was shot at Eniwetok Atoll in the Marshall Islands with the loudspeaker 50 feet below the surface. When the recording began to play, there were no sharks in the area (*a*). Within 30 seconds a few sharks had gathered (*b, c*). After two minutes more than 20 sharks were circling the loudspeaker (*d*). Low-frequency muscle sounds could alert sharks to the presence of prey.

© 1984 SCIENTIFIC AMERICAN, INC

bration with a frequency of about 25 hertz in the fibers of the anterior latissimus dorsi when the tension is changed.

Such experiments strongly suggest it is the vibrations of the fast-twitch fibers that generate the muscle sound. There is another type of vibration in muscles, however, that might be thought to influence the production of sound. It is the vibration known as physiological tremor. Such tremor can be observed by pointing a finger at a word on this page. As a result of complex neurological feedback phenomena the finger will show a barely discernible oscillation of 10 hertz.

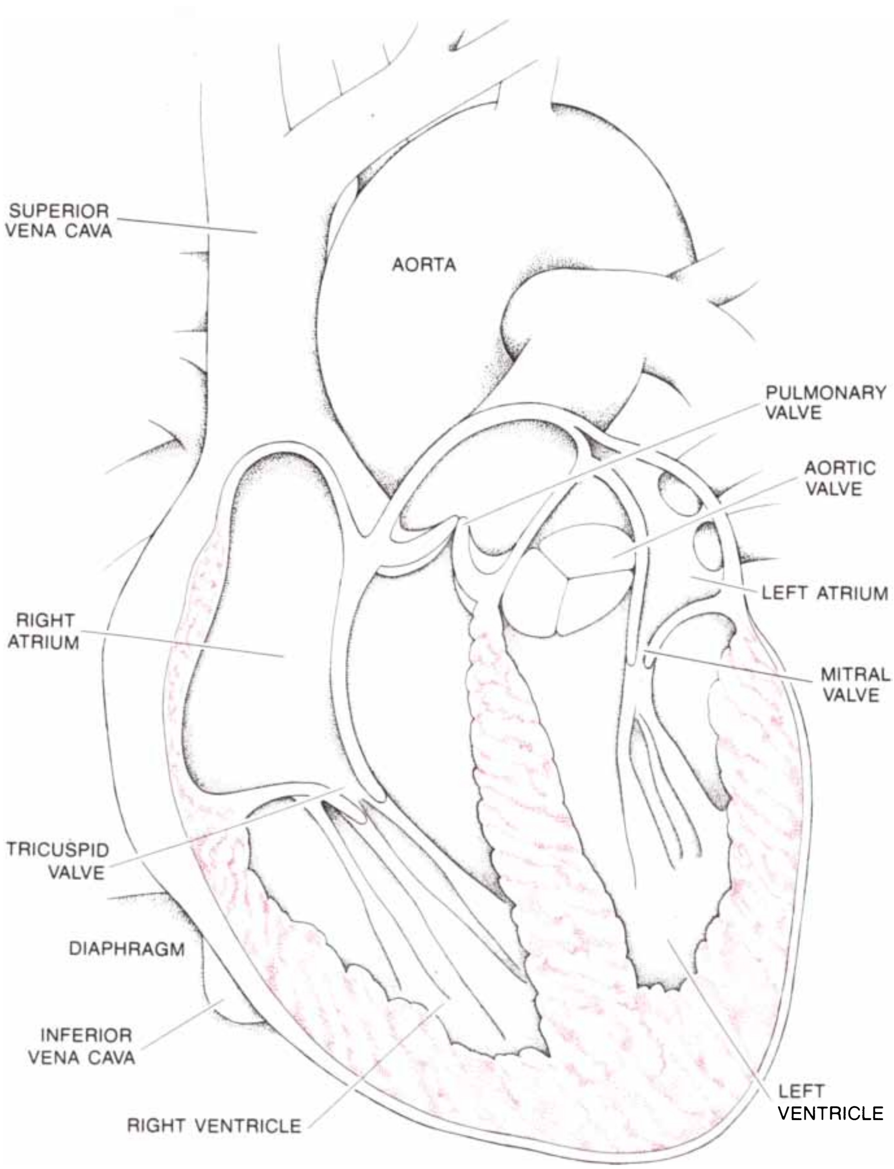
Jaffe and I decided to find out whether or not physiological tremor contributes to muscle sound. Measurements were made of the sound from an arm supporting a weight in a tub of ice water and also from an arm in which the blood flow had been greatly reduced by an inflatable cuff. Such measures eliminate or greatly diminish tremor. They had no effect on the muscle sound.

Any contribution from physiological tremor must therefore be excluded. It is logical to conclude the 25-hertz tone comes from the fast-twitch fibers alone. The many fibers in the bundle that constitutes a muscle can be thought of as a collection of oscillators. All the oscillators have the same frequency of oscillation but they are not necessarily in phase. In a particular muscle some fibers could be in phase with one another, other fibers could be completely out of phase with those fibers and still other fibers could be somewhere in between. The tensing of the muscle causes the bundle to shorten and increase in width, which sets up vibrations in the individual fibers. The motion in turn gives rise to the pressure waves that carry the 25-hertz tone.

When the muscle contracts, the continuous rumbling tone does not always begin immediately. If the muscle is exerted only slightly, what is sometimes heard is clicks, or sharp bursts of tone. As the force increases, the interval between clicks decreases until they merge into the low muscle rumble.

The clicks arise from the action of the muscle subunits called motor units. Each motor unit is made up of the group of fibers activated by a single nerve. The size of the motor unit and its mode of action can vary considerably depending on the function of the muscle.

In the gastrocnemius muscle each motor unit consists of some 2,000 fibers. In walking or running the units do not act simultaneously. The force exerted by the muscle increases slowly as new motor units are recruited. At the same time fatigued units stop functioning in order to recuperate. The alternating pattern continues as long as the muscle is contracting.



HUMAN HEART emits several sounds in each complete pumping cycle. The first heart sound marks the beginning of the systolic (contracting) phase of the heartbeat. It has been suggested that the first sound is caused by the closing of the mitral valve or by the abrupt change in pressure in the left ventricle, the more powerful of the two pumping chambers. It could be, however, that the first sound is the muscle sound emitted by the left ventricle as it contracts.

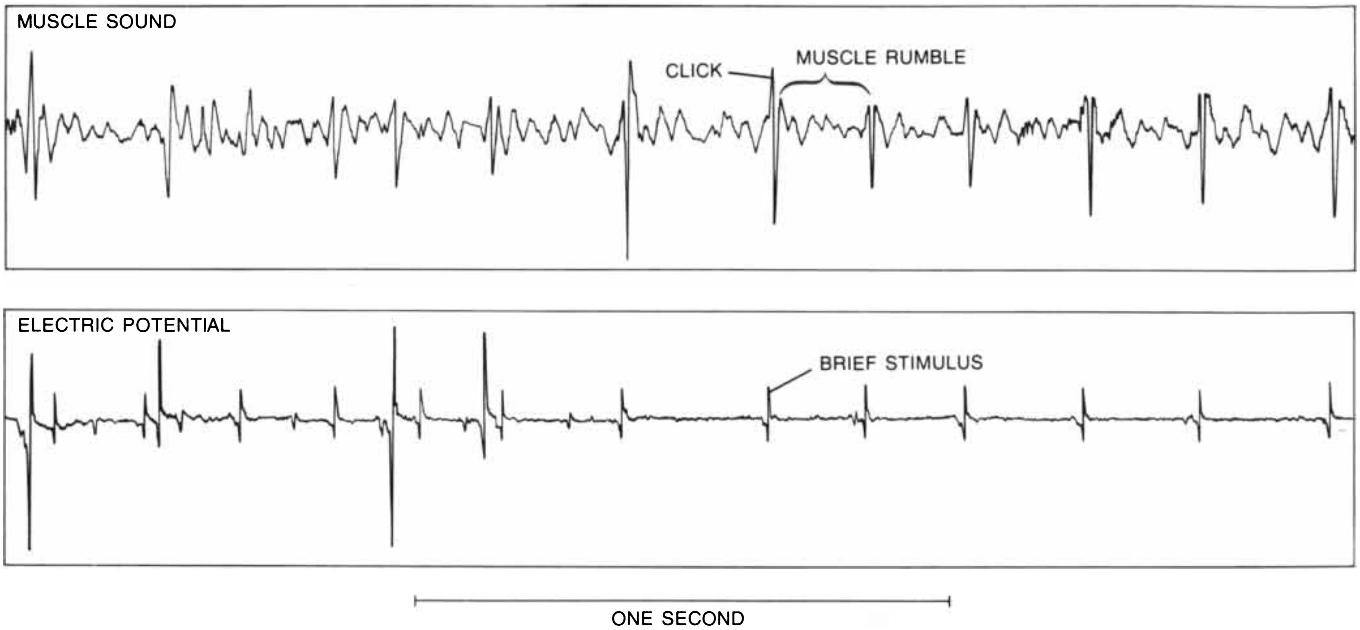
In the orbicularis oculi, which is the eyelid muscle, each motor unit consists of only some 20 fibers. In blinking the eyelid responds in about 50 milliseconds, or a 20th of a second. The response is an all-or-nothing one, with all the motor units firing together. The speed and precision of the operation require that the motor units be small.

The clicks that precede the rumbling sound were noted first in the orbicularis oculi; the observation was reported in 1948 by G. F. Gordon and A. H. S. Holbourn of the University of Oxford, who used a small microphone to study the action of the eyelid. The action of the motor units in the eyelid is so distinct that each motor unit can actually be felt independently of the others. If the middle finger is lightly applied to the eyelid, an occasional throb is felt. This is the contraction of one motor unit. It is this pulse that makes the clicking sound in the air column of the microphone.

Delbert D. Thiessen of the University of Texas at Austin has proposed that gerbils communicate by means of the clicking from the eyelid motor units. If gerbils are kept in separate cages, the animals often begin to groom themselves simultaneously. The grooming is preceded by a rapid fluttering of the eyelids. Several potential signaling mechanisms, including odor, have been ruled out. It appears possible the signaling is accomplished by bursts of sound from the orbicularis oculi.

The connection between the activation of the motor unit and the clicking sound has been confirmed by electromyography. In this technique small electrodes in the form of a needle are

© 1984 SCIENTIFIC AMERICAN, INC



BURST OF MUSCLE SOUND—a click—results from a small nerve stimulus. The upper trace is a record of sound from the vastus medialis, a large muscle in the thigh. The lower trace shows the electric potential in the same muscle. The subject was asked to think about contracting his thigh muscles but not to actually move. The result is a series of small, brief nerve impulses that appear as peaks in the lower trace. Each peak is accompanied by a click in the upper trace. The click is the sound made by the fibers of a single motor unit. A motor unit is the group of fibers activated by a single nerve. Between clicks the upper trace shows the rumble with a frequency of 25 hertz.

inserted into a muscle near a particular motor unit. By means of the electrodes the action potential that accompanies the contraction of the fibers can be detected. The action potential is a brief increase in electric potential of about a hundredth of a volt.

Gordon and Holbourn observed that in the eyelid there are fast peaks in the electromyogram that are coincident with the clicking sounds. They were not able to show limb muscles also emit clicks. Recently, however, D. T. Barry and S. R. Geiringer of the University of Michigan have recorded clicks from leg muscles. They found clicks coming from a thigh muscle of a person in a relatively quiet state. As in Gordon and Holbourn's observations the clicks came at the same time as the peaks in the electromyogram.

The study of sounds from the muscles of the heart might yield important practical benefits. The pumping action of the heart gives rise to several sounds. The two main ones are referred to as the first heart sound and the second heart sound. The first sound marks the beginning of the systolic (contracting) phase of the heartbeat; the second, the beginning of the diastolic (relaxing) phase. I found that the spectrum of the first sound has a maximum at about 22 hertz; the second sound has a somewhat higher overall pitch.

The origin of heart sounds is a controversial subject, but two of the most widely accepted conceptions of the first sound are the valve-snap hypothesis and the water-hammer hypothesis. The valve-snap hypothesis was proposed in 1830 by Pierre Rouanet, a New Orleans physician. He attributed the first sound to reverberations from the abrupt closing of the mitral valve between the left atrium and the left ventricle. Ultrasonic studies have shown, however, that the first heart sound actually begins just before the valve closes.

The water-hammer hypothesis, put forward by Robert F. Rushmer of the University of Washington, is that the first heart sound results from a mechanism like the one making a loud bang in a pipe of a badly designed plumbing system when a faucet is closed too fast. (In a well-designed system there is an air chamber that prevents the water hammer.) When the mitral valve closes, the blood pressure quickly rises. This stretches the left ventricle, causing a sharp reverberation that could yield the first sound.

In such hypotheses the heart muscle itself is relegated to a minor role. The heart sound is loudest, however, when the force of contraction is greatest, suggesting the first heart sound is due to the contraction of the heart muscle without much contribution from the dynamics of the mitral valve or the left ventricle.

This hypothesis is supported by experiments with frogs. I put a miniature microphone in direct contact with the ventricle of a beating frog heart. In spite of the large difference in mass between the frog heart and the human heart, the muscle sounds from the two are quite similar. The water-hammer hypothesis would predict a much higher frequency for the frog-heart muscle sounds because when a small ventricle is stretched, the reverberation should generate a higher-pitched sound than the reverberation of a large ventricle does. The water-hammer hypothesis is further weakened by the observation that when the blood supply to the frog heart is clamped off, the heart sound is not affected. It would therefore seem that the chief source of the first heart sound is muscle contraction.

Since more than a third of all deaths in the U.S. are caused by degeneration of the heart muscle, heart-muscle sounds could become a valuable diagnostic tool. The analysis of the detailed spectrum of the sound from the muscle could reveal the state of the tissue and thereby diagnose a myocardial infarction: heart tissue that has died because of an interruption in its blood supply. Even in the mechanical stethoscope the first heart sound has an abnormally low pitch in people with an advanced infarction; to obtain a more objective record a fast Fourier transform of the heart sound should be made.

Apart from such clinical possibilities, the study of muscle sounds might well make a contribution to the understanding of the basic mechanism of muscle contraction. For example, F. V. Brozovich and Gerald H. Pollack of the University of Washington have found frog muscle emits sounds in a sequence of bursts. This suggests that when a muscle contracts, it does so in small, discrete steps. If this is the case, it is at variance with the current picture that muscle contraction is a continuous process operating at the molecular level.

© 1984 SCIENTIFIC AMERICAN, INC

Chrysler creates LASER XE. The sports car that had to outperform the competition.

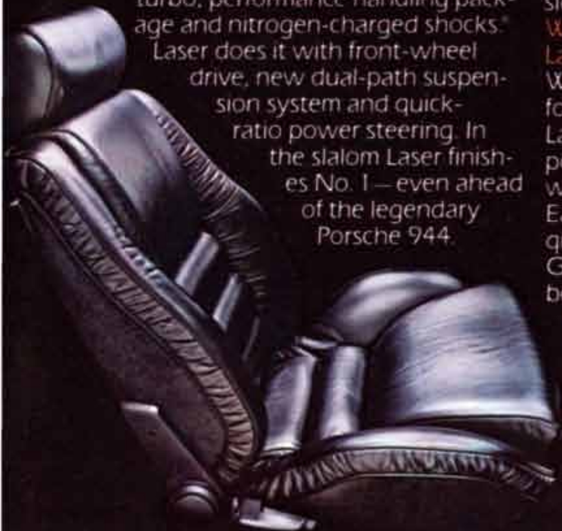
We gave Laser XE world-class performance. In the slalom, Laser beats all entries—from Trans Am to Mustang GT. We built Laser XE to outperform the competition: Camaro Z28, Trans Am, Mustang GT, Supra, Mazda RX-7.

Laser beats Porsche in the slalom.

Laser outperforms Trans Am in braking.

Laser is faster than Camaro Z28 from 0-50 mph.

Laser does it when you equip it with turbo, performance handling package and nitrogen-charged shocks.* Laser does it with front-wheel drive, new dual-path suspension system and quick-ratio power steering. In the slalom Laser finishes No. 1—even ahead of the legendary Porsche 944.



We turbocharged it. From 0-50, Laser XE leaves Camaro Z28 with its shadow. Z28 is a powerhouse—but Laser XE is the sophisticated new wave. Its multi-point injection system "spritzes" fuel in at four points. Its water-cooled bearing reduces a critical turbo temperature by 500° F. Its turbo engine boosts h.p. 45% and moves Laser like light. With 5-speed your time to 50 mph is 5.8 seconds. Z28, Trans Am, Supra and RX-7 are in your remote-controlled side-view mirrors.

We gave it high-performance braking. Laser XE stops where Trans Am doesn't. We think total performance calls for performance braking. So we gave Laser XE semi-metallic brake pads, power brakes all around and optional wide 15" alloy wheels with Goodyear Eagle GT radials. Result: Laser stops quicker than Z28, Trans Am, Mustang GT, Supra, RX-7. Even Porsche 944 can't beat our world class braking.

We gave it a brain—and a performance seat that performs. Laser XE thinks with you. Its 21-feature electronic monitor even talks your language. Its navigator computes for you while its color graphic displays make you a calculating driver. Laser XE's AM/FM stereo remembers what you like to hear and its self-diagnostic system is the nearest thing to an on-board mechanic. Your seat responds with cushions you pump up for thigh and lumbar support, and you can choose a 6-way power seat and Mark Cross world-class leather.

We gave it our best: a 5 year/50,000 mile Protection Plan. We believe a performer has to be a survivor. We build for that. And back your engine and powertrain with a 5 year/50,000 mile Protection Plan, with outer body rust-through protection for the same period.** See dealer for details. Buckle up for safety.

35 Est. Hwy
22 EPA Est. MPG

Chrysler

Laser. A Product of The New Chrysler Technology.

*Based on overall results of USAC sports car certification tests versus 1983 models. Laser XE equipped with optional handling suspension, turbo package and 15" road wheels and tires.

**5 years or 50,000 miles, whichever comes first. Limited warranty. Deductible required. Excludes leases. †Use EPA est. mpg for comparison. Actual mileage may vary depending on speed, trip length and weather. Hwy mileage probably less.

CHRYSLER LASER

"THE COMPETITION IS GOOD. WE HAD TO BE BETTER."

Lee A. Iacocca

© 1984 SCIENTIFIC AMERICAN, INC

The Adaptability of the House Mouse

The reproductive abilities of this small mammal are remarkably flexible. It is for this reason that the species flourishes in a great diversity of habitats, notably those it shares with human beings

by F. H. Bronson

The house mouse is an opportunistic commensal of human beings that has a spectacular history of global colonization. A commensal lives and feeds in close association with another species without either benefiting it or harming it directly. *Mus domesticus* began its coexistence with humans perhaps 8,000 to 10,000 years ago on the steppes of Asia. Since then, in large part through transportation unwittingly provided by people, it has established itself on every major landmass and most of the minor ones. It infests homes, commercial establishments, granaries and other outbuildings from the Equator to the subpolar regions. It can live in manmade climates as different as central-heating ducts and frozen-food lockers.

Yet the house mouse is not a strict commensal. In many places it has reverted to a wild existence, where it lives in total independence of humans. In this feral pattern it can be found in a surprising diversity of habitats: coral atolls, grasslands, deserts, freshwater marshes, seashore dunes and alpine valleys.

The outstanding success of the species in colonizing new environments has made it a subject of great interest to reproductive physiologists. To be introduced into a radically new environment is one thing; to be able to reproduce there and so to establish a new population is quite another. The planetwide spread of the house mouse in both manmade and natural habitats suggests an extreme reproductive adaptability, probably the most extreme among the mammals. Here I shall mainly discuss the uniquely flexible set of reproductive adaptations that make up the core of the house mouse's immensely successful colonization strategy. At an entirely different level, however, I shall be discussing how experiments in the laboratory can complement observations in the field and vice versa. Neither approach by itself has proved adequate in achieving an understanding of the reproductive flexibility of the house mouse.

As a background for perceiving the uniqueness of the reproductive processes of the house mouse it is useful to compare the rules governing reproduction in this species with those operating generally among mammals. Reproduction by a mammal must always proceed in harmony with three interrelated sets of environmental factors: the dietary, the physical and the social. Diet is the core environmental constraint on a mammal's reproduction. Probably the most fundamental dietary factor of all is simply the number of calories available for reproduction.

In other words, all mammals must eat. They must extract energy from their food and then they must partition the assimilated energy to serve a variety of competing needs, only one of which is reproduction. Specific nutrients such as vitamins and amino acids also must be partitioned between reproductive and nonreproductive needs.

Closely related to these dietary factors are two aspects of the physical environment. First, since energy must be expended in order to regulate the body temperature of a mammal, the ambient temperature can exert an important influence on the energy-partitioning process. Second, the daily photoperiod (the length of the daylight phase of the 24-hour cycle) can enable a mammal to predict (or sense) the approach of a season when breeding will be possible. Specific substances associated with young green vegetation can serve the same purpose for some mammals. The alternation of day and night also entrains the 24-hour cycle that pervades the reproductive activities of most mammals. Finally, a mammal's social environment often provides cues that can act either specifically to time events such as ovulation or nonspecifically to promote aversive emotional states that interfere with reproduction.

Set against this background, the first aspect of the house mouse's reproductive flexibility appears in relation to the phenomenon of seasonal breeding. Most mammals live in habitats characterized by seasonal variation in the availability of food, and as a result most of them are seasonal breeders. Most mammals also rely on a predictor such as photoperiod to increase their seasonal efficiency. The house mouse apparently does not rely on a predictor of any kind, and so it is not an obligatory seasonal breeder.

The conclusion that the house mouse does not react to variations in, say, photoperiod rests on observations made in both the field and the laboratory. Field biologists have found that the species breeds either seasonally or continuously throughout the year depending entirely on its habitat. In manmade structures in the Tropics and the Temperate Zone regions it breeds continuously, but in natural habitats it is usually found to breed seasonally.

In one sense this observation is not surprising. Human barns and granaries obviously should buffer the mice to a certain extent against extremes of climate and diet. On the other hand, the observation certainly suggests either that the species as a whole is not photoperiodic or that only house mice living in natural habitats are susceptible to such regulation. Experiments in the laboratory support the first hypothesis. Several wild stocks of house mice, including some from natural habitats, have now been examined, and all of them have been found to breed equally well whether the day is short or long.

It is interesting that at least some of the wild stocks studied in the laboratory breed at maximum efficiency even when they are maintained for months in constant and total darkness. R. C. Stoddart of the University of Glasgow first demonstrated this fact with a wild stock from Scotland. In my laboratory at the

© 1984 SCIENTIFIC AMERICAN, INC

University of Texas at Austin we have confirmed it with two North American stocks, one from Alberta and one from Texas. These results correlate well with incidental observations of house mice colonizing coal mines to depths of 1,800 feet, where the light supplied by the mine operator can be constant, unpredictably intermittent or totally absent.

Although a lack of dependence on photoperiod reveals why the house mouse is not an obligatory seasonal breeder, this fact in itself is not sufficient to explain what inhibits the mouse's breeding during some seasons in wild habitats, nor does it explain the animal's ability to breed successfully in such a wide variety of places. Understanding these phenomena requires an exploration of the potential dietary and climatic inhibitors. Unfortunately it is impossible to separate the relative importance of calories, nutrients and temperature variation under natural conditions because they all vary simultaneously with both location and change of season. Hence the need for controlled experiments in the laboratory arises again.

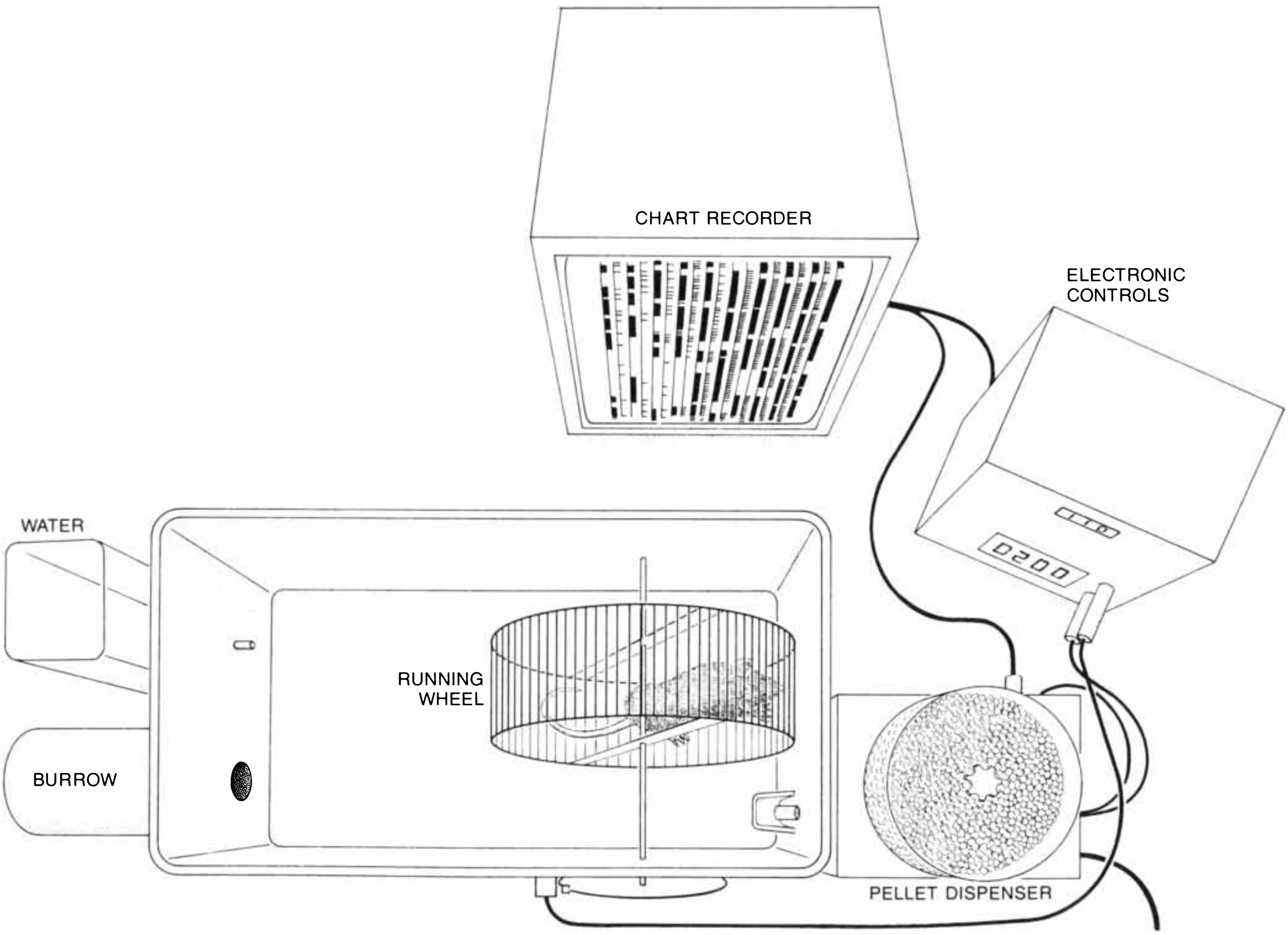
The factors that normally limit the breeding of house mice have been a concern of my laboratory for some time. One line of research—the one dealing with the bioenergetic partitioning I have already mentioned—has been particularly revealing with regard to the colonization strategy of the species.

My colleagues and I view the physiological partitioning of energy in the form of an equation where the number of calories assimilated from food must equal the number partitioned among several competing demands: body maintenance, growth, reproduction, thermoregulation, storage in the form of fat and the locomotor costs of obtaining the food in the first place. An increase in any of these demands can be satisfied only at the expense of one or all of the others or by increasing the intake of food.

Our objective has been to study the mouse's ability to reproduce while we manipulate the equation in ecologically meaningful ways. The amount of food available to an animal can be varied, the animal can be made to forage for a longer or shorter time for its food, or the temperature at which it forages can be varied. The competitive partitioning of energy can then be examined with a view to determining the potential limits on the species' reproduction.

With this approach we have now established several fundamental principles concerning reproductive success in house mice. First, failures to reproduce, whether they are attributable to the season or to the locality, probably often reflect simple imbalances in the bioenergetic equation. Second, the imbalances seem to affect reproduction adversely only in females. Finally, house mice of both sexes exhibit some surprisingly broad tolerances in certain components of the equation.

The conclusion that the female plays the critical role stems from simple experiments on the availability of food.



TEST APPARATUS provided a means of examining the energetics of reproductive development in the house mouse. The purpose of the apparatus was to simulate conditions the mouse would encounter in natural habitats and to isolate the variables that might be relevant. In this cage the animal can live in a thermally buffered burrow (the insulated can at the left) but must enter the open area and run a predetermined number of revolutions on the running wheel in order to obtain a 45-milligram pellet of food from the dispenser at the right. A chart recorder (*top*) keeps track of the number of revolutions of the wheel and the number of food pellets dispensed. The box at the upper right contains electronic controls for regulating the apparatus. The equipment made it possible to examine the relations among temperature, foraging, growth, fat deposition and reproductive development. The test apparatus was designed by the author and Glenn H. Perrigo.

© 1984 SCIENTIFIC AMERICAN, INC

Whenever the amount of food given to a weanling female is restricted, her reproductive development suffers. If her growth is inhibited to any degree, she either fails entirely to develop sexually or she does so with abnormal slowness. Males, in contrast, show absolutely normal reproductive development even when their food is restricted to the point where their growth is completely inhibited. Indeed, young males can even be made to lose weight and many of them will still mature reproductively.

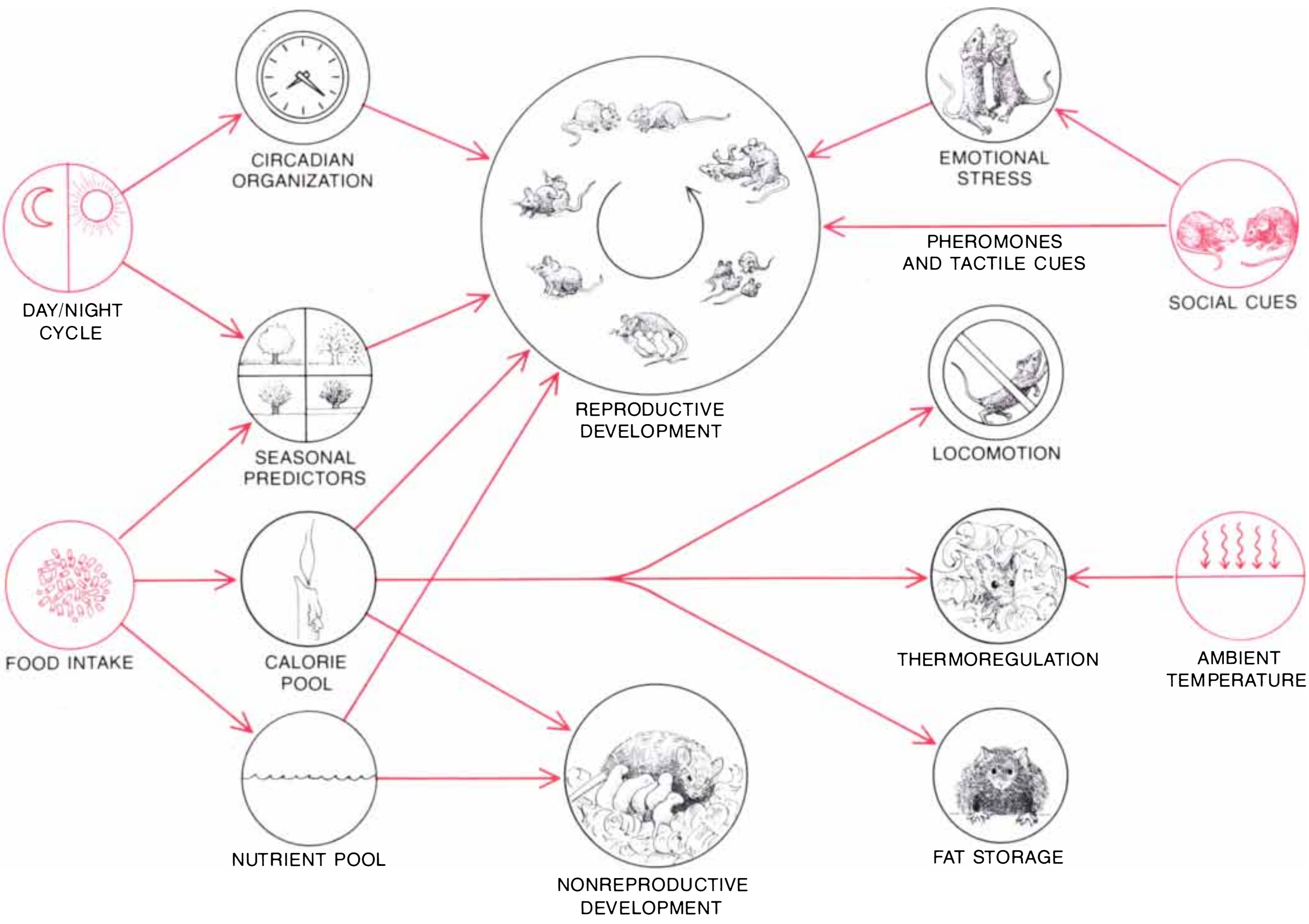
In terms of the bioenergetic equation, then, growth has a higher competitive priority than reproductive development in females but the reverse is true in males. The result is a situation in which a scarcity of food controls the reproduction of a species by acting only on one sex: the female.

Food-restriction studies of this kind are of only limited value in research with broad objectives such as ours. A wild animal seldom faces a food shortage in the absence of climatic change, particularly a change in the ambient temperature, and the animal never faces a food shortage without the ability to forage longer or harder to get more food. In pondering how to duplicate such complex conditions better in the laboratory, Glenn H. Perrigo and I developed a unique caging system. Besides giving us the first significant simulation of the energetic challenges faced by a small mammal in the wild, the caging system has enabled us to dissect the crucial bioenergetic equation that controls reproduction in the female. Hence it has enabled us to uncover what is probably the most fundamental environmental determinant of reproductive success operating in the house mouse: the relation between the availability of food (and hence the time required to forage for a maintenance diet) and the ambient temperature in the female's foraging environment.

The first key to understanding the importance of this relation lies in a simple acknowledgement of the small size of the house mouse. As has been recognized for some time, a small mammal has a large surface area with respect to its body mass, and so it loses heat rapidly when it is exposed to low temperatures. The second key is the realization the house mouse can build a nest that offers a high degree of thermal buffering, and therefore it has to expose itself to outside temperatures only during its daily foraging expeditions.

When food is scarce, the mouse must spend more time foraging. Whether or not this is an energetically profitable venture depends entirely on the temperature at which the animal must forage. Obviously there will be a set of threshold conditions, determined either seasonally or geographically, where food scarcity and low temperature combine in such a way that successful foraging, and hence life, are made impossible for the mouse. As the environmental conditions approach this threshold the bioenergetic equation becomes progressively dominated by the need for thermoregulation, and the first of the competing demands to suffer is the female's reproductive capacity.

The overwhelming importance of this relation (availability of food, duration of foraging and ambient temperature) for reproduction emerged dramatically from one of Perrigo's first experiments



ENVIRONMENTAL FACTORS regulating reproduction in a typical mammal are summarized. The environmental factors are the day-night cycle, the food intake, the social cues and the ambient temperature. The interacting functions they regulate make up the rest of the diagram. A central concern of the author's experiments was the partitioning of energy among competing functions, including growth, reproduction, thermoregulation (the control of body temperature), the storage of fat and the animal's locomotor costs in getting food.

© 1984 SCIENTIFIC AMERICAN, INC

BREAKTHROUGH: COMPUTER GRAPHICS THAT CREATE MODEL PATIENTS FOR SURGEONS.

We're using computers to give reconstructive surgeons a startling new perspective on their work—three-dimensional images that depict the patient's face and skull from any angle, inside or out.

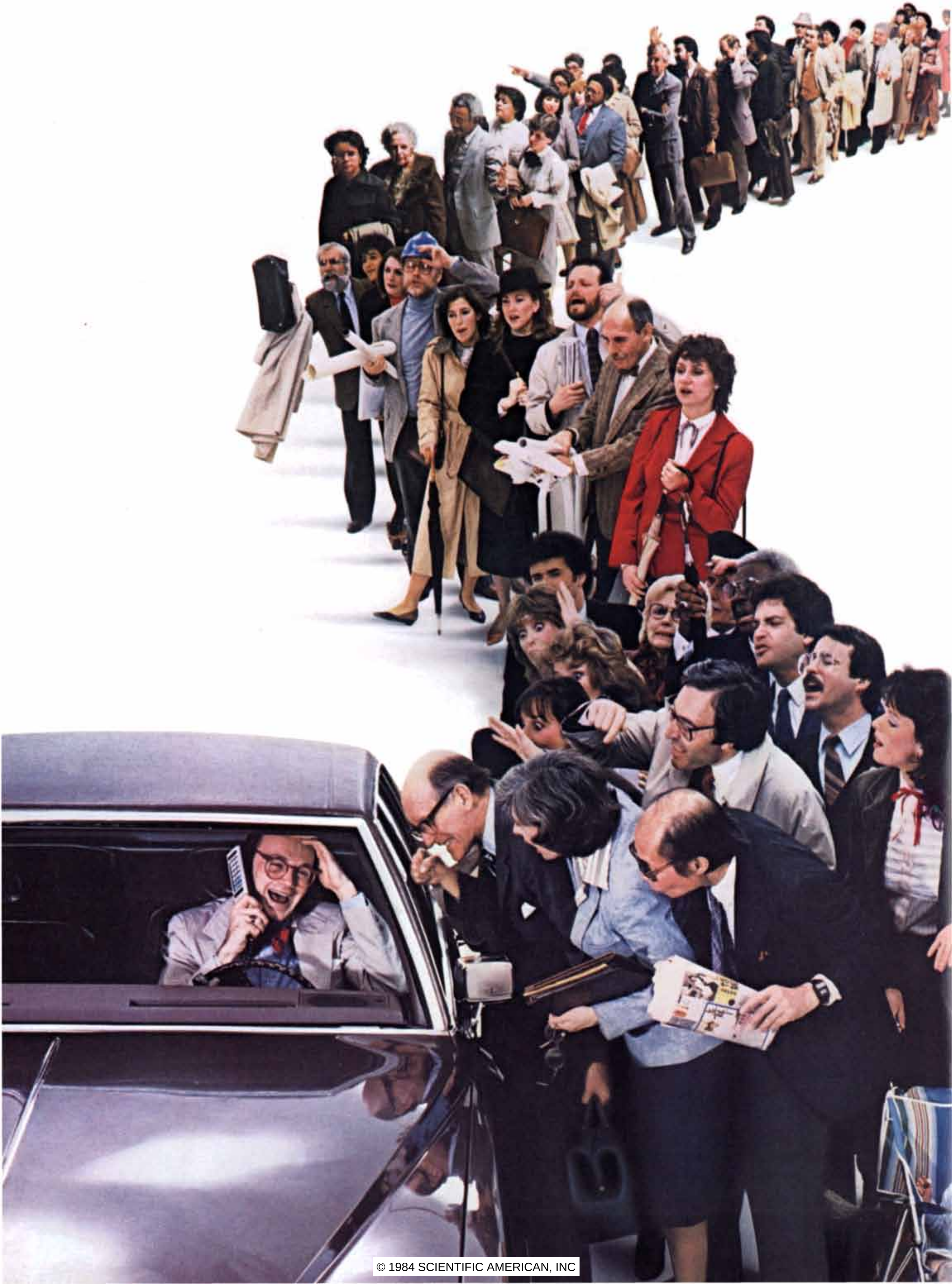
These 3D images show key relationships between bone and soft tissue. They help surgeons predetermine precisely how bone and skin grafts should be placed to achieve the desired results. They can be used to create physical models that let a surgeon perform a trial procedure or envision the results of his work before surgery begins.

Our breakthrough in computer modeling helps surgeons correct birth defects and undo the damage of disfiguring accidents.

We're creating breakthroughs not only in health care and information handling but also in communications and equipment leasing.

We're McDonnell Douglas.

© 1984 SCIENTIFIC AMERICAN, INC



© 1984 SCIENTIFIC AMERICAN, INC

The end of waiting in line for a line.

Ever tried waiting in line for a phone? When there are 5,000 people ahead of you? For years, that's been the problem with conventional mobile phone service: a limitation of 24 channels per coverage area, resulting in a maximum of about a dozen simultaneous conversations and only 700 customers for a city the size of New York.

But twelve years ago, AT&T Bell Laboratories developed the cellular radio concept.

And today, that concept has been developed into the WESTERN ELECTRIC® AUTOPLEX™ system. It allows telephone companies and others to give almost anyone access to high-quality mobile phone service.

In fact, the AUTOPLEX system is switching calls right now in Chicago. The first commercial operation in the country.

Advanced computer hardware and software make AT&T cellular technology possible. They allow the company offering mobile or portable phone service to divide a city into a honeycomb of cells. Each cell has a low-powered transmitter and receiver tied into the local phone network. Because the cells are small, radio frequencies used in one of them can be used again and again in others, making more channels available.

As you move across the city, AT&T electronic control equipment will pass your call from cell to cell. Without missing a syllable. And with AT&T cellular telephones transmission quality is as good as at home or in the office.

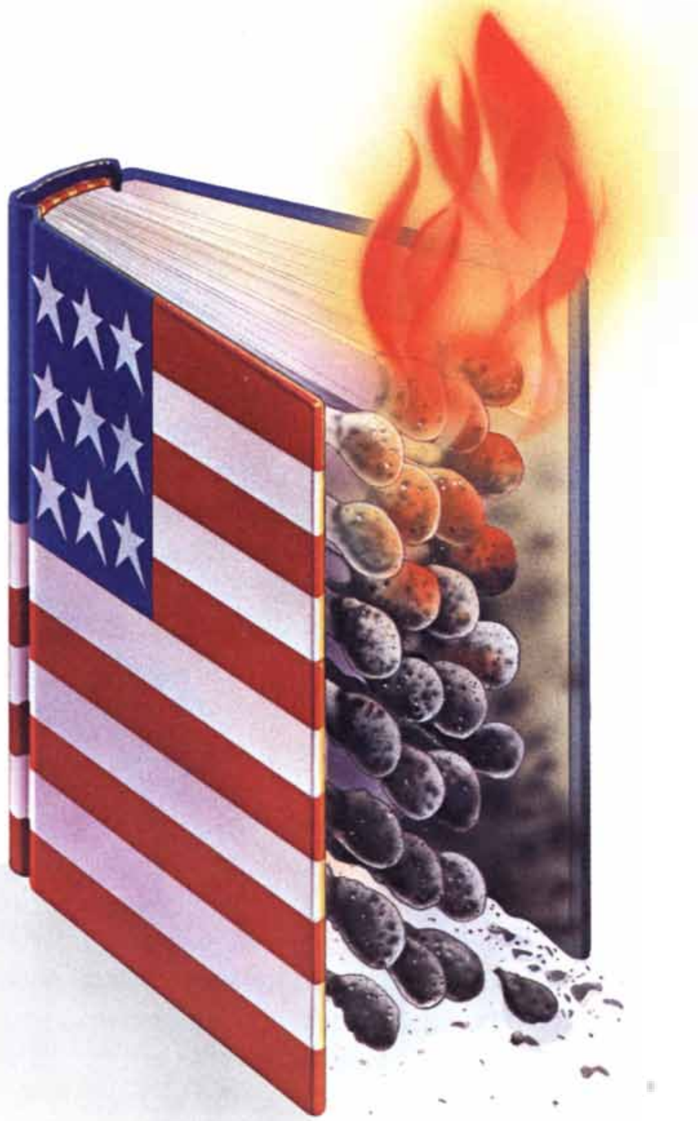
Meanwhile, AT&T Bell Laboratories is continuing to pave the way for entirely new forms of portable telecommunications.

AT&T. Applying the technologies of microelectronics, lightwave and software to make the dream of the Information Age a reality.

©1984 AT&T Technologies, Inc.

AT&T

© 1984 SCIENTIFIC AMERICAN, INC



CENSORSHIP IN A FREE SOCIETY. IT'S A BAD MATCH.

Censorship is the greatest tragedy in American literature. It constricts the mind, teaches fear and leaves only ignorance and ashes.

Today, all over the country, books are being banned, burned and censored. Teachers, students, librarians, and book and magazine publishers are being harassed. The attacks of these self-appointed censors are endorsed by our silence.

The freedom to read is one of our most precious rights. Do something to protect it.

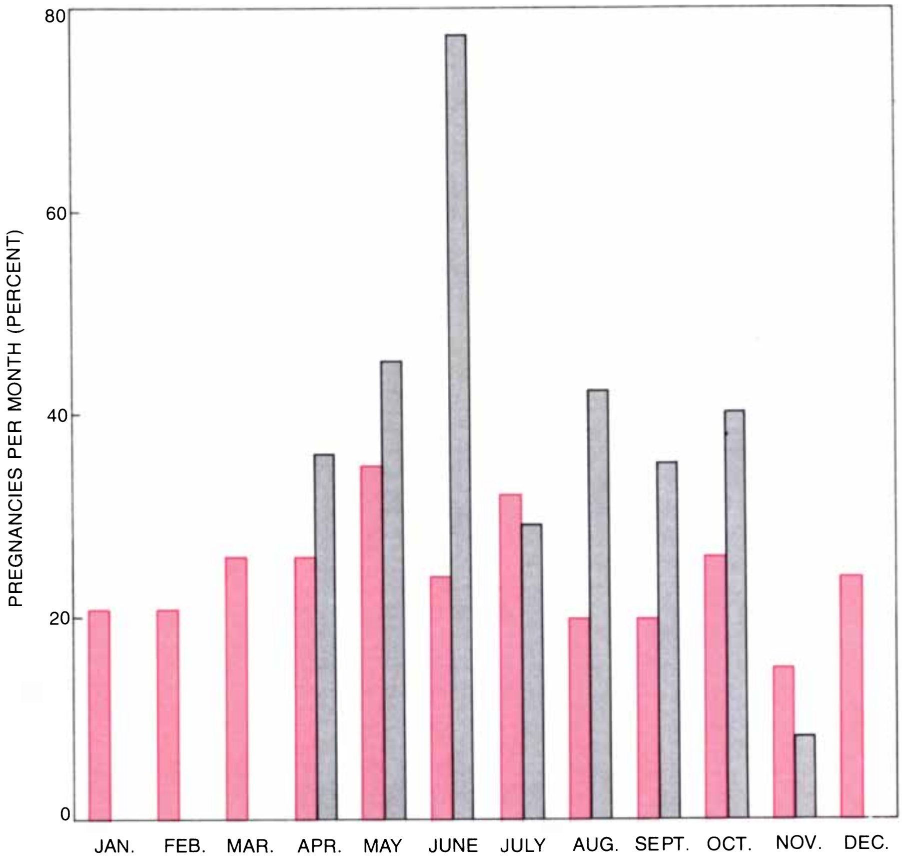
Contact:
People For The American Way, P.O. Box 19900
Washington, D.C. 20036 or call 202/822-9450.

People For The American Way

© 1984 SCIENTIFIC AMERICAN, INC

with our new caging system. Weanling females were adapted gradually to a moderately low temperature (10 degrees Celsius rather than the laboratory normal temperature of 23 degrees, or 50 degrees Fahrenheit rather than 73) and then were given excess food or were required to run either 100 or 200 revolutions on the running wheel for each pellet of food they received. These conditions mimic what mice in many natural habitats face normally with each change of season. Winter is a time of food shortage, and the mouse must forage longer and harder (and at lower temperatures) to get the same amount of food it could get easily and more comfortably in the summer.

A house mouse living at 10 degrees C. needs about 50 percent more food than it does at 23 degrees if it is to gain weight normally. This is the extra cost of thermoregulation that must be inserted into the bioenergetic equation of the female. Since the factors in the equation are competitive, the result is an increase in the time required to reach puberty. At 10 degrees our females take several more weeks to attain their first ovulation, even when excess food is available (a fact previously explored much more extensively by S. A. Barnett of the Australian National University).



EFFECT OF HABITAT on the breeding of house mice is evident in a comparison of mice living in houses and human outbuildings in Oxford, England (*color*), with mice living in a salt marsh near San Francisco Bay (*gray*). As was shown by the ecologists who collected the data, the commensal house mice breed all year but the seasons affect the breeding of the marsh mice.

As long as our young females are able to get their food without foraging for long periods of time, living at 10 degrees C. does not alter the rate at which they deposit fat, nor does it permanently impair them reproductively. If they are made to forage intensively, however, without enough opportunity to retreat periodically to their temperature-buffered burrows, the result is a total imbalance of the bioenergetic equation toward thermoregulation and away from body growth, fat deposition and reproductive development. Indeed, when our young mice are subjected to these conditions, they absolutely refuse to stay out of their burrows long enough to obtain the extra food required by the cooler temperatures, even though they are giving up any vestige of growth and of reproductive development.

This experiment was particularly rewarding in that our mice told us precisely what was and was not important for their reproductive development. One of the surprises to emerge from the experiment was the relative unimportance of the locomotor costs associated with an increased demand for foraging.

Even mice given an unlimited supply of food ran from 20,000 to 30,000 revolutions per night at 23 degrees C. This effort translates as from six to nine miles of spontaneous activity, an immense distance for a mammal less than four inches long (not counting its tail). A 150-pound man would have to run from 20,000 to 30,000 miles a day to equal the relative distance a 20-gram mouse normally moves its mass in the same period of time. Clearly the house mouse is well adapted for almost unlimited locomotor costs, and greatly increasing its activity, as would be necessary when food is scarce, costs the animal nothing in terms of growth and reproductive development. The temperature at which the increased foraging must be done, however, is crucial.

This experiment and others like it have strengthened our conclusion that a critical environmental feature limiting breeding in this species, both seasonally and geographically, is the amount of time the mouse must spend foraging at low temperatures. The size of the animal and its inability to store much energy in the form of fat leave it always in a precarious position energetically, even at temperatures that are mild by human standards.

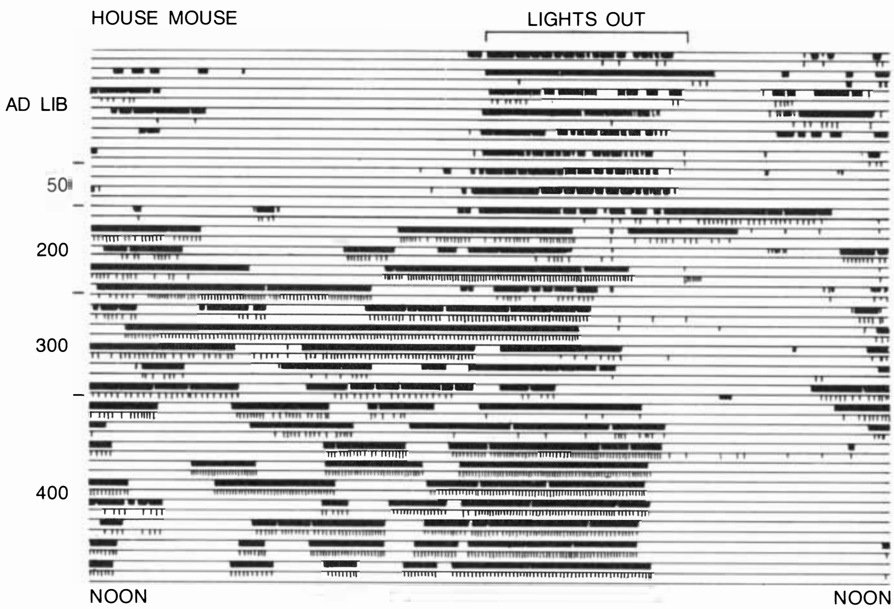
If the core environmental regulator of reproduction in the house mouse is the relation involving food availability, foraging time and ambient temperature, what adaptations have enabled the animal to achieve such widespread success as a colonizer? Two weapons in the house mouse's continuing battle with the vagaries of food supply and ambient temperature have been discovered so far. One weapon is that the house mouse, unlike most small mammals, is not strictly nocturnal. The animal can and often does forage during the day, when the temperature is higher than it is at night. The other is that once full body size and puberty have been attained the house mouse is actually quite resistant to low temperatures, provided ample food is near at hand. We find that if adult mice are slowly adapted to low temperatures and then housed in dry bedding with excess cotton available for nesting, and if excess food is placed only inches away, most of them can reproduce well at temperatures as low as −6 degrees C. Most of our mice are also able to breed at 34 degrees, only three degrees below body temperature. This is a truly amazing range of temperatures over which a small mammal can breed efficiently.

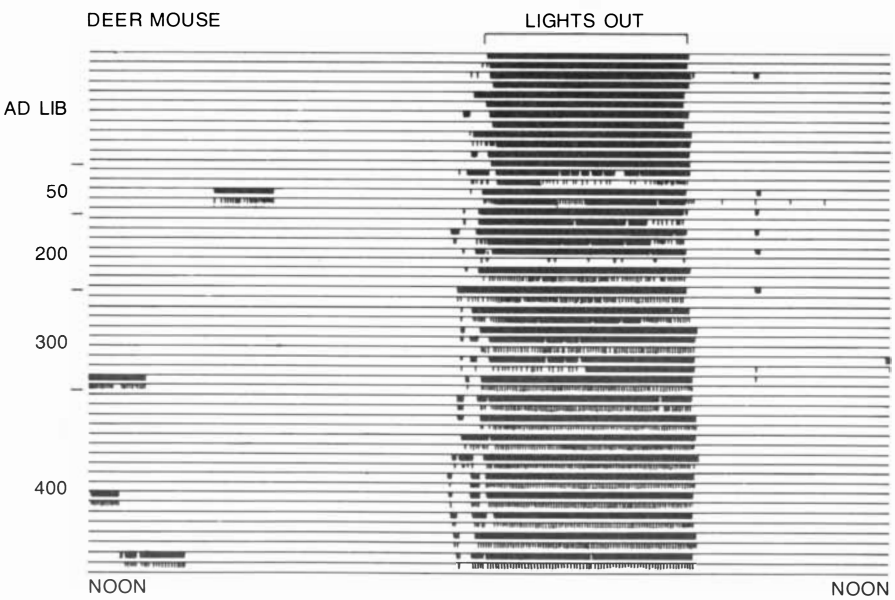
Another clue to the successful colonization efforts of the house mouse becomes evident when one examines the animal's social environment. The importance of social cues emerged entirely from laboratory studies; it had not

© 1984 SCIENTIFIC AMERICAN, INC

been suspected by field biologists. What emerges is that the house mouse has evolved a highly sophisticated system for timing its ovulation. One result is that pregnancy can be delayed until major migratory movements have ended, rather than beginning before or during such movements.

Like many other mammals, the house mouse depends heavily on olfactory cues for communication. One source of such cues is urine. As has been shown by my colleague Claude Desjardins, wherever the house mouse (the male in particular) explores or forages it deposits urine marks on the ground. The marks contain a variety of substances that can be detected by other mice. In this way a mouse can determine the species, the sex and the state of sexual readiness of another mouse even long after the other mouse has left.





DAILY ACTIVITY of a house mouse and a deer mouse was recorded for 27 days while the mice were (separately) in the test apparatus. In the apparatus every turn of the running wheel and every discharge from the pellet dispenser actuated a pen on the constantly moving chart recorder. Each 24-hour record was put one below another to show what the mouse, which was a female, did over the entire 27-day period. The top line of each record reveals the mouse's running activity during the first day; the second line is her pellet-discharge record for the same day, and then the record for each succeeding day is presented similarly. For the first few days the mice were given excess food (*ad lib*); then they were placed on a steadily increasing foraging requirement that ended with a requirement of 400 revolutions for a pellet of food. The mouse would face a similar situation as the seasons changed from summer to winter. The deer mouse, a more typical small mammal, is strictly nocturnal whereas the house mouse is not.

Some of the urinary substances can act as priming pheromones, that is, they can hormonally prime another animal for reproductive activity (or conversely they can block such activity). As was first shown in the laboratory of John G. Vandenbergh of North Carolina State University, the time when the pubertal ovulation occurs in mice is profoundly flexible. The presence of an adult male accelerates a young female's ovulation, whereas the presence of other females decelerates it. Early in the female's life the accelerating effect of a male is completely blocked if other females are around; later the male's accelerating action easily overrides the presence of other females. All such effects can be duplicated in the laboratory by periodic exposure of the animals to a droplet of urine from the appropriate sex, even though these pheromonal cues normally act in conjunction with tactile stimuli.

The adaptive significance of the mouse's system for regulating the pubertal ovulation socially seems to reside in its relation to dispersal. After they have been weaned young house mice often disperse from the area where they were born, sometimes traveling long distances to find a new home. It is an arduous journey; indeed, the probability of survival is low. The insemination of the young female by her father is prevented before dispersal by the presence of her sisters and mother. In this way the distance the young female can disperse is greatly increased.

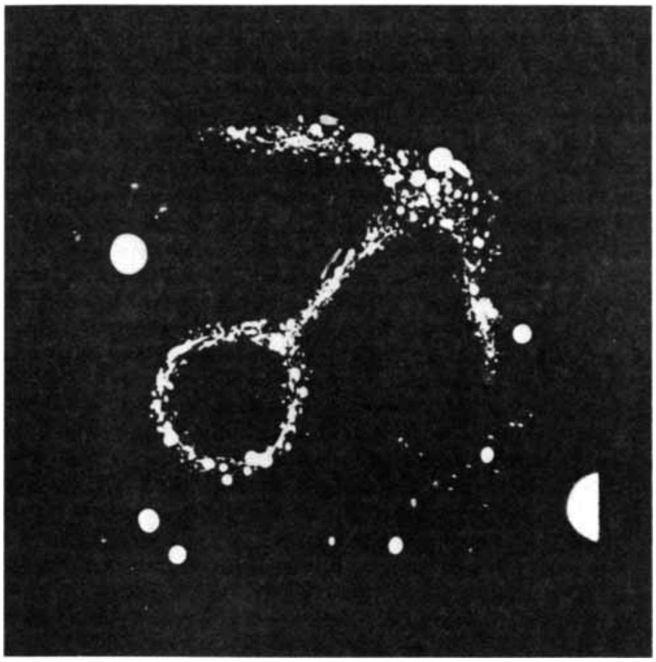
Dispersing females surely must often encounter the urine markings of a male as they travel. Such chance encounters probably exert only slight effects on a young female's sexual development, however, because of the absence of the male's synergistically acting tactile cues. It is only when the female finds a new home, cohabits with a male and is stimulated by him both pheromonally and tactilely that she attains her rapidest rate of sexual development.

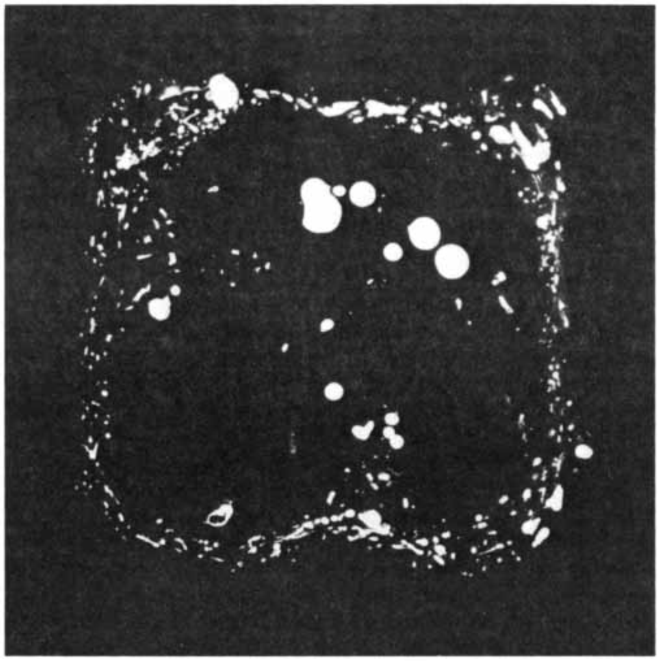
The reproductive flexibility that arises from the house mouse's particular reactions to dietary, physical and social controls is probably unparalleled among mammals. When this flexibility is coupled with some of the animal's other characteristics, particularly its capability of living commensally with man, the result is a blueprint for an opportunistic way of life and, on a grander scale, for global colonization.

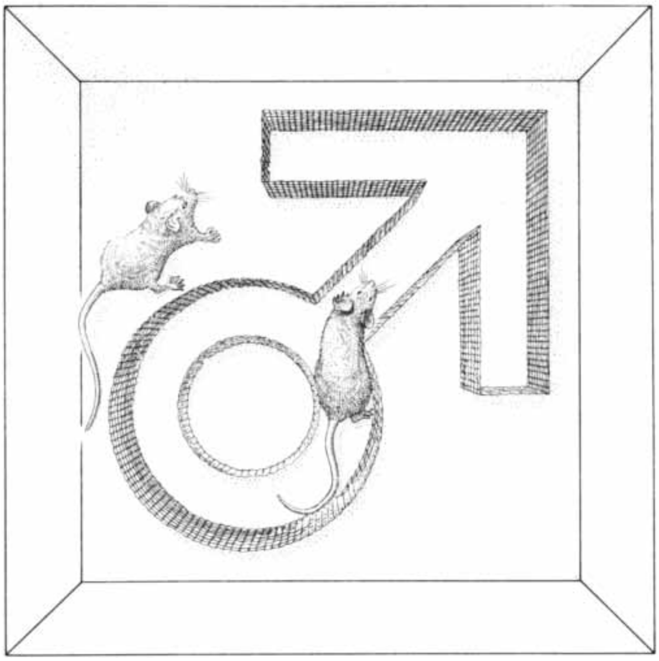
One can visualize the temporal dimension of this blueprint by starting with a commensal population. Given the buffered dietary and climatic conditions of most manmade structures, house mice will reproduce rapidly throughout the year until they reach the structure's carrying capacity. Thereafter the young animals must disperse.

Dispersing mice are sometimes car-

© 1984 SCIENTIFIC AMERICAN, INC







SOCIAL CUES affecting the reproductive development of the female house mouse are present in the urine of both sexes. The practice of marking the ground with urine was examined in the apparatus portrayed at the left. It is a cage with filter paper as the floor covering and an inner fenced-in area (shaped like the biological symbol for a male). Above that illustration is a photograph of a piece of the filter paper. The photograph was made in ultraviolet radiation, which causes the urine deposited by a mouse to fluoresce. The pattern of fluorescence in the photograph represents the urine deposited by a male house mouse that was inside the inner enclosure for 30 minutes while a female was outside. The other photograph shows the marks of a female that was inside the enclosure when a male was outside.

ried long distances by man. A key adaptation here is seen in the water metabolism of the house mouse, which is like that of a desert rodent. Because of this trait the mouse can exist in one of the driest environments on the earth: the homes, granaries and vehicles of man.

A second important adaptation is the species' wide nutritional limits. House mice can survive and breed on a vast array of foodstuffs. Wheat seeds are a particularly good breeding diet, even at low temperatures. Thus house mice have been carried around the world, often following the spread of wheat but always taking advantage of man's food and shelter to reproduce and start new colonies wherever they find themselves.

When the house mouse reinvades a natural habitat, the character of its environmental challenges changes drastically. Seasonal reproduction usually is dictated there because of the energetic imbalances afflicting the female. The house mouse becomes quite nomadic under these conditions, much more so than most other small mammals. In many natural locales it shifts about more or less constantly in its attempt to find a reproductively acceptable relation between temperature and the availability of food. The pheromonal cuing system is probably of great importance in enabling these nomadic adults to time their reproduction with respect to such movements.

Underlying all the animal's success as a colonizer is its naturally high fecundity. Given the right combination of environmental conditions, populations of house mice can reproduce rapidly to plague proportions. (More than 30 million mice were trapped at one location in Australia over a four-month period in 1917.) All the species' offspring are potential dispersers, and it is through this agency that the species continuously "tries out" new habitats. Most such attempts end in failure, but because of the species' unique array of reproductive advantages, a few animals succeed in starting new colonies. Thus has arisen the worldwide distribution of the species and the remarkable diversity of habitats in which it can be found.

Good biological stories never quite end. They often seem to raise new ques-

© 1984 SCIENTIFIC AMERICAN, INC

A logical extension of the Buick philosophy.

Introducing the Buick Century Estate Wagon.

What makes it exceptional is that, in addition to the spaciousness and convenience of a traditional station wagon, it also offers you all the advantages of a very advanced Buick.

It has the tight, firm control of front-wheel drive. The confident, dependable performance of a computer-controlled V-6 engine. And the

© 1984 SCIENTIFIC AMERICAN, INC

smooth, quiet ride provided by its computer-tuned suspension system.

Beyond these highly advanced characteristics, you will also appreciate the careful attention to workmanship that goes into the new Century Wagon–and into all new Buicks.

Of course, you will also discover the plush, comfortable, convenient interior that discriminating car owners have come to expect in Buicks.

If you're looking for a wagon that offers more than just extra space, visit your Buick dealer and buckle yourself into the Century Estate Wagon.

Not only is it logical. It's also very, very comforting.

Official Car of the XXIIIrd Olympiad
Los Angeles 1984

BUICK

Wouldn't you really rather have a Buick?

© 1984 SCIENTIFIC AMERICAN, INC

- A Self Evident Truth -
Krell Offers the Best Educational Software

for Acorn, Apple, Atari, Commodore, Kaypro, IBM, Radio Shack

KRELL'S SAT* PREP SERIES

70 Point SAT* Increase Warranty

A series of 42 programs offering complete coverage of all SAT* topics including The Test of Standard Written English. All materials are presented in SAT* format and at the same level of difficulty encountered in SAT* Exams. Scoring and explanations are provided instantly. Krell's unique logical design customizes this multi-disk set for each individual user. **$299.95**

KRELL'S LOGO

THE M.I.T. AUTHORIZED VERSION. Comprehensive 4-disk set includes two copies of LOGO for Apple II™, all utility programs and Sprite drivers, all M.I.T. demonstration programs, shape editor, music editor, 21 program *Alice in Logoland* Tutorial Series, and massive documentation including full color wall chart. THIS IS THE GENUINE ARTICLE! Unlike the version marketed by Apple Corporation, Krell's LOGO offers the full package of M.I.T. features including the ability to save pictures. ***Spectacular Price $89.95***

Krell's Best Sellers include:

Linear Equations
Isaac Newton and FG. Newton
Connections
The Black Death
Adventures in Flesh
Botticelli
Pythagoras and the Dragon
Alexander the Great
Plato's Cave
Descartes' Delight
Galactic Magellan
Galileo

Call or write for our new catalog

"SAT" and "College Board" are registered trademarks and service marks of the College Entrance Examination Board. Krell Software Corp. has no affiliation with the CEEB and is solely responsible for these programs. Krell's M.I.T. LOGO © 1981, Massachusetts Institute of Technology.

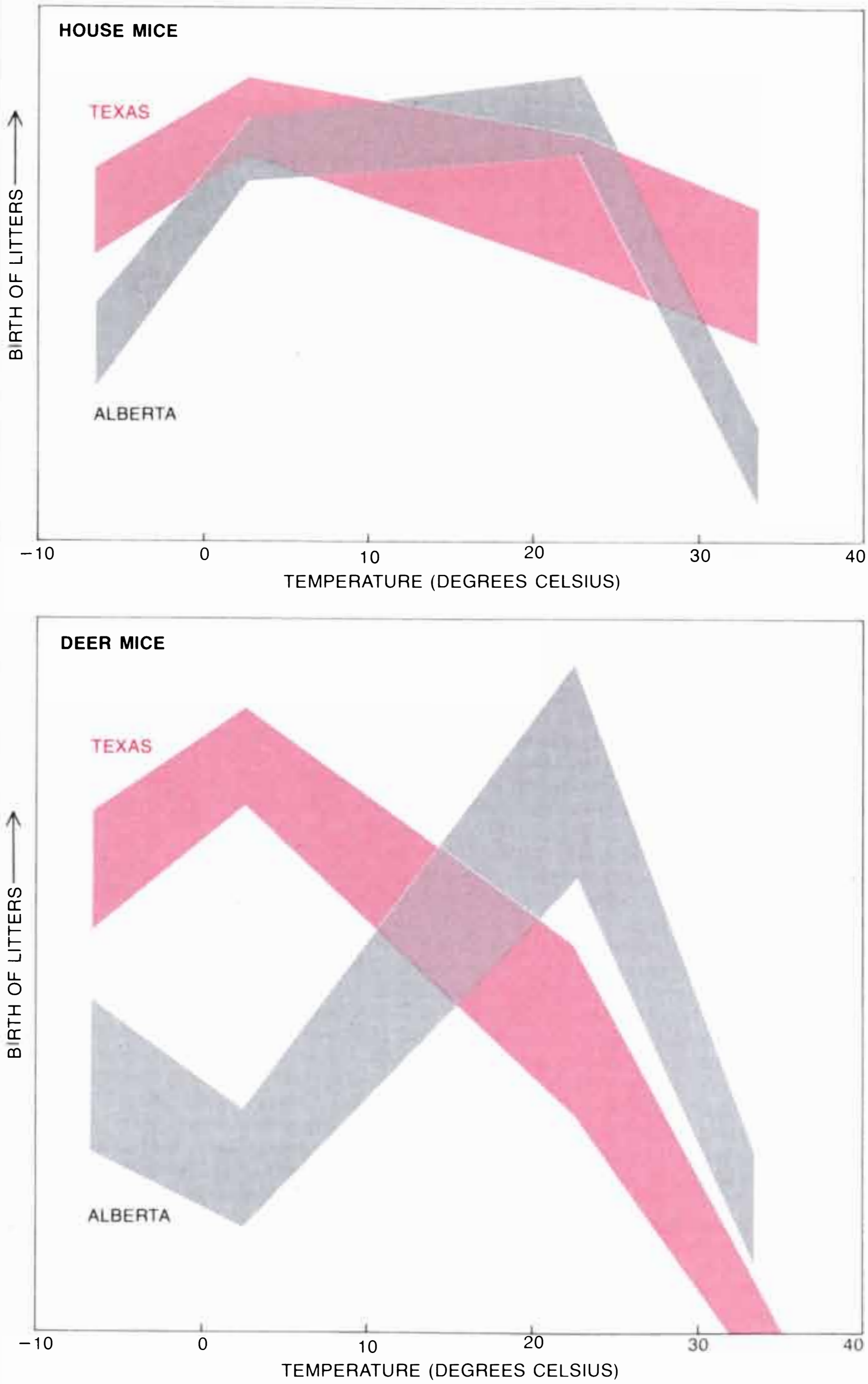
Trademarks of Apple Computer Corp., Tandy Corp., Commodore Corp., IBM, Atari Corp., Non-Linear Systems, Inc.
N.Y. residents add sales tax
Prices slightly higher outside U.S.

FOR ORDERS OUTSIDE NY
800-24KRELL
For other information call (516) 751-5139

KRELL SOFTWARE CORP.
1320 Stony Brook Road
Stony Brook, New York 11790

tions while dealing with old ones. Our research has turned up several new questions.

One question is how the house mouse retains its reproductive flexibility in the face of natural selection for efficiency under local conditions. The more typical mammal remains in one geographic region and one type of habitat, and natural selection then channels its reproductive controls to fit the unique environmental demands of those particular conditions. The typical small mammal evolves so that there is a limited temperature range within which it can breed, and its reproduction becomes dependent on a limited diet. Usually the animal becomes photoperiodic; if it is a



EFFECT OF TEMPERATURE on the breeding of house mice and deer mice obtained from different latitudes on the North American continent is charted. The mice were from Texas (*color*) and Alberta (*gray*). When the more typical deer mouse is given excess food and challenged to breed at different temperatures in the laboratory, it shows the effect of geographic selection. The winter-breeding deer mouse of Texas is insensitive to low temperatures and sensitive to high ones. As long as plenty of food is available the more adaptable house mouse breeds well at all the temperatures tested, regardless of where it lives under normal conditions.

© 1984 SCIENTIFIC AMERICAN, INC

small mammal, it also usually becomes strictly nocturnal. In contrast, success for a colonizer requires that the animal retain its reproductive flexibility in order to be able to exploit new habitats.

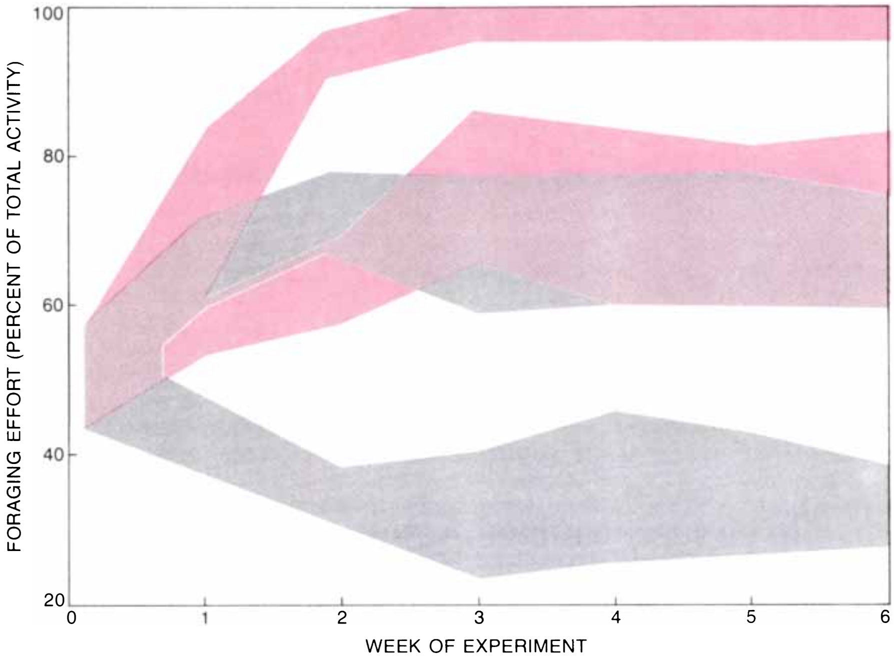
Does the house mouse in some way avoid channeling of its reproductive controls by selection? Perhaps yes, perhaps no. Certainly, as R. J. Berry of University College London has shown, many of this species' nonreproductive traits are susceptible to geographic selection. With regard to its reproductive controls, however, maybe the successful colonizers always originate from environmentally buffered buildings, never from natural populations, and therefore the world's commensal populations act as a continuing source of genetically "average" animals. Perhaps so, but in most cases commensal and noncommensal mice are genetically similar; indeed, they are often the same nomadic individuals. Hence there may actually be a genetic basis for maintaining reproductive flexibility in the face of local selection.

Another question raised by this research has to do with the uncoupling of reproductive capacity from dietary controls in the male house mouse. Does anything other than the male's social interactions with other males control its reproductive success? There are no "free lunches" in evolution, and so a related question is: What has the male given up in promoting his fitness that the more resource-oriented female has not?

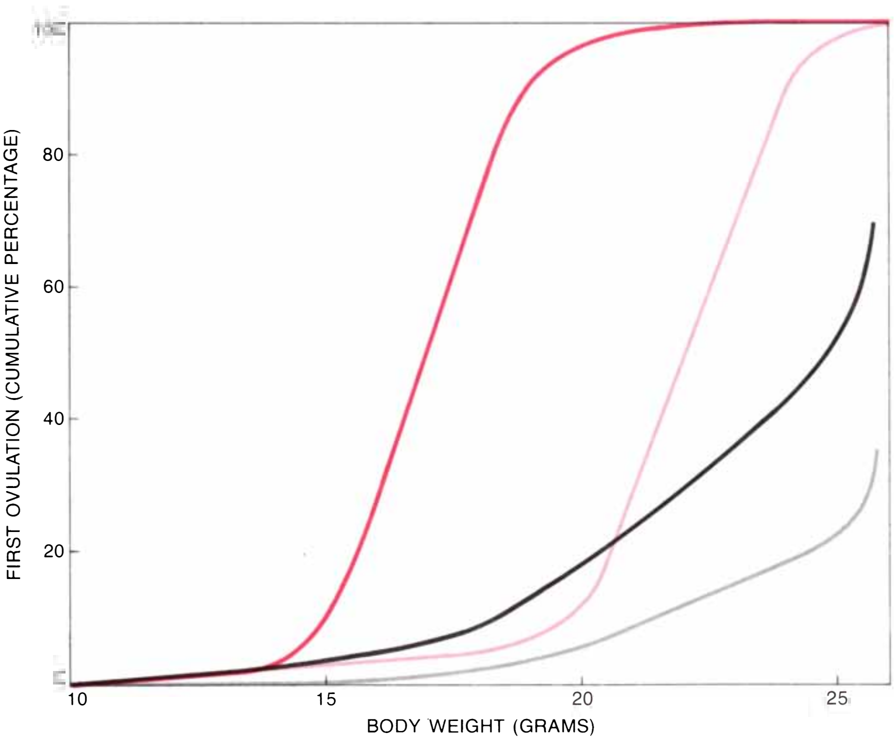
On a larger scale the mouse's amazing reproductive flexibility likewise must be paid for by giving up something else. Put another way, if the house mouse's opportunistic way of life is so successful, why have other species not adopted it? What the house mouse has given up in order to retain its reproductive flexibility remains unknown.

One final result of our work merits comment. In the course of this effort we have become increasingly aware that relatively little is known generally about the environmental control of reproduction in mammals. Indeed, one of the major problems in our work has been to find a comparable blend of ecological and physiological information about another mammal.

All mammals are subject to environmental regulation; even human beings confront dietary and emotional controls. Yet modern reproductive biology has tended to direct its attention almost exclusively to the molecular and tissue-level aspects of reproduction, relying on only a few domestic mammals as model systems. Perhaps research on the wild house mouse can lead the way to a renewed concern with the role of environmental regulation and, ultimately, to a broader view of mammalian reproduction generally.



TEMPERATURE AND FORAGING are closely related. When a house mouse is required to run 100 revolutions at a temperature of 23 degrees Celsius for a pellet of food, it devotes only a small fraction of its daily activity to foraging (*lower gray area*). When the mouse runs 200 revolutions per pellet at the same temperature (*upper gray area*) or 100 revolutions at nine degrees C. (*lower colored area*), it has less time for nonforaging activity. When it is required to run 200 revolutions at nine degrees (*upper colored area*), it reduces its total activity (not shown) and devotes all its activity to foraging. Over the first two weeks of the experiment reflected here the animals were being slowly adapted to temperature; most were developing rapidly.



VARIATION IN ONSET OF PUBERTY among female house mice is seen in certain conditions of housing. The conditions were a single female paired with a male (*dark color*), females living in a group with a single male (*light color*), isolated females (*black*) and females in a group without a male (*gray*). Young females cohabiting at weaning with a male mature rapidly and achieve their first ovulation early, whereas in the presence of other females the first ovulation is delayed. The phenomenon represents one of the components of the house mouse's reproductive adaptability. Young mice usually disperse, and the delay in first ovulation caused by the presence of other females prevents pregnancy before or during dispersal. Only when the female has found a new home and cohabited with a male does she attain sexual maturity rapidly.

© 1984 SCIENTIFIC AMERICAN, INC

Where the

To be successful today, you not only need to know what you're doing. You also need to know what the people around you are doing.

Which isn't always easy. Unless you have The Smart Desk from IBM.

Because The Smart Desk is any desk with an IBM computer or terminal on it. Our 3290 display station, for example, can give you plenty of powerful connections.

It can connect you to customer files, inventory levels, production status and the text of a letter confirming an order—all at the same time, all on the same screen.

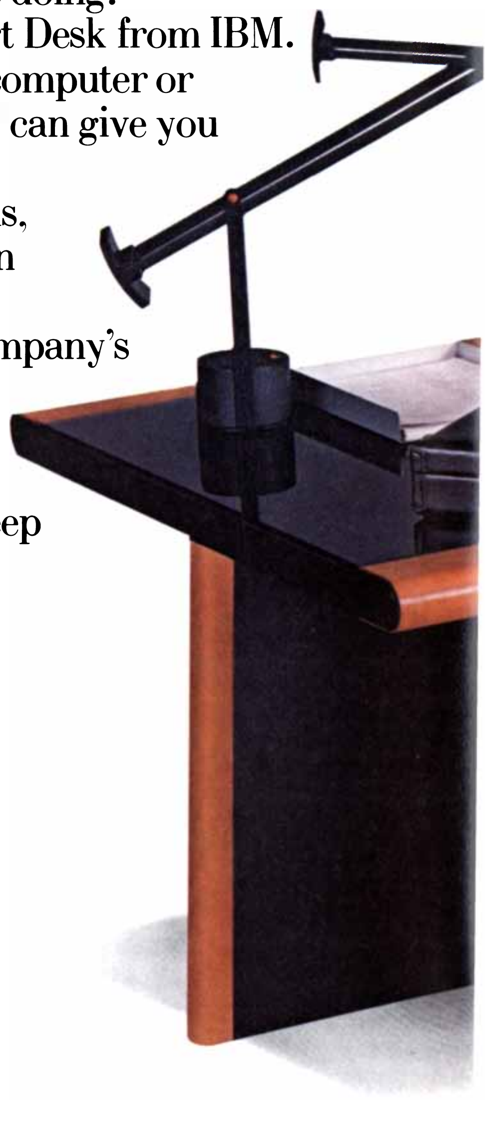
The Smart Desk can put you in touch with your company's central data files and office systems. Or with The Smart Desks of managers, secretaries and other professionals in your office.

It can help produce graphs, compile reports and keep your calendar. Even send and receive electronic mail.

IBM Smart Desk products are easy to use, so they won't try even the busiest person's patience. And if a question arises, relax. Our famous service and support network is behind you.

To make your desk The Smart Desk, talk to your IBM representative. And get all the right connections.

IBM®



The Smart Desk from IBM

© 1984 SCIENTIFIC AMERICAN, INC

successful stay calm, cool and connected.

© IBM CORPORATION

© 1984 SCIENTIFIC AMERICAN, INC

Florence Nightingale

She saved the lives of thousands of soldiers in the Crimea and was one of the founders of modern medical care. She was also a pioneer in the uses of social statistics and in their graphical representation

by I. Bernard Cohen

Florence Nightingale is remembered as a pioneer of nursing and a reformer of hospitals. She herself saw her mission in larger terms: to serve humanity through the prevention of needless illness and death. For much of her long life (1820–1910) she pursued this mission with a fierce determination that gave everything she did a singular coherence. Her greatest contributions were undoubtedly her efforts to reform the British military health-care system and her establishment, through the founding of training programs and the definition of sound professional standards, of nursing as a respected profession. Much of what now seems basic in modern health care can be traced to pitched battles fought by Nightingale in the 19th century. Less well known, because it has been neglected by her biographers, is her equally pioneering use of the new advanced techniques of statistical analysis in those battles.

Nightingale learned at first hand as chief nurse during the Crimean War (1854–56) that improved sanitary conditions in military hospitals and barracks could sharply cut the death rate and save thousands of lives. Her battle was to convince skeptical men in power. At a time when the collection and analysis of social statistics was still uncommon Nightingale recognized that reliable data on the incidence of preventable deaths in the military made compelling arguments for reform. Thus in addition to advancing the cause of medical reform itself she helped to pioneer the revolutionary notion that social phenomena could be objectively measured and subjected to mathematical analysis.

Nightingale's achievements are all the more impressive when they are gauged against the background of social restraints on women in Victorian England. Her father, William Edward Nightingale, was an extremely wealthy landowner, and the family moved in the highest circles of English society. In those days women of Nightingale's class did not attend universities and did not pursue professional careers; their purpose in life was to marry and bear children. Nightingale was fortunate: her father believed women should be educated, and he personally taught her Italian, Latin, Greek, philosophy, history and—most unusual of all for women of the time—writing and mathematics. When in her early twenties Nightingale expressed an interest in nursing, her father took that interest seriously enough to consult with physicians about the suitability of such a career.

If pursuing any career was a radical step for a woman of Nightingale's social class, however, taking up nursing seemed out of the question even in her enlightened family. It was not "the physically revolting part of a hospital" that offended William Nightingale so much as what seemed to be overwhelming evidence of the dissolute habits of nurses. Nurses in those days lacked training; they were almost always coarse and ignorant women, given to promiscuity and drunkenness. Nightingale herself later told her father she had been informed by the head nurse in a London hospital that she "had never known a nurse who was not drunken" and that most of the nurses engaged in "immoral conduct" with the patients in the wards. Not surprisingly, her parents hoped their daughter would give up her unusual ambition, marry and settle down.

By all accounts Florence Nightingale was an attractive young woman, and it was not for lack of opportunity that she rejected marriage. Indeed, she once was tempted to accept a suitor, but after a long courtship she reluctantly concluded that she could never satisfy her "moral" and "active" nature "by spending a life with him in making society and arranging domestic things." Conventional marriage, she wrote in her diary, meant "to be nailed to a continuation and exaggeration" of her "present life," a prospect that seemed to her "like suicide." God, she decided, had envisioned for her a different fate. She was one of those whom he "had clearly marked out... to be single women."

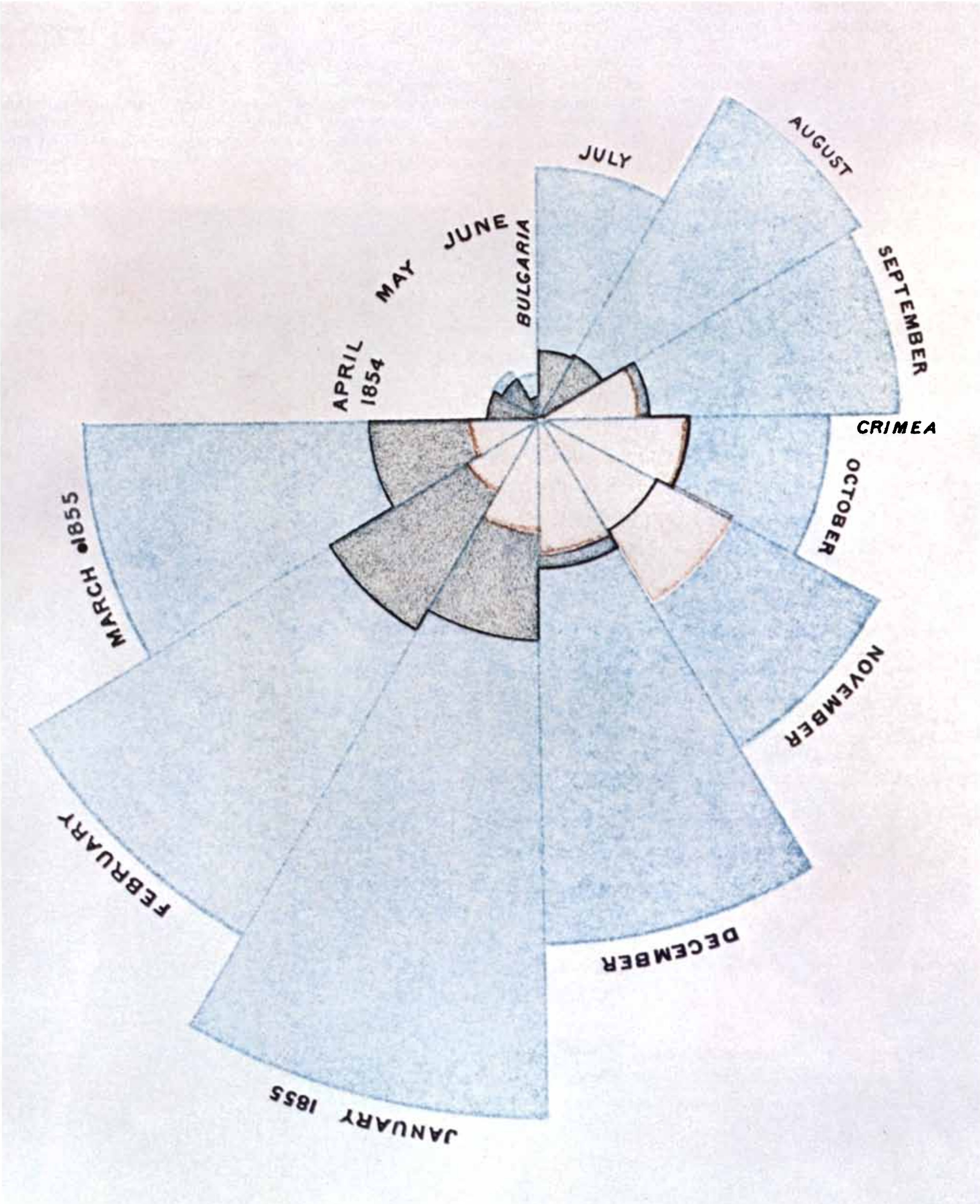
When her parents forbade her to take up nursing, Nightingale turned for comfort to religion. It was to remain a driving force in her life. Her religious feelings, however, centered on the conviction that the best way to serve God was through service to mankind. Thus in the difficult years of her twenties she did not give up her ambition to pursue a career; she read voraciously on medicine and health care, spent some time inspecting hospitals in London and worked privately with children of the slums, whom she called her "little thieves at Westminster." Still, she was frustrated.

Finally in 1851 Nightingale was able to break away from home, spending three months near Düsseldorf in Germany at a hospital and orphanage run by a Protestant order of "deaconesses." Later, in spite of the protests of her family, she served an apprenticeship at another hospital, this one operated by the Sisters of Mercy in St. Germain, near Paris. At the age of 33 she was at last starting out in her chosen profession.

Returning to London in 1853, Nightingale soon got her first "situation" (an unpaid one) as superintendent of a London "establishment for gentlewomen during illness." Her job was to supervise the nurses and the functioning of the physical plant and to guarantee the purity of the medicines. Although she succeeded in creating a model institution by the standards of the day, one that was open to patients of all classes and religions, she was disappointed that she could not accomplish what even then she had come to consider her primary aim: the establishment of a formal training school for nurses.

Nightingale stayed only a year at her first job, because greater opportunities awaited her. In September, 1854, British and French troops invaded the Crimea, on the north coast of the Black Sea, in support of Turkey in its dispute with Russia. (Russia had long had territorial ambitions in Turkey, particularly with

© 1984 SCIENTIFIC AMERICAN, INC



POLAR-AREA DIAGRAM was invented by Florence Nightingale to dramatize the extent of needless deaths in British military hospitals during the Crimean War (1854–56). She called such diagrams "coxcombs." The area of each colored wedge, measured from the center, is proportional to the statistic being represented. Blue wedges represent deaths from "preventable or mitigable zymotic" diseases (contagious diseases such as cholera and typhus), pink wedges deaths from wounds and gray wedges deaths from all other causes. Mortality in the British hospitals peaked in January, 1855, when 2,761 soldiers died of contagious diseases, 83 of wounds and 324 of other causes, for a total of 3,168. Based on the army's average strength of 32,393 in the Crimea that month, Nightingale computed an annual mortality rate of 1,174 per 1,000. The diagram is taken from Nightingale's book *Notes on Matters Affecting the Health, Efficiency and Hospital Administration of the British Army* (1858); half of the diagram, representing the period from April, 1855, to March, 1856, does not appear.

© 1984 SCIENTIFIC AMERICAN, INC

regard to Constantinople, the Orthodox holy city; one of the proximate causes of the Crimean War was the Russian demand that it be given a protectorate over the Orthodox subjects of the Turkish sultan.) The allied forces scored a quick victory at the Battle of the Alma River on September 20, and then began a siege of the Russian naval base at Sevastopol. Public jubilation in Britain soon turned to dismay when the Crimean correspondent of *The Times,* William Howard Russell, reported that sick and wounded British soldiers were being left to die without medical attention. Not only were there too few surgeons and "not even linen to make bandages" but also there was not a single qualified nurse in the British military hospital at Scutari (near Constantinople). The French, on the other hand, had sent 50 Sisters of Mercy to the Crimea.

It was a golden opportunity for the ambitious Nightingale. She immediately wrote to a longtime friend, Sidney Herbert, the "Secretary at War," to volun-

FLORENCE NIGHTINGALE is pictured in a photograph taken in later life. Beginning soon after her return from the Crimea in 1856 until her death in 1910 at the age of 90, she lived as an invalid, largely confined to her bedroom. Her illness may not have been organic, but it did not prevent her from exercising her influence by receiving frequent visitors and by maintaining an extensive correspondence.

© 1984 SCIENTIFIC AMERICAN, INC

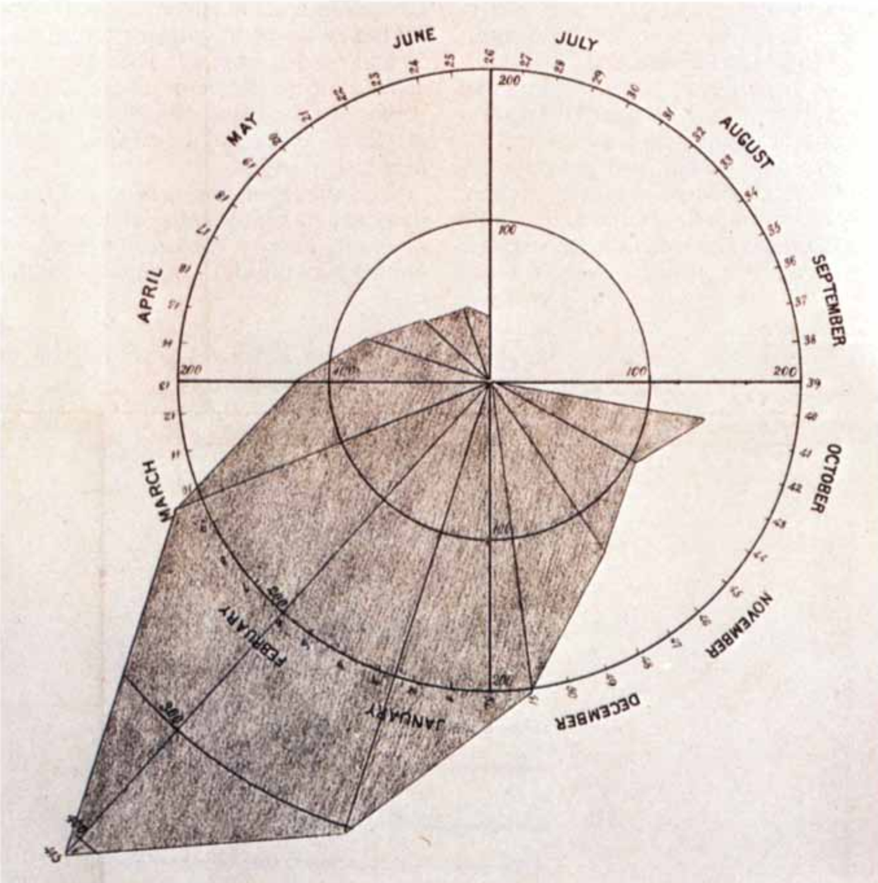
teer her services. As it happened, a letter from Herbert was already on its way to her, asking her to recruit a corps of trained nurses and lead them to Scutari. When Nightingale left for Turkey on October 21, 1854, accompanied by 38 nurses, she had the official backing of the government (although not of the army) and, perhaps more important, the private financial support of a special fund raised by *The Times*. Besides making her an international heroine, her work in the Crimea and the conditions she saw there were to determine her mission for the rest of her life.

The conditions Nightingale and her party found when they arrived at Scutari on November 5, the day of the major Battle of Inkerman, were appalling. The hospital barracks was infested with fleas and rats. Under the buildings, as a commission of inquiry later reported, "were sewers... loaded with filth... through which the wind blew sewer air up the pipes of numerous open privies into the corridors and wards where the sick were lying" on straw mats, in a state of overcrowding that got even worse after Inkerman. The canvas sheets, according to Nightingale, were "so coarse that the wounded men begged to be left in their blankets"; moreover, the laundry was done in cold water, with the result that many linens returned as clean were so "verminous" that they had to be destroyed. Essential surgical and medical supplies were lacking, or their distribution was blocked by military red tape.

These were the conditions that awaited patients arriving at Scutari after a slow sea voyage across the Black Sea and through the Bosporus, weak and emaciated, suffering from frostbite and dysentery as well as from their wounds. In fact, the resulting epidemics of cholera and typhus, and not the injuries themselves, caused the greatest loss of life at Scutari. In February, 1855, the mortality rate at the hospital was 42.7 percent of the cases treated.

In her efforts to establish an effective hospital in Turkey, Nightingale showed real skill as an administrator. At every step, however, she was hampered by the military authorities, who resisted any change that might seem to be a concession of their own errors or incompetence. The military men resented the fact that Nightingale's authority was independent of the armed services, that she was a civilian and—far worse—that she was a woman. Hostility to her mission ran so high that at first her nurses were not allowed on the wards. Even after she had achieved greater acceptance she had to struggle against petty officials, such as a supply officer who refused to distribute badly needed shirts from his store until the entire shipment of 27,000 could be inspected by an official of the Board of Survey.

In the face of such impediments it was Nightingale's independence from the military and, above all, her private source of funds that enabled her to accomplish so much at Scutari. She established her own laundry, including boilers to heat the water; she installed extra kitchens in the hospital; she became, finally, the supplier of the entire hospital, "a kind of General Dealer in socks, shirts, knives and forks, wooden spoons, tin baths, tables and forms, cabbage and carrots, operating tables, towels and soap, small tooth combs, precipitate for destroying lice, scissors, bed pans and stump pillows." The money for these supplies and for the staff she recruited came not only from *The Times* fund but also from other philanthropists and from her own private funds.



MORTALITY RATE at Scutari, the main British hospital in the Crimean War, declined sharply after sanitary improvements were made under Nightingale's influence. In the winter of 1854–55 the British army besieging the Russian fortress at Sevastopol was ravaged by malnutrition, exposure and disease: dysentery, cholera, typhus and scurvy. The death rate at Scutari, calculated here by Nightingale on an annual basis as a fraction of the patient population, reached 415 percent in February. Sanitary reforms began in March. This diagram is taken from the report of a Royal Commission set up after the war to investigate sanitary conditions in the army.

While Nightingale was carrying out her administrative duties she still found time to attend to the sick herself, late at night, on endless rounds that gave rise to the legend of the "ministering angel" of the Crimea. At night she banned all other women from the wards (she had been obliged to send some of her nurses home for delinquent behavior) and made her way, according to the commissioner of *The Times* fund, "alone, with a little lamp in her hand," through "those miles of prostrate sick." Longfellow immortalized this "lady with a lamp" image in his poem of 1857 ("Lo! in that house of misery / A lady with a lamp I see"). There is, however, a more significant measure of Nightingale's accomplishment, one that she herself stressed: by the spring of 1855, half a year after she arrived at Scutari, mortality in the hospital had dropped from 42.7 percent to 2.2 percent.

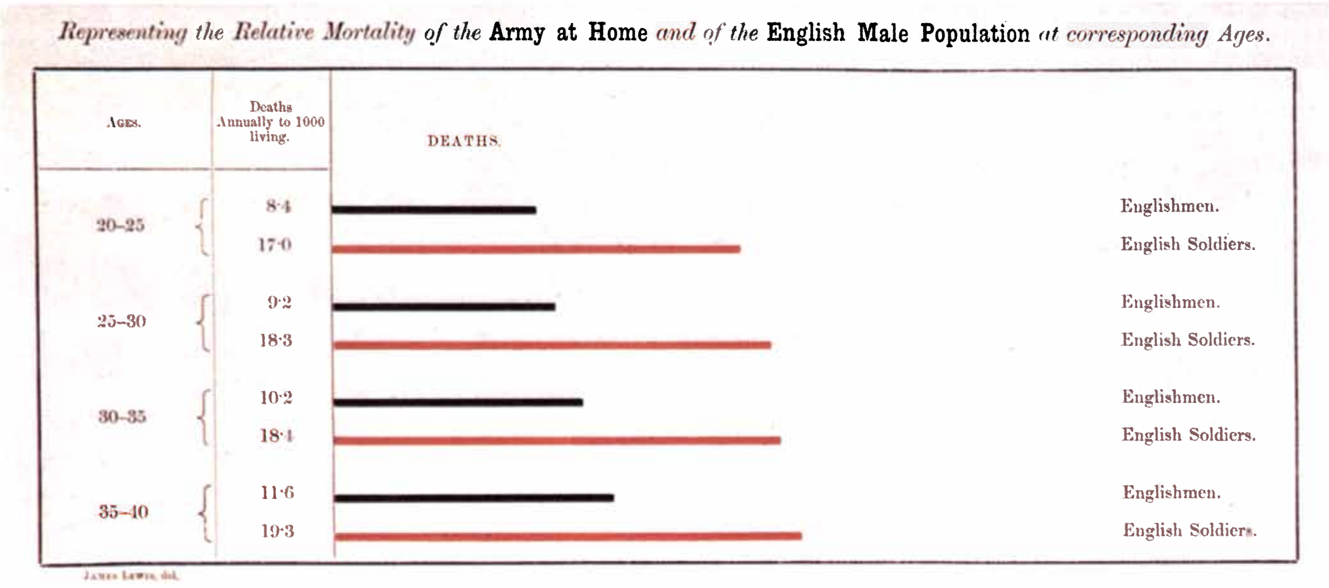
Nightingale returned to England in July, 1856, four months after the end of the war. By that time, at the age of 36, she was a world-famous and revered figure. She nonetheless shunned all attempts to honor her publicly, deciding instead that the most appropriate recognition for her services would be the establishment of a commission to investigate military medical care. In the Crimea, she wrote, some 9,000 soldiers were lying "in their forgotten graves," dead "from causes which might have been prevented." The tragedy of needless death was continuing in every army

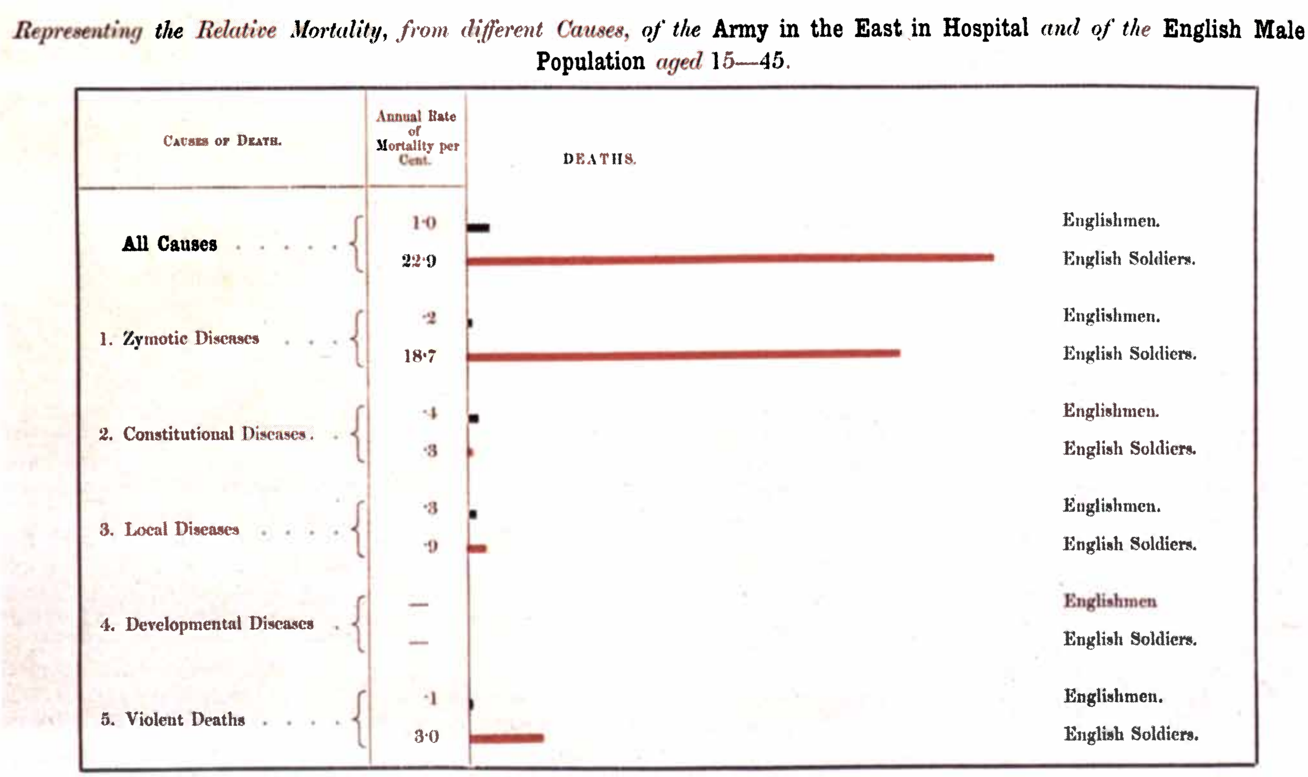
© 1984 SCIENTIFIC AMERICAN, INC

barracks and hospital, even in peacetime. It could be ended only by instituting throughout the Army Medical Service the same sanitary reforms that had saved so many lives at Scutari. This was the task Nightingale set herself.

How could she convince people of the need for reform? Nightingale saw that the most compelling argument would be statistical. The idea of using statistics for such a purpose—to analyze social conditions and the effectiveness of public policy—is commonplace today, but at that time it was not. The science of social statistics was in its infancy, and in promoting the cause of medical reform Nightingale became a promoter of the new tool as well.

Seen simply as the collection of numerical data, statistics have a long history (going back at least to the Book of Numbers of the Old Testament), but the analysis of such data is only as old as the scientific revolution of the 17th century. Early attempts to analyze data on social phenomena were hampered by inadequacies both in the data themselves and in the mathematical tools of analysis. According to the historian of statistics Helen M. Walker, the rise of modern statistics in the 19th century had three roots: the development of the mathematical theory of probability, the emer-



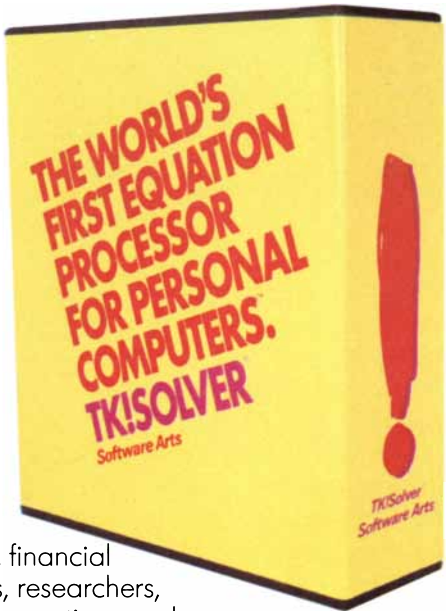


LINE DIAGRAMS from the Royal Commission's report compare conditions in the army to those in civilian life. Mortality in the peacetime army in Britain was nearly twice as high as it was among civilians (*top*). In the Crimean War "zymotic" diseases were the main causes of death and were far more prevalent than they were in England (*bottom*). Figures in the top diagram are percentages; those in the bottom one are per 1,000. The report led to the adoption of a sanitary code for the army and to a series of physical improvements in military buildings. Like other diagrams in report, these are examples of Nightingale's innovative approach to representation of statistics.

© 1984 SCIENTIFIC AMERICAN, INC

EQUATIONS PROCESSED

NO PENCIL. NO PAPER. NO MANUAL LABOR.



The TK!Solver® program will take on your toughest problems—linear, quadratic, simultaneous equations, whatever. Then stand back. Because TK!Solver turns your personal computer into a simple, yet powerful, desktop equation processor.

Whether your problem is a simple formula or a model consisting of many equations, TK!Solver can help improve your productivity. Once the equations are written, enter the known values, press the ! key, and TK!Solver gives you the answer.

Engineers, scientists, architects, financial analysts and planners, educators, researchers, and other professionals who use equations and mathematical models can work more creatively with TK!Solver.

TK!SOLVER GIVES YOU: BACKSOLVING

If the programs you use now require you to rewrite the same equation to solve for different unknowns, TK!Solver can dramatically improve your productivity. Enter your problem once and then solve for the unknowns no matter where they are in your equation.

ITERATIVE SOLVING

If TK!Solver can't solve an equation directly, take an educated guess at the answer. Type the ! key and the TK!Solver program starts with your guess and performs repeated approximations to converge on the answer.

LIST SOLVING

Given a list of input values, TK!Solver automatically calculates the equation for every value in your list. For example, if you want to know how different interest rates will affect monthly loan payments, enter a list of interest rates and let TK!Solver calculate the payment amount for each value.

UNIT CONVERSIONS

Any type of unit conversion—Fahrenheit to Celsius, meters to feet, dollars to deutschemarks, newtons to dynes—can be made without altering your equations. Just define the numerical relationship between two units of measurement and the TK!Solver program automatically converts the variable value to the unit you specify.

TABLES AND PLOTS

Quickly generate tables and plots of your results on your screen or printer.

AVAILABLE NOW

You can run the TK!Solver program on the IBM® PC and XT and compatible machines, the Digital™ Professional™ 350, the Digital™ Rainbow™ 100, the Wang Professional Computer, Apple® //e, and on the following personal computers using MS™-DOS: TI Professional Computer, GRiD Compass Computer,™ Canon AS-100, Eagle® 1600, Toshiba T300, and the Zenith Z-100.™

SEE IT TODAY

There's more. Lot's more. But you'll have to see it to believe it. And that's easy. Bring your own equations into your nearest computer retailer and ask to see the TK!Solver program in action.

The world's first equation processor for personal computers.

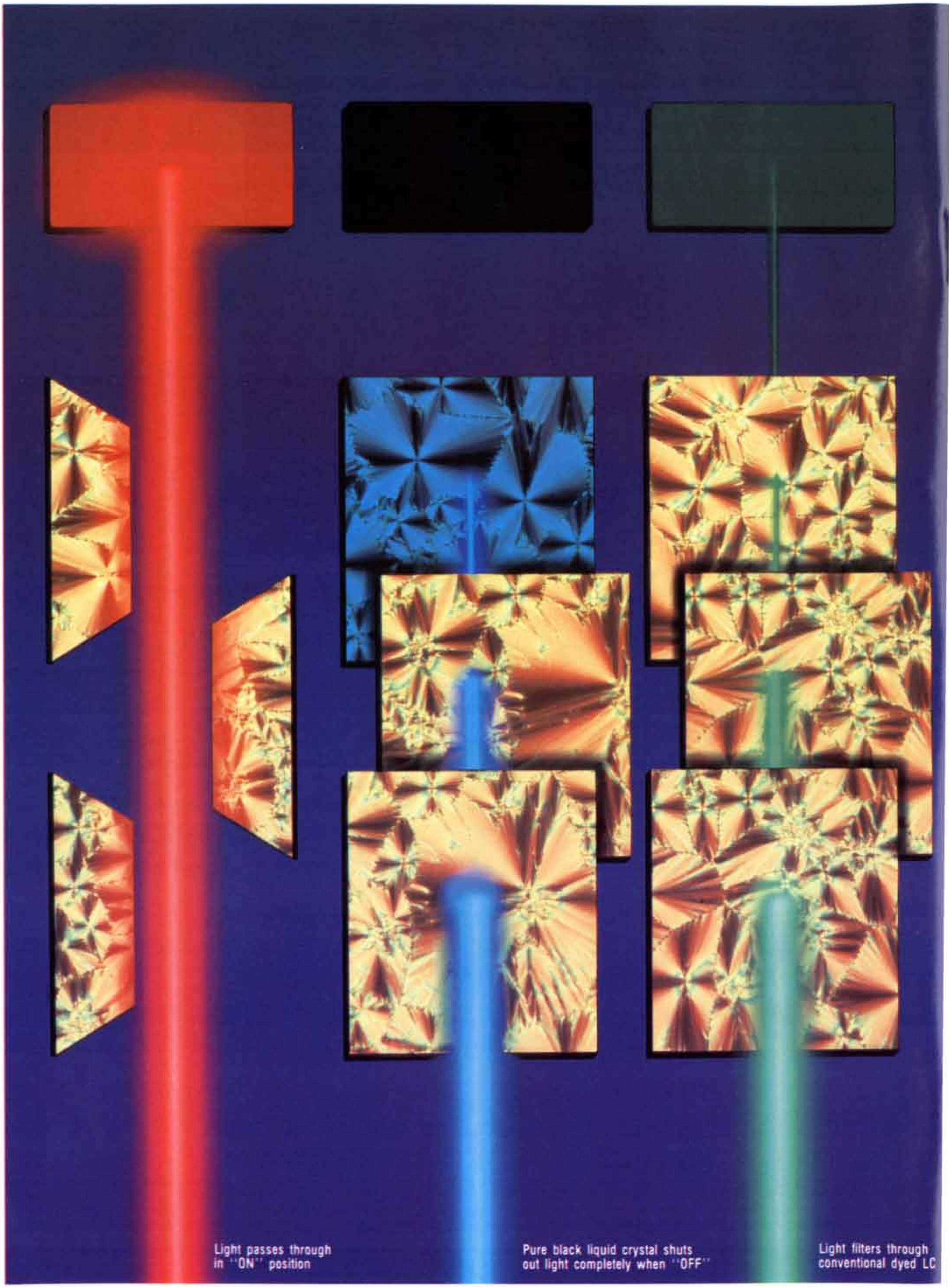
TK!Solver®

By Software Arts,™ creators of VisiCalc®

27 Mica Lane, Wellesley, Massachusetts 02181 617-237-4000

TK, TK!, TK!Solver, the stylized ! and the slogan "THE WORLD'S FIRST EQUATION PROCESSOR FOR PERSONAL COMPUTERS" are trademarks or registered trademarks of Software Arts, Inc. Software Arts is a trademark of Software Arts, Inc. and Software Arts Products Corp. The TK!Solver program is a product of Software Arts, Inc., which is solely responsible for its contents. VisiCalc is a registered trademark of VisiCorp. Apple is a registered trademark of Apple Computer Inc. IBM is a registered trademark of International Business Machines Corporation. GRiD Compass Computer is a trademark of GRiD Systems Corporation. Z-100 is a trademark of Zenith Data Systems. Eagle is a registered trademark of Eagle Computer, Inc. Digital, Professional, and Rainbow are trademarks of Digital Equipment Corporation. WANG is a registered trademark of Wang Laboratories Inc. MS is a trademark of Microsoft Corporation.
Copyright © 1984 Software Arts, Inc. All rights reserved.

© 1984 SCIENTIFIC AMERICAN, INC



© 1984 SCIENTIFIC AMERICAN, INC

Adding new color to the future of visual displays

LCDs should be colorful, not just flat

Panel-type televisions that hang on the wall, portable computers with ultra-thin color-graphic screens, high-resolution stadium scoreboards and electronic road maps—turning such dreams as these into commonplace reality requires a new approach to visual display. That's why so much attention is being focused on liquid crystal displays (LCDs), which allow low power consumption, flat panel construction, compact designs and economical production. The only major problem is that standard LCDs provide insufficient color contrast for such demanding applications.

Darkening crystals at the molecular level

Dye molecules used in LCDs function like tiny swinging doors, allowing light to pass through when turned "ON" and filtering light when turned "OFF." Thus dyes give color to the liquid crystals in the "OFF" state. But to produce high contrast, LCDs need hues as distinct as printer's ink and pure black liquid crystals that let no light escape over the entire visible spectrum.

Hitachi takes a new angle on color display

To achieve this goal, Hitachi engineers developed new types of dichroic dyes that exhibit light or dark tones depending on the angle the molecules are turned, rather like the different tones created by brushing suede. LCDs using such dyes yield 30 to 50 percent better color contrast than those using conventional types. And thanks to these new dichroic dyes—absorbing light evenly at wavelengths from 400 to almost 700 nm—red, green and blue filters can be mounted on black-crystal LCDs to produce full-color displays. Even now, Hitachi is marketing dichroic-dye LCDs in brilliant red, green, yellow and other customized colors, for use in automobile dashboards, portable measuring devices and other equipment requiring numeric displays.

Technical excellence is embodied in all Hitachi products

The development of high-contrast color LCDs is just one case demonstrating Hitachi's technological strength. You'll find other examples in all of Hitachi's products, from high-resolution CRTs and dot-matrix LCD modules to a whole host of high-quality electronic appliances. Our comprehensive technological expertise is your guarantee of convenience, ease of operation and high reliability in every product that bears the Hitachi brand.

HITACHI

A World Leader in Technology

© 1984 SCIENTIFIC AMERICAN, INC

In the celebrated world of quarter horses, the value of a stallion often runs into the millions. So it's no wonder that Bob Norris, former President of the American Quarter Horse Association, and so many other distinguished figures in equestrian circles, trust the safe transport of their valued cargos to the Jeep® Grand Wagoneer.

There's a special feeling of confidence behind the wheel of a Grand Wagoneer. There's authority in its power, security in its traction and prestige in its appearance.

With its remarkable Selec-Trac 2-wheel/4-wheel drive system, the Wagoneer artfully masters most any terrain, even under adverse conditions...from summer downpours to winter snowstorms.

Yet, for all its impressive strength and dependability, classic comforts are standard in a Grand Wagoneer. From its plush interior fabrics and leathers to its premium stereo sound system, quality is never compromised.

And because it does so *many* things without compromise, Jeep Grand Wagoneer makes perfect horse sense, no matter who your traveling companion is.

Jeep® Grand Wagoneer. The Ultimate Wagon.

FROM JEEP CORPORATION. AVAILABLE FOR PURCHASE OR LEASE AT PARTICIPATING JEEP DEALERS.

© 1984 SCIENTIFIC AMERICAN, INC

gence of the modern state with its agencies for collecting information on its citizens and their activities, and the theoretical interest of political economists in finding causes for human social behaviors. These "three movements," Walker wrote, were pulled together in the career of the mid-19th-century Belgian astronomer-statistician Lambert-Adolphe-Jacques Quételet, widely regarded as the founder of modern social statistics. In 1841 Quételet organized Belgium's central statistical bureau, which became a model for similar agencies in other countries, and his international leadership in statistical research continued until his death in 1874.

Nineteenth-century scholars trying to make a science out of the study of human behavior faced a dilemma: the model science of those days was classical physics, with its deterministic laws describing natural phenomena, but human behavior seemed individual and indeterminate. Quételet's resolution of the problem bypassed the question of the individual with the concept of an "average man." He showed that whereas there are no laws determining individual behavior, there are regularities in the attributes and behavior of groups, and that these regularities could be characterized mathematically by the laws of probability. Quételet was convinced that even mental and moral traits, if only they could be measured accurately, would also follow regular laws of statistical distribution.

Quételet's most original and most startling work was his analysis of the influence of such factors as sex, age, education, climate and season on the French crime rate (1831). The data did not allow a prediction of who would commit what crime, but according to Quételet they did display regularities that would enable a scientist to "enumerate in advance how many individuals will stain their hands in the blood of their fellows, how many will be forgers, how many will be poisoners." The discovery of these regularities led Quételet to the radical conclusion that "it is society which, in some way, prepares these crimes, and the criminal is only the instrument that executes them."

Although Quételet's work was highly regarded by many scholars, it was abhorred by others. The determinism of his "social physics" was an anathema to people committed to the prevailing doctrines of free will and individual responsibility. John Stuart Mill, for example, wrote at length against probability in general and its application to social science in particular. Another vocal opponent of the statistical view of man and society was Charles Dickens. His novel *Hard Times* was meant to satirize those people, Dickens later said, who could see nothing but "figures and averages," those "addled heads" who would use the yearly average temperature in the Crimea "as a reason for clothing a soldier in nankeens [silks] on a night when he would be frozen to death in fur." Dickens disliked the statistical view because he thought it was dehumanizing, and in *Hard Times* he portrayed the regularities found by statisticians in the rate of insanity, crime, suicide and prostitution as a "deadly statistical clock."

Nightingale, on the other hand, was an ardent admirer of Quételet's work, and she early displayed a predilection for collecting and analyzing data. At Scutari, apart from the all-important sanitary reforms she instituted, she also systematized the chaotic record-keeping practices; until then even the number of deaths was not known with accuracy. When she returned to England in 1856, she met William Farr, a physician and professional statistician. Under Farr's guidance Nightingale soon recognized the potential of the statistics she had gathered at Scutari, and of medical statistics in general, as a tool for improving medical care in military and civilian hospitals.

Throughout military history until the 20th century the main cause of death in war was disease rather than wounds sustained in battle, and the Crimean War was no exception. Nightingale's numbers still speak eloquently. During the first months of the Crimean campaign there was "a mortality among the troops at the rate of 60 percent per annum from *disease* alone," a rate exceeding that of the Great Plague of 1665 in London and higher also "than the mortality in cholera to the attacks" (that is, the mortality among those who had contracted the disease). In January, 1855, the mortality in all British hospitals in Turkey and the Crimea, measured in relation to the entire army in the Crimea but not including men killed in action, peaked at an annual rate of 1,174 per 1,000. Of this number 1,023 deaths per 1,000 were attributable to "zymotic" disease (a category introduced by Farr including epidemic, endemic and contagious disease). This means that if mortality had persisted for a full year at the rate that applied in January, and if the dead soldiers had not been replaced, disease alone would have wiped out the entire British army in the Crimea.

Nightingale's various methods of calculating mortality dramatized both the impact of disease and the effects of improved sanitary conditions. Calculated on an annual basis as a percentage of the patient population, the death rate at the Scutari hospital reached an incredible 415 percent in February, 1855. In March, however, Nightingale's sanitary reforms began to be implemented and mortality among the patients declined precipitously. By the end of the war, according to Nightingale, the death rate among sick British soldiers in Turkey was "not much more" than it was among healthy soldiers in England; even more remarkable, the mortality among all British troops in the Crimea was "two-thirds only of what it [was] among our troops at home."

The comparison suggested that the soldiers at home were living in their barracks under unhealthy conditions. After Farr had made Nightingale aware of the significance of mortality tables, she at once thought of comparing the mortality among civilians to that among soldiers. She found that in peacetime soldiers in England between the ages of 20 and 35 had a mortality rate nearly twice that of civilians. It is just as criminal, she wrote in 1857, "to have a mortality of 17, 19, and 20 per thousand in the Line, Artillery and Guards in England, when that of Civil life is only 11 per 1,000, as it would be to take 1,100 men per annum out upon Salisbury Plain and shoot them." (The 1,100 represented 20 per 1,000 of an enlisted force of 55,000.) Clearly the need for sanitary improvements in the military was not limited to hospitals in the field. By pressing her case with these statistics Nightingale eventually gained the attention of Queen Victoria and Prince Albert, as well as of the prime minister, Lord Palmerston. In spite of the passive resistance of the War Office, Nightingale's wish for a formal investigation of military health care was granted in May, 1857, with the establishment of a Royal Commission on the Health of the Army.

It would not have been possible at that time for a woman to serve on such a board. Nightingale nonetheless strongly influenced the commission's work, both because some of its members were her friends (including Sidney Herbert, the minister who sent her to the Crimea) and because she provided it with much of its information. As a statement of her own views she wrote and had privately printed an 800-page book titled *Notes on Matters Affecting the Health, Efficiency and Hospital Administration of the British Army,* which included a section of statistics accompanied by diagrams. Farr called it "the best [thing] that ever was written" either on statistical "Diagrams or on the Army."

Nightingale was a true pioneer in the graphical representation of statistics: she invented polar-area charts, in which the statistic being represented is proportional to the area of a wedge in a circular diagram. Nightingale used these diagrams, which she called her "coxcombs" because of their vivid colors, to dramatize the extent to which deaths in the Crimea campaign had been preventable. Farr was impressed with *Notes,* and much of Nightingale's work found its way into the statistical charts and diagrams he prepared for the final report of

© 1984 SCIENTIFIC AMERICAN, INC

"Last year's training manual says '*push* the reset button', J.P., ... the kid pushed the reset button . . . . ."

© 1984 SCIENTIFIC AMERICAN, INC

Not so long ago, most companies could give their new employees a training manual, a pep talk from the foreman, and let them pick up the rest of the job by trial and error.

Well, not any more. When you look at the kind of equipment on the factory floor today, a good training program is about the best insurance you can ask for. And now, thanks to a company called Wicat, the technology of training has finally come into its own.

HOW TO DOUBLE TRAINING PRODUCTIVITY.

At Wicat, we use the power of computers to make training an intense, individualized experience. Instead of getting just a fraction of an instructor's time, every trainee gets the computer's undivided attention. Instead of outdated manuals, they work with up-to-the-minute information.

Besides turning out better trained workers, you can also look forward to some very impressive cost savings. In fact, some of the world's largest corporations put Wicat learning stations in front of their trainees. And get the job done in less than half the time. For less than half the cost.

LASERDISCS VS. MANUALS AND VIDEOTAPES.

Since there's nothing like realism to make training stick, Wicat was the first to incorporate the extraordinary graphics, sound, and simulation capability of the laserdisc into computer-based training.

Just imagine, for example, a trainee flipping through a manual on sophisticated shop floor equipment. Now imagine the same employee going through a one-on-one laserdisc session that shows how the equipment works, then leads him through a practice session. Since it's only a simulation, even a colossal mistake is a positive learning experience. Instead of a total catastrophe.

COMPUTERIZE COURSES WITHOUT PROGRAMMING.

Until now, computer-based training systems demanded sophisticated programming skills. Wicat doesn't.

For example, with most other systems, it takes a skilled programmer several weeks to computerize one course. To computerize the same course on Wicat equipment, it takes a non-programmer several days.

THE TOTAL TRAINING SOLUTION.

At Wicat, we design and manufacture our own hardware. In fact, we spent four years in development to make this the most productive training system possible.

Our WISE authorizing system sets new standards for ease of course development. And our large staff is ready to create custom courseware for any application.

Our system even automates training program management: everything from scheduling to progress reporting to testing.

Best of all, you can get started for under $1,000 a month.

Call us at (801) 224-6400, ext. 209 for details. Or write to Wicat Systems, P.O. Box 539, 1875 South State Street, Orem, UT 84057.

We'll show you how to put the trial and error into your training program. And not on the shop floor.

WICATsystems™

© 1984 SCIENTIFIC AMERICAN, INC

Speak German like a Diplomat!

What sort of people need to learn a foreign language as quickly and effectively as possible? *Foreign service personnel,* that's who.

Now you can learn to speak German with the same materials used by the U.S. State Department—the Foreign Service Institute's *Programmed Introduction to German.*

The FSI spent thousands of dollars and many years developing these materials for use by members of the United States diplomatic corps. Today people in all walks of life who need to learn to speak a foreign language are turning to this outstanding audio program.

The FSI's German Course is by far *the most effective way* to learn German at your own convenience and pace. It consists of a series of cassettes, recorded by native German speakers, and accompanying textbook. You simply follow the spoken and written instructions, listening and learning. By the end of the course you'll find yourself learning and speaking entirely in German!

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

This FSI Programmed Course comes in two volumes, each shipped in a handsome library binder. Order either, or save 10% by ordering both:

- ☐ **Volume I. Programmed Introduction to German,** 10 cassettes (13 hr.), and 647-p. text, $125.
- ☐ **Volume II. Basic Course Continued,** 8 cassettes (8 hr.), and 179-p. text, $110.

(Conn. and N.Y. residents add sales tax.)

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

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

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

112 courses in 35 other languages also available. Write us for free catalog. Our 12th year.

Audio-Forum
Room 237
On-the-Green
Guilford, CT 06437
(203) 453-9794

AUDIO-FORUM®

the Royal Commission. As part of her "flank march" against the forces of resistance to medical reform, Nightingale had the statistical section of the report printed as a pamphlet and distributed widely in Parliament, the government and the army. She even had a few copies of the diagrams framed for presentation to officials in the War Office and in the Army Medical Department.

Nightingale's efforts were not in vain. Four subcommissions were established to carry out the reforms recommended in the report of the Royal Commission. The first presided over physical alterations to military barracks and hospitals: improvements in ventilation, heating, sewage disposal, water supply and kitchens. Other subcommissions drafted a sanitary code for the army, established a military medical school and reorganized the army's procedures for gathering medical statistics.

Nightingale next turned her attention to the health of soldiers in India. She and Farr began to study the sickness and mortality records of the India Office, and she sent inquiry forms to the various British stations in India for information on sanitary conditions there. In 1858 and 1859 she lobbied successfully for the establishment of another Royal Commission to look into the Indian question. Two years later she submitted to the commission a report, based on the inquiries sent to the stations in India, on the conditions that were causing among the troops in India a death rate six times higher than the rate among civilians in England: defective sewage systems, overcrowding in the barracks, lack of exercise and inadequate hospitals, among other things. The commission completed its own study in 1863. After 10 years of sanitary reform, in 1873 Nightingale reported that mortality among the soldiers in India had declined from 69 to 18 per 1,000.

Statistics, as Nightingale so effectively demonstrated, provide an organized way of learning from experience, and medical statistics can teach far more than the simple fact that unsanitary conditions kill. Uniform and accurate hospital statistics, she wrote, would "enable the value of particular methods of treatment and of special operations to be brought to statistical proof"; in short, statistics would lead to improvement in medical and surgical practice. The problem was that the statistics kept by hospitals in Nightingale's day were neither uniform nor consistently accurate. To remedy this she developed, with the aid of Farr and other physicians, a Model Hospital Statistical Form. The form was approved at the International Congress of Statistics, held in London in the summer of 1860.

The new scheme set out the basic categories of data that hospitals should collect: the number of patients in a hospital at the beginning and end of a year and the number of patients admitted during the year, the number of patients who had recovered or who had been either discharged as incurable or dismissed at their request, the number of patients who had died and the mean duration of hospital stays. Yet although the ideal of gathering uniform hospital statistics was clearly a good one, and far ahead of its time, the new scheme was never put into general practice. The proposed form itself was overly complex, and it included an idiosyncratic system for the classification of diseases devised by Farr with which many pathologists strongly disagreed. In medical science, unfortunately, Nightingale did not display the same understanding that led her to recognize the value of medical statistics; for instance, she showed no interest in the new germ theory of disease and its implications for the treatment of contagious diseases.

Nightingale's commitment to statistics transcended her interest in health-care reform, and it was closely tied to her religious convictions. To her, laws governing social phenomena, "the laws of our moral progress," were God's laws, to be revealed by statistics. Quételet's science, she taught, was "essential to all Political & Social Administration," yet political leaders were for the most part completely untrained in the interpretation of statistics. The result of such ignorance, in Nightingale's view, was legislation that was "not progressive but see-saw-y," written by officials who "legislate without knowing what [they] are doing." That is why she experimented with graphical representations, which everyone could understand, and why she struggled to get the study of statistics introduced into higher education, although her dream of a university chair in statistics did not become a reality until after her death. Even today society has not come around fully to Nightingale's point of view, as is clear from the fact that statistics has yet to become a mandatory part of public education.

Something of the religious fervor Nightingale felt for statistics is evident from her annotation of her copy of Quételet's book *Physique Sociale.* On the title page she incorporated the title into a statement of her own creed:

> The sense of infinite power
> The assurances of solid certainty
> The endless vista of improvement
> from the principles of
> PHYSIQUE SOCIALE
> if only found possible to apply on
> occasions
> when it is so much wanted

To Nightingale, Quételet was the founder of "the most important science in the whole world," because "upon it

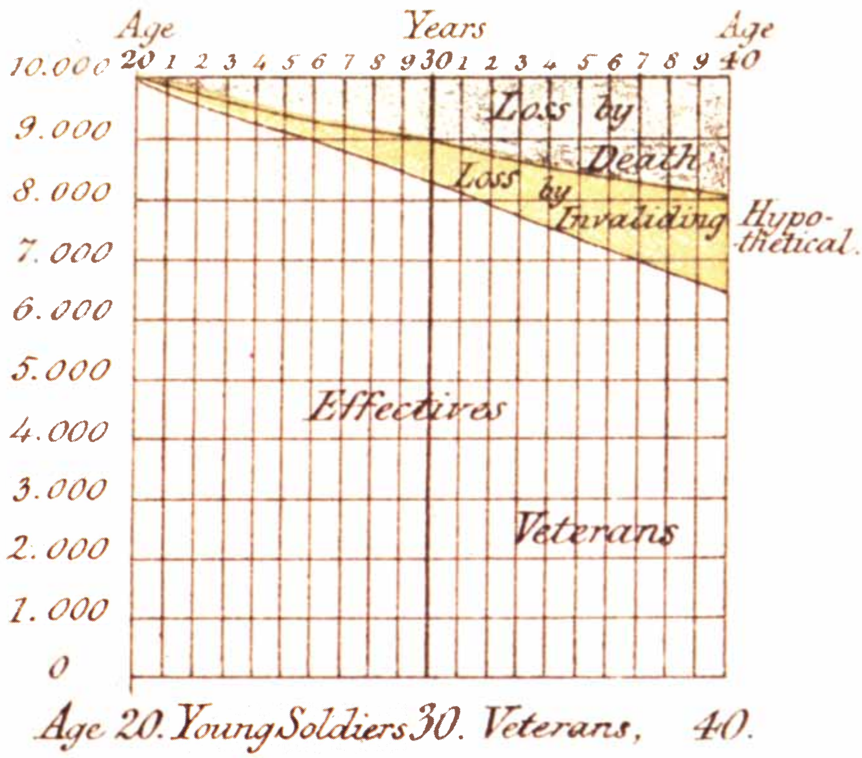
© 1984 SCIENTIFIC AMERICAN, INC

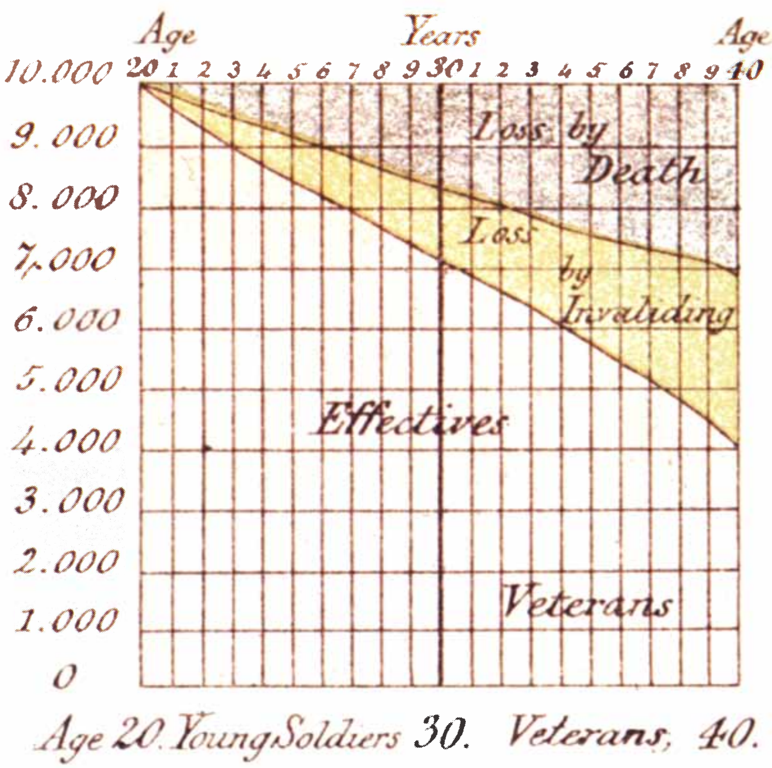
depends the practical application of every other [science]." Judging from their correspondence, the respect seems to have been mutual.

Although statistics were important to Nightingale, during her later years of being "an influential" she by her own account yearned to return to nursing, her chosen profession, her first "call from God." She could not, however, because she lived a good part of her life after her return from the Crimea as an invalid, practically bedridden.

Although Nightingale's poor health may have been related to a fever she contracted while she was in the Crimea, some have suggested that she did not have an organic illness at all, that her invalidism was neurotic or even intentional. In any event confinement to her bedroom, where she received a steady stream of visitors, did not diminish her influence or keep her from establishing the professional status of modern nursing. With money from the Nightingale Fund (almost 50,000 pounds, raised by public subscription to honor "the Popular Heroine") she was able to realize an early goal, founding the Nightingale Training School for Nurses in 1860. She could not, as she had hoped, superintend the school, but it followed her principles: "(1) That nurses should have their technical training in hospitals specially organized for that purpose; (2) That they should live in a home fit to form their moral life and discipline."

Both principles were radical in their time. That they are accepted as commonplace today is testimony to Florence Nightingale's service to nursing, which did as much as any scientific advance to improve the general quality of medical care. In view of her other passion, it is appropriate that another telling indicator of that service is statistical: in 1861 the British census found 27,618 nurses in Britain, and it listed that figure in the tables of occupations under the heading "Domestics"; by 1901 the number had increased to 64,214, and it was listed under "Medicine."





LOSS OF MANPOWER in the British army due to excess mortality and invaliding is illustrated by diagrams from the report of the Royal Commission. Both graphs assume that 10,000 20-year-old recruits are added to the force annually and that a healthy soldier's career lasts for 20 years. Each small rectangle represents 1,000 men. Under the existing unhealthy conditions (*bottom*) death (*brown*) and invaliding (*yellow*) reduce the strength of the army (*beige*) to 141,764 from its maximum size of 200,000, a loss of 29 percent. If mortality were as low as it was in the civilian population and the relation between mortality and the invaliding rate stayed the same, the report concluded, the strength of the army would increase significantly, to 166,910 (*top*).

© 1984 SCIENTIFIC AMERICAN, INC

THE AMATEUR SCIENTIST

What is a fish's view of a fisherman and the fly he has cast on the water?

by Jearl Walker

A fly fisherman who sees a fish in the water confronts the problem of where to cast the fly. The received wisdom is that the best place is just above the fish. Does the cast have to be that accurate? If it misses by a few centimeters, will the fish still see it as a fly? Robert Harmon and John Cline, who are patent attorneys in Chicago, have been looking into the optics of fly fishing. They believe the cast does have to be fairly accurate, otherwise the image seen by the fish might be too distorted by the refraction of the light rays at the surface of the water.

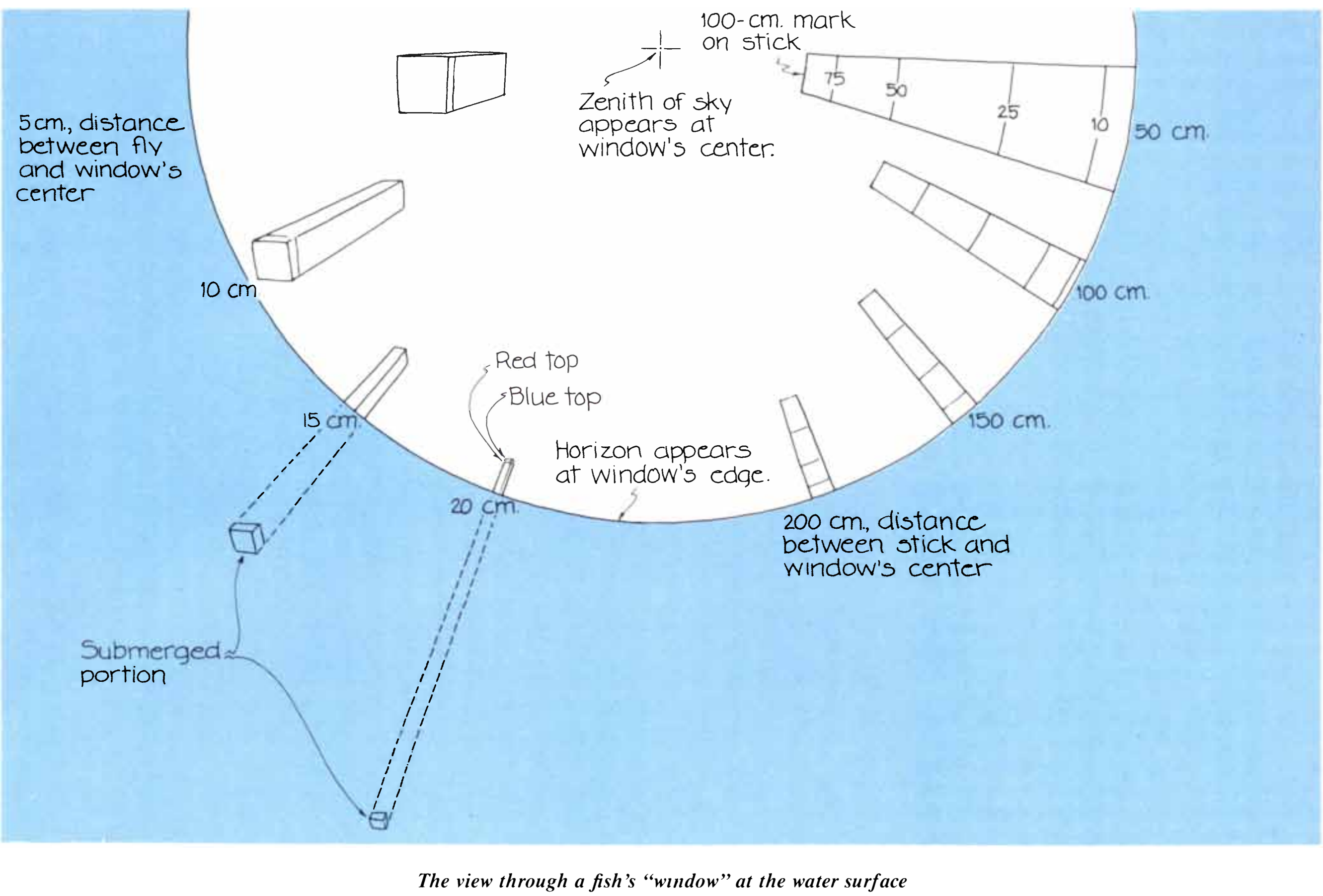
Light in a vacuum travels at only one speed (3×10^8 meters per second). Through any transparent medium, however, its speed is lower because the light interacts with the molecules of the medium. Each interaction can be considered as a brief absorption of the light.

The easiest way to describe the net delay in the passage of the light through the medium is to say the light is moving slower. For this purpose every transparent medium is assigned an index of refraction. The effective speed of light through the medium is equal to the speed of light in a vacuum divided by the index of refraction. The index of refraction of water is approximately 1.331 and of air slightly more than 1. Hence the effective speed of light through air is almost the same as it is through a vacuum, whereas through water the speed is considerably lower.

When a light ray passes through the surface of water, it is refracted (changes its direction) because of the change in its effective speed. By convention the orientation of a ray is measured with respect to a line perpendicular to the surface crossed by the ray. Suppose the ray is incident on the water surface at an angle of 42 degrees with respect to the vertical. Part of the light is reflected from the surface at the same angle from the vertical. The rest of the light refracts into the water as a ray 30 degrees off the vertical.

Other angles of incidence yield other angles of refraction. The relation is set out in the rule named for Willebrord Snell, who proposed it in 1621. According to Snell's rule, the sine of the angle of refraction in the water is equal to a fraction of the sine of the incident angle in the air. The fraction is the ratio of the respective indexes (air : water). In every case except one the angle of refraction is smaller than the angle of incidence. The exception is when the ray is incident along the vertical line; then it goes into the water with no change in direction.

Consider a ray that comes to a fish from an object a short distance above the surface of the water. If the fish can assign a position to the origin of the ray (as a human being can), it interprets the object as lying somewhere along the ray.



The view through a fish's "window" at the water surface

© 1984 SCIENTIFIC AMERICAN, INC

The object therefore appears to be at an angle in the sky higher than the true angle. The error is small if the angle of the incident ray is small and large if the angle is large.

The ray refracted the most comes from near the horizon; it is incident on the water at an angle of slightly less than 90 degrees. The angle of refraction is approximately 48.7 degrees. (The exact angle depends on the index of refraction of the water.) No ray from above the water can reach the fish at a larger angle of refraction. Hence all the rays reaching the fish from above the water fall within an imaginary cone with its apex centered on the fish's eye and with its sides 48.7 degrees off the vertical.

Harmon and Cline call the intersection of this cone with the surface of the water the "window" through which the fish sees objects above the water. A ray from the horizon passes through the edge of the window and then down the side of the cone. The size of the window varies with the fish's depth in the water. When the fish is at a depth of 10 centimeters, the radius of the window is 11.3 centimeters. A greater depth gives a wider window but cannot alter the angular size of the cone. That size is set by the refraction of the rays from the horizon.

The view of the external world that arrives at the fish is anamorphic: the magnification differs in each of two perpendicular directions. Refraction warps and repositions objects in the fish's view. Perhaps a fish can interpret the anamorphic view, realizing that the objects appearing in the window lie at some distance above the surface of the water. Perhaps instead the fish regards the objects as being on the surface. In either case what does the fisherman look like to the fish?

I investigated the question by computer, calculating what the refraction would be from each of four vertical sticks at several distances from a fish. I programmed my home computer to make the calculation on the basis that each stick extends one meter above the water and 20 centimeters below it, which is about right to simulate a fisherman standing in shallow water. The fish is assumed to be 10 centimeters below the surface, which is a reasonable depth for a feeding fish.

I first considered a stick two meters from the fish horizontally. A ray from the submerged part of the stick is not refracted and is perceived (if the fish can see that far) in its proper place. A ray from just above the waterline on the stick passes through the edge of the window and travels along the side of the imaginary cone that marks the limit of the rays reaching the fish from above the water. The fish might interpret this ray as originating somewhere back along a line making the same angle with the vertical. If it does, the waterline of the stick would seem to lie along a line 48.7 degrees from the vertical.

A ray from the top of the stick passes slightly closer to the center of the window. Its angle of refraction is about 42 degrees. The fish might see the ray as originating along a line that is a rearward extrapolation of the refracted ray. If the fish does, the top of the stick would seem to lie on a line 42 degrees off the vertical. Hence if the fish has depth perception, the stick would seem to lie somewhere in the air between 42 and 48.7 degrees off the vertical.

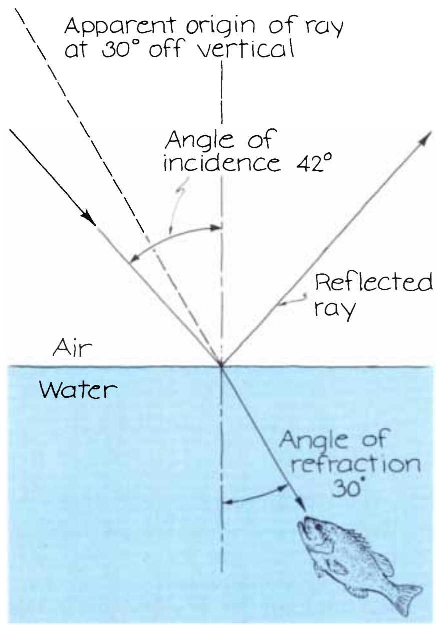
The situation is represented in the bottom illustration on the next page. The image of the stick curves between those angles. In order to leave room for the other components of the illustration the image is shown as being separated from the window by about as much as the stick actually is.

Do not take the drawing literally. I do not know if the fish can mentally extrapolate light rays. I also do not know if it can even recognize a stick for what it is. Surely a fish cannot conclude that the seemingly warped object is a vertical, rigid stick. Much of a human being's ability to assign depth and shape to objects comes from experience with those objects.

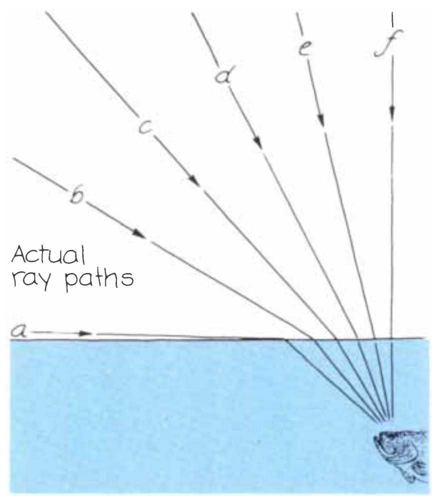
With my computer I calculated angular positions for three other sticks. In all four cases the fish sees two images of the stick. The part above the surface of the water is seen through the window. The submerged part is seen in its true position and is well separated from the image of the part above the surface. As I move a stick closer to the window the images of the two parts get closer to each other, finally merging when the stick reaches the edge of the window.

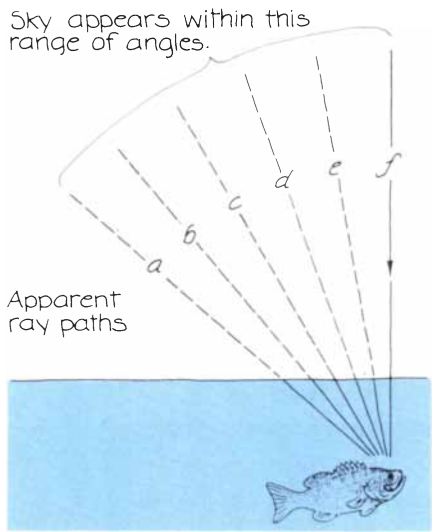
The illustration on the opposite page offers a flat view of the sticks as they are seen through the refraction of the window. A fish without depth perception or any understanding of what it is seeing probably depends on such a flat picture of the external world. To keep the sticks from overlapping in the illustration I have repositioned them so that they lie in a circle around the fish. The sizes of the sticks and the distances from them to the fish are the same as before. The submerged parts are not shown because they are too far away to fit into the illustration. Marks on the sticks indicate several heights above the waterline.

In the illustration the bottom of the part of a stick above the surface appears at the edge of the window and the top appears along a radial line and closer to the center of the window. A stick two meters from the fish is compressed into a small area. The bottom of the part of the stick above the surface is compressed more than the top because of the strong refraction of the rays from the bottom. The image of the stick takes up less than 2.5 centimeters along a radius of



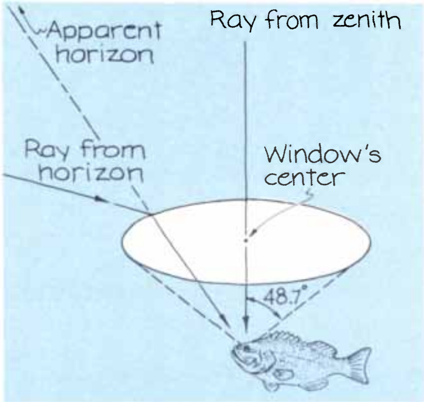
Refraction at a water surface





Actual and apparent light rays

© 1984 SCIENTIFIC AMERICAN, INC



The geometry of the window

the window. Since many other objects around a body of water would show up along the edge of the window, the stick might be lost in the clutter.

Less compression is apparent in the sticks 1.5 meters and one meter from the fish. Since they extend more toward the center of the window, however, they are noticeably tapered. The stick 50 centimeters from the fish is even more tapered and distorted. The full image of the part above the surface takes up about 70 percent of a radial line in the window and therefore must be quite noticeable to a fish.

My stick is equivalent to a short fisherman. Such a fisherman two meters or more from the fish is compressed into a miniature that occupies only a small part of the window and may be lost in the clutter at the edge. As the fisherman moves closer to the fish he takes up more angle in the fish's field of view and occupies more of the window. The submerged part of the fisherman also gets larger in the fish's field of view.

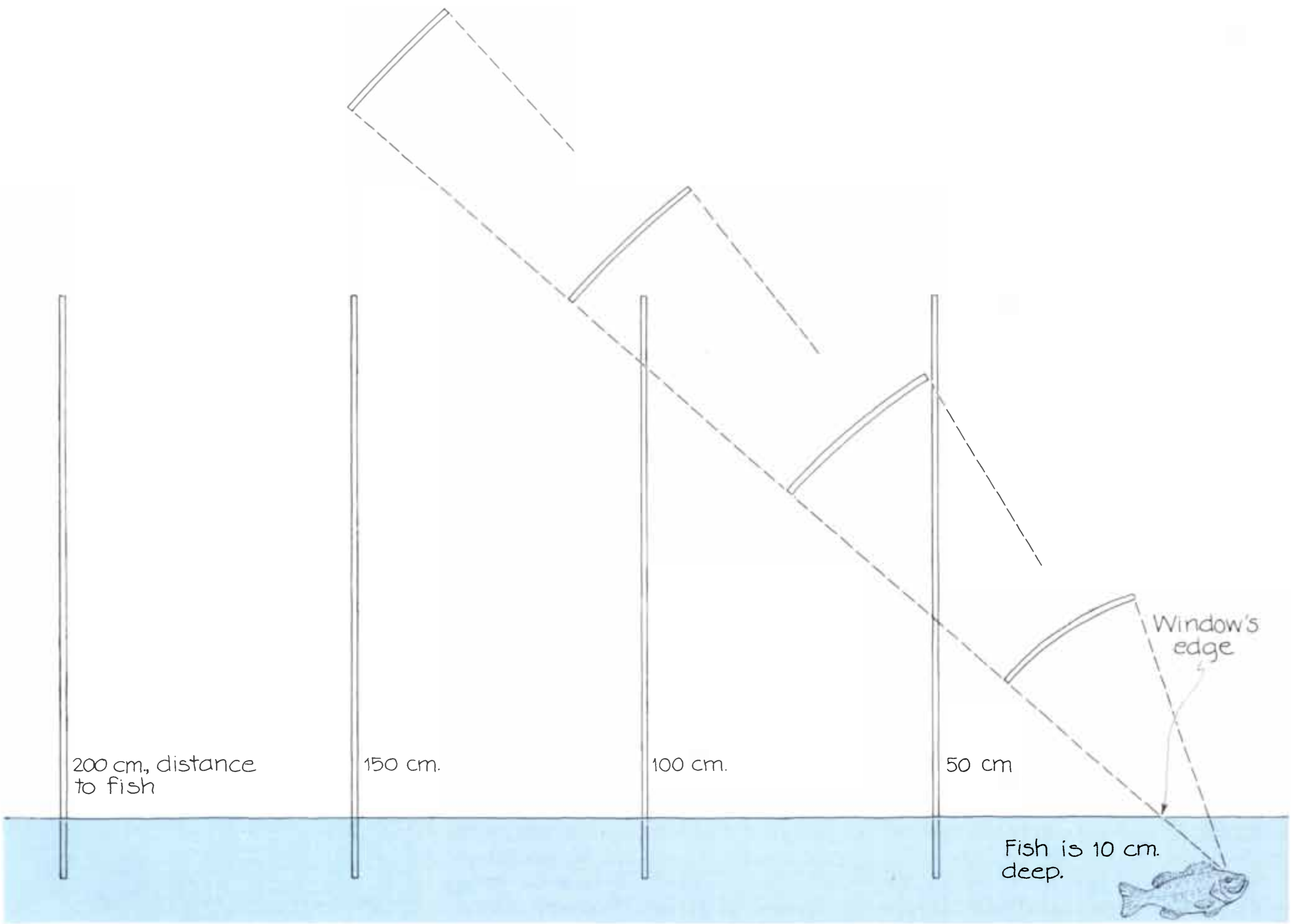
At some point the motion of one of these images warns the fish of possible danger. The motion of the part of the fisherman above the water shows up as an image that starts at the edge of the window and grows radially toward the center. Perhaps the fish watches for motion that looks as though it might cast a full image from the edge to the center.

Similar optics applies to the appearance of a fly cast near a fish. Some possibilities are represented on the left side of the illustration on page 138. For the sake of convenience I have considered a narrow rectangular fly extending 2.5 centimeters above the surface and .2 centimeter below it. (The height is about the same as that of a size 4 dry fly. The width of the fly along the surface is not important.) Although a rectangular fly is not likely to be inviting to a fish, it serves to demonstrate the distortion caused by refraction.

I programmed my computer to find the image the fly makes in the window. If the waterline of the fly is five centimeters from the center of the window, the part of the fly above the surface of the water lies across only 1.3 centimeters of a radial line in the window. The part below the surface, which is compressed, merges into the image of the part above the surface.

As the fly moves closer to the edge its image stretches. For example, when the fly's waterline is 10 centimeters from the center of the window, the image of the part of the fly above the surface takes up three centimeters along a radial line. That is more than the true height of the fly. The image of the part below the surface, still attached to that of the part above the surface, is also stretched slightly, which should make the fly more noticeable to the fish.

When the fly moves past the edge of the window, the image of the part above the surface begins to contract and that of the part below the surface separates from it. The illustration shows the situation when the fly is 15 centimeters from



How sticks in the water might look to a fish

© 1984 SCIENTIFIC AMERICAN, INC

the center of the window. The top of the part below the surface is seen at its proper distance from the center. The bottom of the part above the surface appears at the edge. The top of the fly, which is actually 2.5 centimeters above the waterline, shows up only 1.9 centimeters from the edge of the window. The fly is no longer easy to see.

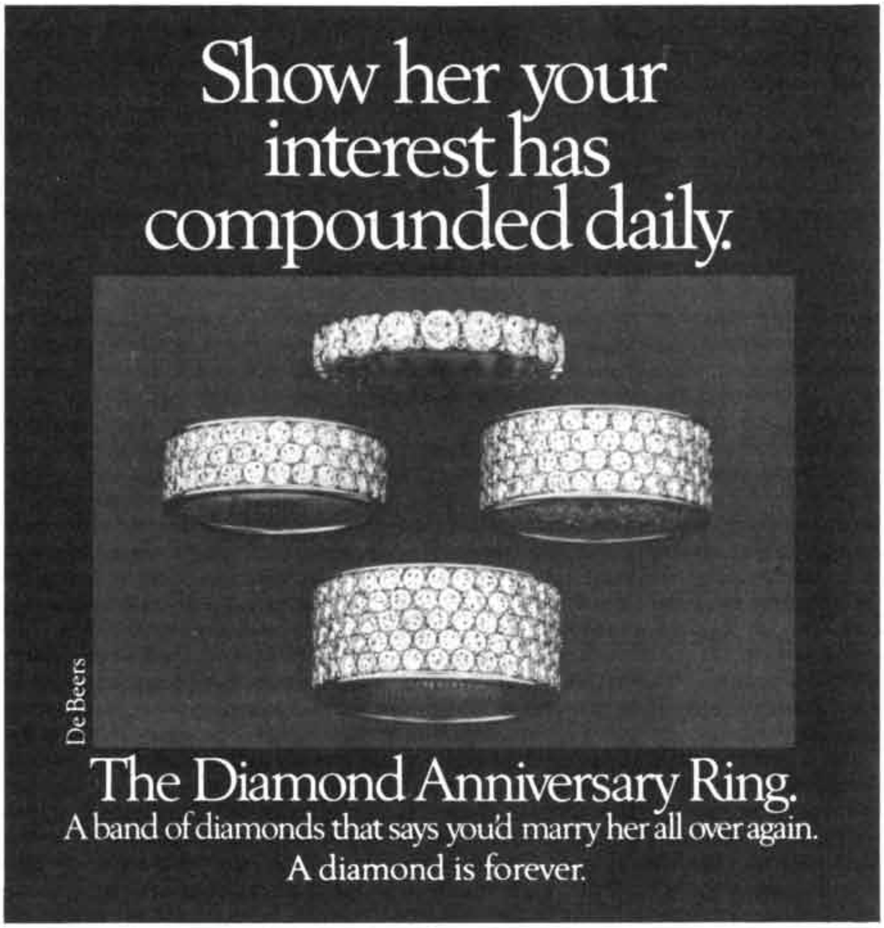
When the fly is moved to 20 centimeters from the center of the window, the apparent contraction of the part above the surface is greater. The bottom of that part still lies at the edge of the window and the top now appears at about .8 centimeter from the edge. This contraction of the image of the part above the surface gives the fish a highly distorted view of the fly. Moreover, the image of the part above the surface may be lost in the clutter at the edge of the window. Recognizing the fly is now more difficult. In addition the image of the part below the surface is well removed from the image in the window. Even if both images are still perceptible, a fish is likely to see two objects, both of them small.

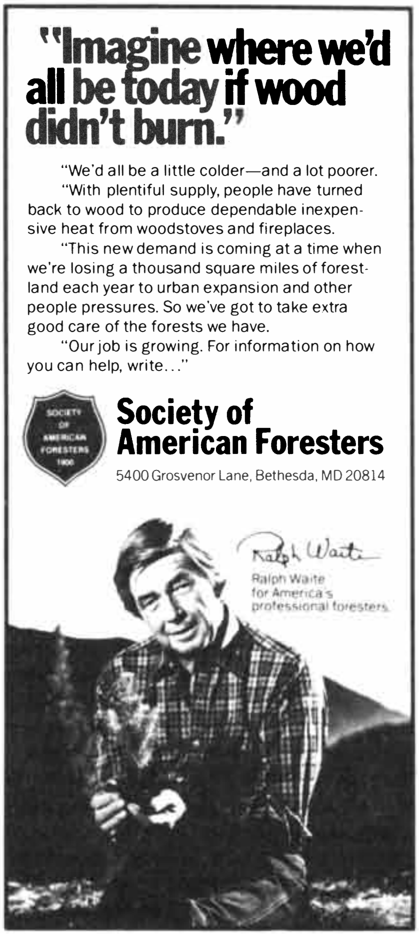
Harmon and Cline say that if you are fishing with a fly and can see the fish, cast the fly as close to it as you can. If you can put the fly within the fish's window, it may be recognizable as a fly. At least the images of the part of the fly above the water and of the part below the water are merged. If the fly lies inside the window near the edge, the image of the part above the water is magnified in the sense that its length along a radial line of the window is larger than the true height of the fly.

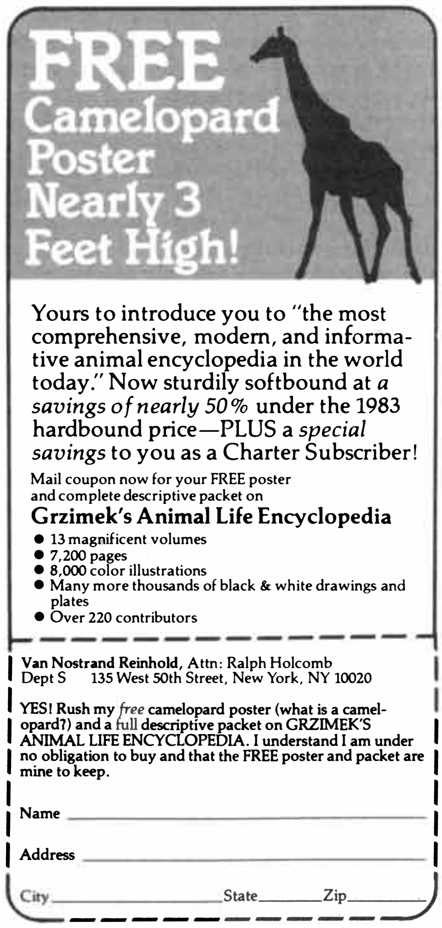
If your cast is off by a few centimeters, the fly may be outside the fish's window. The separation of the images of the part below the surface and of the part above makes the fly look less like a fly. The compression of the image of the part above the surface may even make that part so small that it is lost in the clutter at the edge of the window.

The problem is particularly difficult if the fisherman is in the same direction from the fish as the fly is; his image adds to the clutter. In this situation his only chance of attracting the fish is with the image of the part of the fly below the surface, which the fish will see without distortion by refraction. Harmon and Cline suggest it would be well if that part of the fly were brightly colored.

So far I have assumed that the index of refraction of water has a single value. In reality it differs at different wavelengths of light. Red light, at the long-wavelength end of the visible range, has an index of about 1.331. Blue light, at the short-wavelength end, has an index of about 1.343. Suppose a ray of white light, consisting of all the colors, passes into water. Refraction spreads the colors through a small range of angles. The ray with the smallest angle of refraction is blue; the one with the largest angle of







© 1984 SCIENTIFIC AMERICAN, INC

refraction is red. The colors at intermediate wavelengths have intermediate angles of refraction. This separation of colors is called dispersion.

Harmon and Cline point out that dispersion plays a minor role in the image a fish sees in its window. To investigate dispersion I considered the rays of white light extending from the top of my imaginary rectangular fly. One ray refracts at the water surface to send a red ray to the fish. Another ray refracts slightly closer to the center of the window to send a blue ray. The fish sees a colored image where the rays cross through the window. Although the blue image is slightly closer to the center of the window, the dispersion of the colored image is weak unless the fly is well outside the window. Even then the spread amounts to no more than about a millimeter in the window.

What the fish sees on the surface of the water outside the window is largely a reflection of rays that have scattered off the bottom. Although any refraction of light through the surface and into the air must obey Snell's rule, for some rays refraction is impossible. Whether or not a ray refracts depends on the angle of incidence. If the angle is less than 48.7 degrees, part of the light refracts through the surface and the rest reflects downward. According to Snell's rule, the angle of refraction (now in the air) must be larger than the angle of incidence. The angle of refraction can be as much as 90 degrees, however, which it is when the refracted ray barely skims over the surface of the water.

If the incident angle is larger than 48.7 degrees, refraction is impossible. The light can only reflect, a situation that is called total internal reflection since the light is unable to escape from the water. Any light that reflects to the fish from the underside of the window must have an angle of incidence smaller than 48.7 degrees. There part of the light also refracts into the air. A ray that reflects just at the window's edge has an angle of incidence of 48.7 degrees, sending a refracted component along the surface of the water. Any light that reflects to the fish from the rest of the surface must have an angle of incidence larger than 48.7 degrees. All this light is internally reflected. The reflections from the window region are likely to be lost in the glare of light from the sky, but the reflections elsewhere might be bright enough to give the fish a mirrorlike picture of the bottom.

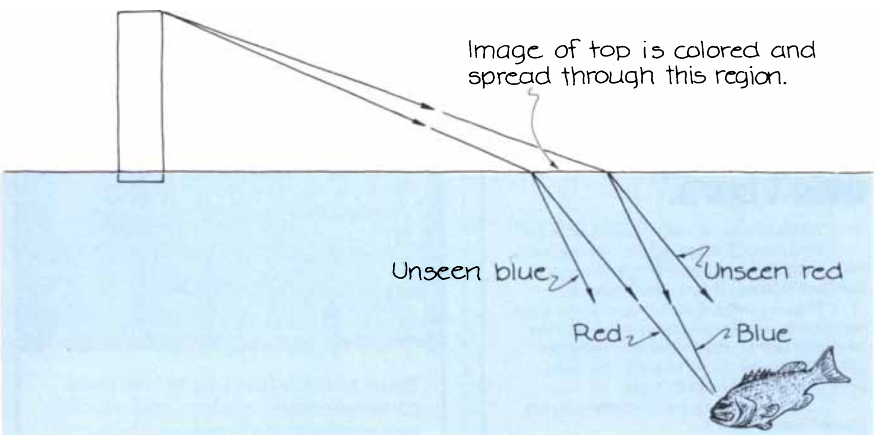
The optics I have been discussing applies to a situation in which a fish looks out through the sides of an aquarium. Here, of course, the window is in a vertical plane. The anamorphic distortion resulting from refraction would change the geometry of objects outside the aquarium. For example, an object that is in fact square would have the shape of a pincushion.

The human eye open in water does not see any of these optical distortions because it is adapted for vision in air. About two-thirds of the refraction necessary for focusing normally takes place at the surface of the eye. Since the eye has almost the same index of refraction as water, a submerged eye loses that refraction. It cannot focus on objects imaged in the window. You can regain focus if you wear a face mask to trap air next to your eye. Is there a window then? There is none if the plane of the mask is parallel to the surface of the water. When the rays pass from the water into the air in the mask, the refraction reinstates their original directions of travel. The cone limiting the rays is eliminated and therefore so is the window. You might want to investigate other orientations of the face mask.

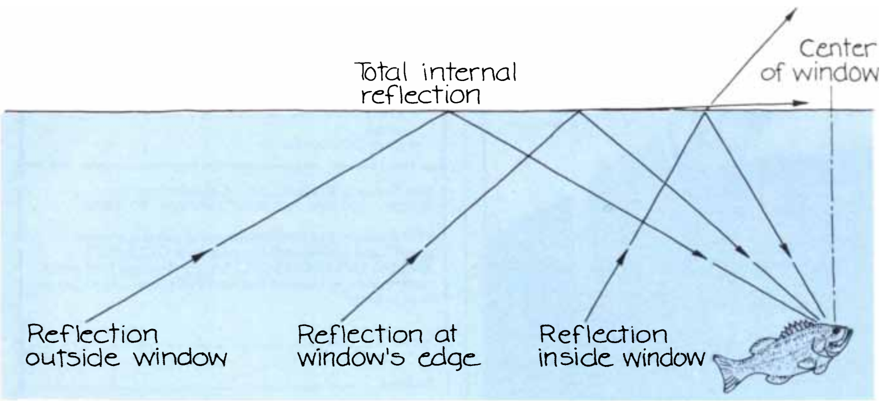
I have briefly considered another refraction problem common to fishing. Can you see a fish in its true location? The problem is crucial if you fish, as a few people do, with a bow and arrow. Should you aim the arrow directly at the fish as you see it? The answer is no. Unless the fish is just below the surface, you should aim lower in your field of view. The rays reaching you from the fish refract according to Snell's rule, ending up with larger angles with respect to the vertical than they had initially. When you receive one of the rays, you mentally extrapolate back along it to find the source, being misled into thinking that the fish is in that direction.

Lawrence E. Kinsler analyzed similar problems about the refraction of rays from a submerged object. He pointed out that the depth of an object is misjudged even when your view is from directly above it. Much of your decision about the distance to the object derives from the angle through which each eye must turn so that the eyes together can converge their lines of sight on the object. Since the rays of light are refracted before they reach the eyes, the point of convergence lies above the object, leaving you with the illusion that the object is not as deep as it actually is.

Observations from other angles also involve such an error in the assignment of depth. Kinsler's results (for a fish) are summarized in the lower illustration on the opposite page. One ray is included to represent the light that travels from the fish to the observer. Actually each eye receives a ray from a slightly different direction. The observer believes the fish lies along a rearward extrapolation

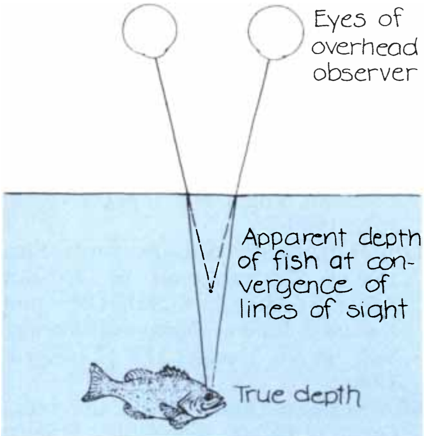


The dispersion of light rays

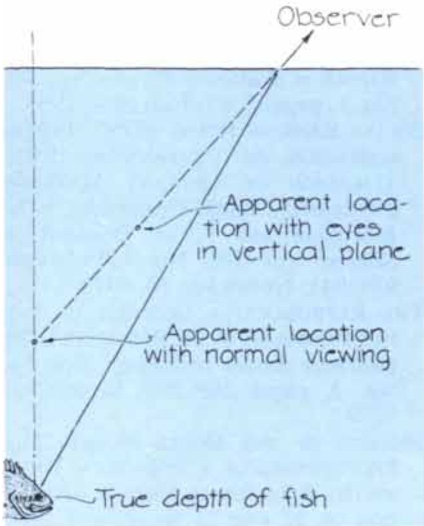


Reflections off the underside of the water surface

© 1984 SCIENTIFIC AMERICAN, INC



The illusion of depth



Where a fish is and appears to be

of the rays. In the illustration the extrapolation is indicated for the single representative ray. The convergence of the lines of sight from the eyes determines where along the extrapolation the fish appears to be. The result is that the fish seems to be higher on a vertical line running through its true location.

Such is the illusion for a normal view of a fish. Suppose the observer lies on a dock with his eyes directed downward in a vertical plane. As before the fish seems to be on a rearward extrapolation of the rays reaching the eyes. This time they seem to come from a place higher and closer to the observer.

You can check these illusions with a simple demonstration. Fill a tub with water. Look at a coin on the bottom. When your line of sight is well off the vertical, the apparent depth of the coin is obviously inconsistent with the depth of the tub. When you then move your head so that your eyes are in a vertical plane, the apparent position of the coin immediately shifts so that the coin seems to be higher and closer to you.

Consider the one bond that keeps growing each year.

The Diamond Anniversary Ring.
A band of diamonds that says you'd marry her all over again.
A diamond is forever.

To preserve your copies of SCIENTIFIC AMERICAN

A choice of handsome and durable library files—or binders—for your copies of SCIENTIFIC AMERICAN. Both styles bound in dark green library fabric stamped in gold leaf.

Files **Each file holds 12 issues. Price per file $5.95; three for $17.00; six for $30.00, postpaid.** *(Add $2.50 each outside U.S.A.)*

Binders **Each binder holds 12 issues. Issues open flat. Price per binder $7.50; three for $21.75; six for $42.00, postpaid.** *(Add $2.50 each outside U.S.A.)*

To: Jesse Jones Box Corp., P.O. Box 5120, Philadelphia, Pa. 19141
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).*

file binder

Name ________________
(please print)
Address ________________
City ________________
State ________ Zip ________

NOTE: Satisfaction guaranteed or money refunded. Allow four to six weeks for delivery.

© 1984 SCIENTIFIC AMERICAN, INC

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

COMPUTER RECREATIONS

PHYSICS AND COMPUTATION. Tommaso Toffoli in *International Journal of Theoretical Physics,* Vol. 21, Nos. 3/4, pages 165–175; April, 1982.

STATISTICAL MECHANICS OF CELLULAR AUTOMATA. Stephen Wolfram in *Reviews of Modern Physics,* Vol. 55, No. 3, pages 601–644; July, 1983.

CELLULAR AUTOMATA. Stephen Wolfram in *Los Alamos Science,* pages 2–21; Fall, 1983.

DOING PHYSICS WITH MICROCOMPUTERS. Per Bak in *Physics Today,* Vol. 36, No. 12, pages 25–28; December, 1983.

WHEELS, LIFE, AND OTHER MATHEMATICAL AMUSEMENTS. Martin Gardner. W. H. Freeman and Company, 1984.

CELLULAR AUTOMATA: PROCEEDINGS OF AN INTERDISCIPLINARY WORKSHOP. Edited by Stephen Wolfram, J. Doyne Farmer and Tommaso Toffoli, in *Physica D,* Vol. 10D, Nos. 1 and 2; January, 1984.

"NO FIRST USE" OF NUCLEAR WEAPONS

NUCLEAR WEAPONS AND THE ATLANTIC ALLIANCE. McGeorge Bundy, George F. Kennan, Robert S. McNamara and Gerard C. Smith in *Foreign Affairs,* Vol. 60, No. 4, pages 753–768; Spring, 1982.

NUCLEAR WEAPONS AND THE PRESERVATION OF PEACE: A RESPONSE TO AN AMERICAN PROPOSAL FOR RENOUNCING THE FIRST USE OF NUCLEAR WEAPONS. Karl Kaiser, Georg Leber, Alois Mertes and Franz-Josef Schulze in *Foreign Affairs,* Vol. 60, No. 5, pages 1157–1170; Summer, 1982.

ALLIANCE SECURITY: NATO AND THE NO FIRST USE QUESTION. Edited by John D. Steinbruner and Leon V. Sigal. The Brookings Institution, 1983.

THE DEFENSE BUDGET. William W. Kaufman in *Setting National Priorities: The 1983 Budget,* edited by Joseph A. Pechman. The Brookings Institution, 1983.

NO FIRST USE. Union of Concerned Scientists. 1983.

THE DYNAMIC ABYSS

THE ATMOSPHERE AND THE OCEAN. R. W. Stewart in *Scientific American,* Vol. 221, No. 3, pages 76–86; September, 1969.

THE FACE OF THE DEEP. Bruce C. Heezen and Charles D. Hollister. Oxford University Press, 1971.

DEEP CIRCULATION OF THE WORLD OCEAN. Bruce A. Warren in *Evolution of Physical Oceanography: Scientific Surveys in Honor of Henry Stommel,* edited by Bruce A. Warren and Carl Wunsch. The MIT Press, 1981.

THE ENVIRONMENT OF THE DEEP SEA. Edited by W. G. Ernst and J. G. Morin. Prentice-Hall, Inc., 1982.

HOW GENES CONTROL AN INNATE BEHAVIOR

A FAMILY OF GENES THAT CODES FOR ELH, A NEUROPEPTIDE ELICITING A STEREOTYPED PATTERN OF BEHAVIOR IN APLYSIA. Richard H. Scheller, James F. Jackson, Linda Beth McAllister, James H. Schwartz, Eric R. Kandel and Richard Axel in *Cell,* Vol. 28, No. 4, pages 707–719; April, 1982.

A SINGLE GENE ENCODES MULTIPLE NEUROPEPTIDES MEDIATING A STEREOTYPED BEHAVIOR. Richard H. Scheller, James F. Jackson, Linda B. McAllister, Barry S. Rothman, Earl Mayeri and Richard Axel in *Cell,* Vol. 32, No. 1, pages 7–22; January, 1983.

IN SITU HYBRIDIZATION TO STUDY THE ORIGIN AND FATE OF IDENTIFIED NEURONS. Linda B. McAllister, Richard H. Scheller, Eric R. Kandel and Richard Axel in *Science,* Vol. 222, No. 4625, pages 800–808; November 18, 1983.

THE DESCENT OF HOMINOIDS AND HOMINIDS

CURRENT ARGUMENT ON EARLY MAN. Edited by L.-K. Konigsson. Pergamon Press, 1980.

NEW INTERPRETATION OF APE AND HUMAN ANCESTRY. Edited by R. L. Ciochon and R. S. Corruccini. Plenum Press, 1983.

EXCITONIC MATTER

EXCITONS. D. L. Dexter and R. S. Knox. Interscience Publishers, 1965.

ELECTRON-HOLE CONDENSATION IN SEMICONDUCTORS. Carson D. Jeffries in *Science,* Vol. 189, No. 4207, pages 955–964; September 19, 1975.

THE ELECTRON-HOLE LIQUID IN SEMICONDUCTORS: EXPERIMENTAL ASPECTS. J. C. Hensel, T. G. Phillips and G. A. Thomas in *Solid State Physics,* Vol. 32, pages 88–314; 1977.

THE ELECTRON-HOLE LIQUID IN SEMICONDUCTORS: THEORETICAL ASPECTS. T. M. Rice in *Solid State Physics,* Vol. 32, pages 1–86; 1977.

THERMODYNAMICS OF EXCITONS. J. P. Wolfe in *Physics Today,* Vol. 35, No. 12, pages 46–54; March, 1982.

MUSCLE SOUNDS

A GUIDED TOUR OF THE FAST FOURIER TRANSFORM. G. D. Berglund in *IEEE Spectrum,* Vol. 6, No. 7, pages 41–52; July, 1969.

LOW FREQUENCY SOUNDS FROM SUSTAINED CONTRACTION OF HUMAN SKELETAL MUSCLE. Gerald Oster and Joshua S. Jaffe in *Biophysical Journal,* Vol. 30, No. 1, pages 119–127; April, 1980.

MUSCLES AND THEIR NEURAL CONTROL. Geoffrey Hoyle. John Wiley & Sons, Inc., 1983.

THE ADAPTABILITY OF THE HOUSE MOUSE

ADAPTATION OF MICE TO COLD. S. A. Barnett in *Biological Reviews,* Vol. 40, No. 1, pages 5–51; February, 1965.

SOCIAL RANK IN HOUSE MICE: DIFFERENTIATION BY ULTRAVIOLET VISUALIZATION OF URINARY MARKING PATTERNS. Claude Desjardins, J. A. Maruniak and F. H. Bronson in *Science,* Vol. 182, No. 4115, pages 939–941; November 30, 1973.

THE REPRODUCTIVE ECOLOGY OF THE HOUSE MOUSE. F. H. Bronson in *The Quarterly Review of Biology,* Vol. 54, No. 3, pages 265–299; September, 1979.

BIOLOGY OF THE HOUSE MOUSE: THE PROCEEDINGS OF A SYMPOSIUM HELD AT THE ZOOLOGICAL SOCIETY OF LONDON ON 22 AND 23 NOVEMBER 1979. Edited by R. J. Berry. Academic Press, 1981.

FLORENCE NIGHTINGALE

NOTES ON NURSING: WHAT IT IS, AND WHAT IT IS NOT. Florence Nightingale. Harrison & Sons, 1859.

THE LIFE OF FLORENCE NIGHTINGALE. Sir Edward Cook. Macmillan and Co., 1913.

STUDIES IN THE HISTORY OF STATISTICAL METHOD, WITH SPECIAL REFERENCE TO CERTAIN EDUCATIONAL PROBLEMS. Helen M. Walker. The Williams & Wilkins Company, 1929.

VICTORIAN SOCIAL MEDICINE: THE IDEAS AND METHODS OF WILLIAM FARR. John M. Eyler. The Johns Hopkins University Press, 1979.

THE AMATEUR SCIENTIST

IMAGING OF UNDERWATER OBJECTS. Lawrence E. Kinsler in *American Journal of Physics,* Vol. 13, No. 4, pages 255–257; August, 1945.

AT THE EDGE OF THE WINDOW. Robert Harmon and John Cline in *Rod and Reel,* No. 7, pages 41–45; March/April, 1980.

© 1984 SCIENTIFIC AMERICAN, INC

Today's Chevrolet

Cavalier. The hottest-selling Chevrolet has a cool, low price.

Chevy Cavalier. The hottest-selling Chevrolet for a lot of good reasons. Hottest-selling because of some of the most advanced design and engineering you'll find on a front-drive car today. Hottest-selling because of high-level fit and finish, thanks to dedicated workers using computerized robots and lasers. Hottest-selling because Cavalier offers more horsepower and more sedan room than the three top-selling imports. Hottest-selling because of a cool, low price.

Hot-selling performance. Cavalier goes with a high-compression, electronically fuel-injected 2.0-liter engine with an on-board computer that constantly fine-tunes your engine while you drive.

Upbeat style. Cavalier is putting the heat on the competition with new front-end styling which makes its clean wedge shape even cleaner looking. And Cavalier's handsome interior is laid out in a fashion that allows you to drive comfortably and enjoy the challenges of the road.

All at a cool, low price. Best of all, Cavalier gives you all this at a low Chevrolet price. So if you're looking for enlightened thinking and a low price in a responsive, fun-to-drive automobile, grab a hot-selling Chevrolet Cavalier.

Today's Chevrolet. Bringing you the cars and trucks you want and need. That's what Taking Charge is all about.

Some Chevrolets are equipped with engines produced by other GM divisions, subsidiaries, or affiliated companies worldwide. See your dealer for details.

© 1984 SCIENTIFIC AMERICAN, INC

© 1984 SCIENTIFIC AMERICAN, INC