SCIENTIFIC AMERICAN

ANCIENT CHINESE SHIP

ONE DOLLAR

January 1972

© 1971 SCIENTIFIC AMERICAN, INC

artist: trova

The real

During one summer week the air in Tokyo, Sydney, New York, Buenos Aires and Milan was officially unsatisfactory to breathe.

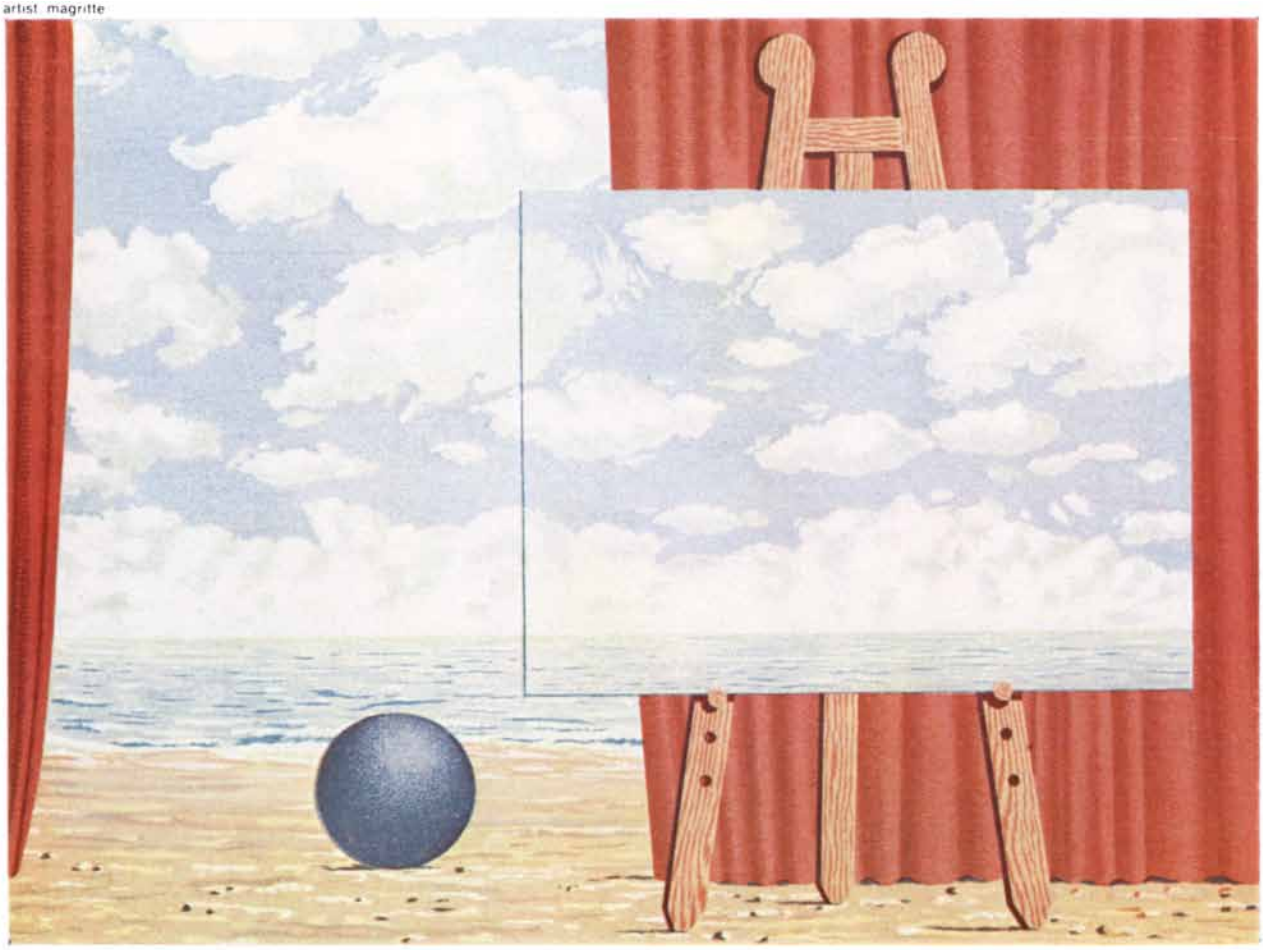
The ideal

Man will forever, for everyone, strip away the smothering blanket that separates him from the blue skies.

The cities suffer first. But air pollutants don't stop at city limits. They're blown to the four corners of the earth and the quality of air everywhere is degraded. Air pollution is everyone's problem. All of us must work at the solution. Governments, corporate management, educationai institutions, research centers and individuals must combine forces to clear and protect the air. All of us together must make blue sky thinking an urgent new reality, or live—perhaps briefly—to regret our complacency.

AtlanticRichfieldCompany

artist: magritte



© 1971 SCIENTIFIC AMERICAN, INC

Emory University got Supercomputer to take on their Chemistry Department for under $10 an hour.

It's an amazing sight.

In one of the labs you can watch a grad student analyzing complicated NMR spectra. Like Carbon 13. Or polymer structures. Checking this. Comparing that. Modifying experiments while they happen.

In another room, a group of students is learning how to work with computers. But there's no instructor. They're teaching themselves.

In yet another room, a post-doctorate fellow is performing molecular orbital calculations. Watching how a slight change in molecular structure affects the energy surface.

But the really incredible thing is that they're all working directly with one computer. All at the same time. With no computer staff.

Emory's Chemistry Department has a DECsystem-10. Supercomputer.

They figure their DECsystem-10 is as powerful as the computers most comp centers have.

But it cost half as much.

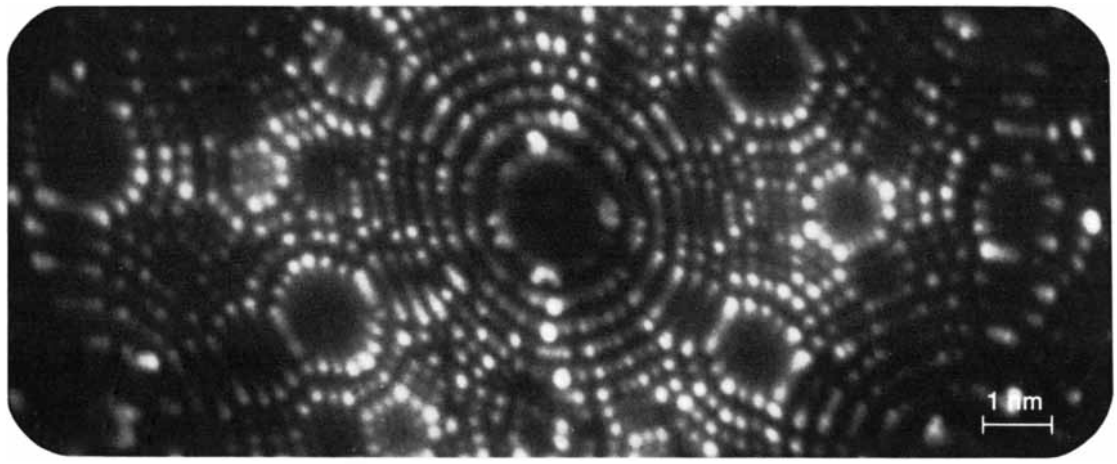
Now stop and think about that for a minute.

If Emory's Chemistry Department can get a computer that big for under $10. an hour, maybe you can afford one after all.

Look into it. Write for DECsystem-10 literature, Digital Equipment Corporation, 146 Main Street, Maynard, Massachusetts 01754.

digital

© 1971 SCIENTIFIC AMERICAN, INC



What's the matter with field ion microscopy?

It's hard to visualize an atom. Since the late fifties, physicists have been able to "glimpse" at the atoms on the surface of certain metal crystals — "see" is much too strong a word.

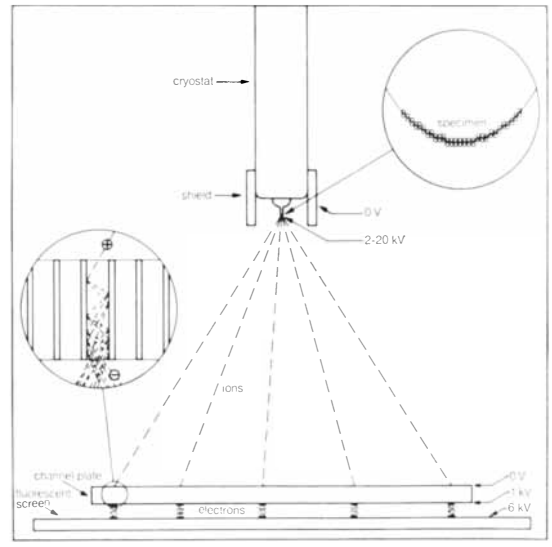
The instrument, which enabled them to do so, was Dr. E. W. Müller's field ion microscope. In this, gas atoms (usually helium) are ionized in high electrical fields near the surface of a specimen, which is in the form of an extremely fine, positively charged needle. The radius of curvature of the tip of this needle is smaller than 100 nm. Ionization occurs at less than 1 nm from the surface atoms. The ions accelerate radially towards a fluorescent screen. Here they produce a highly magnified picture of the atomic lattice at the surface of the tip. To achieve this resolution, the specimen has to be cooled to say 20 K - 80 K.

In practice, however, direct observation of the image was difficult because of its low intensity. Photographs required long exposures — in the order of minutes. Consequently, rapid changes could not be observed. Another limitation was that the use of helium restricted the choice of materials to "high boiled" metals, like tungsten and platinum. This was because the high field strength needed to ionize helium, pulls atoms from the specimen before the gas is ionized (field evaporation). Heavier imaging gases, which ionize more easily, are less efficient at producing light on the screen.

Naturally, researchers could not let this situation continue and examined various methods of improving and extending the versatility of the microscope. Dr. A. van Oostrom of Philips Research Laboratories, Eindhoven, the Netherlands, in collaboration with scientists from Cambridge University, succeeded by inserting a channel plate image intensifier in the path of the ions. This device was obtained from Mullard Research Laboratories, England.

The schematic drawing longside shows the field ion microscope with the channel plate. The specimen faces this plate, which is of metal coated glass with a large number of parallel channels about 40 diameters long. Positive ions produced by field ionization are accelerated towards the channel plate. When an ion enters a channel, it strikes the wall and produces secondary electrons. These electrons are accelerated by the electrical field and strike the wall again, producing more electrons. This multiplication process continues until the electrons finally emerge from the channel and accelerate towards the fluorescent screen. The resulting light intensity can be as much as 10^5 times greater than that of the original field ion microscope.

With liquid nitrogen cooling, the bright pattern of tungsten, imaged in neon gas, can be viewed directly (see top photo). Van Oostrom has operated the microscope at low pressures (10^{-6} torr) and with a series of imaging gases: helium, neon, argon, hydrogen, methane and nitrogen. The introduction of the channel plate enables scientists to study a wider range of materials, the adsorption behaviour of gases and the structure of biological molecules. Now field ion microscopy shows us what's the matter!



In the Research Laboratories of the Philips group of companies, scientists work together in many fields of science. Among these are: Acoustics, Cryogenics, Information Processing, Mechanics, Nuclear Physics, Perception, Solid State, Telecommunications and Television.

PHILIPS

PHILIPS

© 1971 SCIENTIFIC AMERICAN, INC

Established 1845 **SCIENTIFIC AMERICAN** January 1972 Volume 226 Number 1

ARTICLES

DEPARTMENTS

BOARD OF EDITORS Gerard Piel (Publisher), Dennis Flanagan (Editor), Francis Bello (Associate Editor), Philip Morrison (Book Editor), Jonathan B. Piel, David Popoff, John Purcell, James T. Rogers, Armand Schwab, Jr., C. L. Stong, Joseph Wisnovsky
ART DEPARTMENT Samuel L. Howard (Art Director), Alan D. Iselin, Edward Bell
PRODUCTION DEPARTMENT Richard Sasso (Production Manager), Leo J. Petruzzi and Doreen Trager (Assistant Production Managers), Pauline Ray Lipsher
COPY DEPARTMENT Sally Porter Jenks (Copy Chief), Sally Susan Fields, Dorothy Patterson, Julio E. Xavier
GENERAL MANAGER Donald H. Miller, Jr.
ASSISTANTS TO THE PUBLISHER George S. Conn, Stephen M. Fischer

PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. COPYRIGHT © 1971 BY SCIENTIFIC AMERICAN, INC. ALL RIGHTS RESERVED. 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: $10 PER YEAR. EUROPEAN SUBSCRIPTION COPIES DISTRIBUTED C/O INTERNATIONAL DISTRIBUTORS, VLAGSTRAAT 128, BORGERHOUT 2200, BELGIUM.

© 1971 SCIENTIFIC AMERICAN, INC

Every record you buy is one more reason to own a Dual.

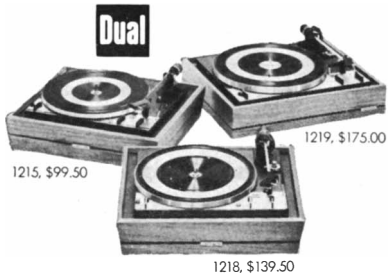
If you think of your total investment in records—which may be hundreds, even thousands of dollars—we think you'll agree that those records should be handled with the utmost care.

Which brings us to the turntable, the component that actually handles your precious records and tracks their impressionable grooves with the unyielding hardness of a diamond.

For many years serious music lovers have entrusted their records to one make of automatic turntable Dual. In fact, nearly all professionals, including record reviewers, use a Dual in their own stereo component systems.

If you'd like to know why, just mail the coupon. We'll send you an article from a leading music magazine that tells you what to look for in record playing equipment. We'll also send you a series of independent test reports on Dual.

When you read them, you'll know why Dual has made an impression on the experts and none on their records.



United Audio Products, Inc.
120 So. Columbus Avenue,
Mt. Vernon, New York 10553.

Exclusive U.S. Distribution Agency for Dual.

Please send your free literature on turntables, including test report reprints.

Name _____

Address _____

City _____

State _____ Zip _____

United Audio, 120 So. Columbus Ave., Mt. Vernon, N.Y. 10553. Dept. S.A.

THE COVER

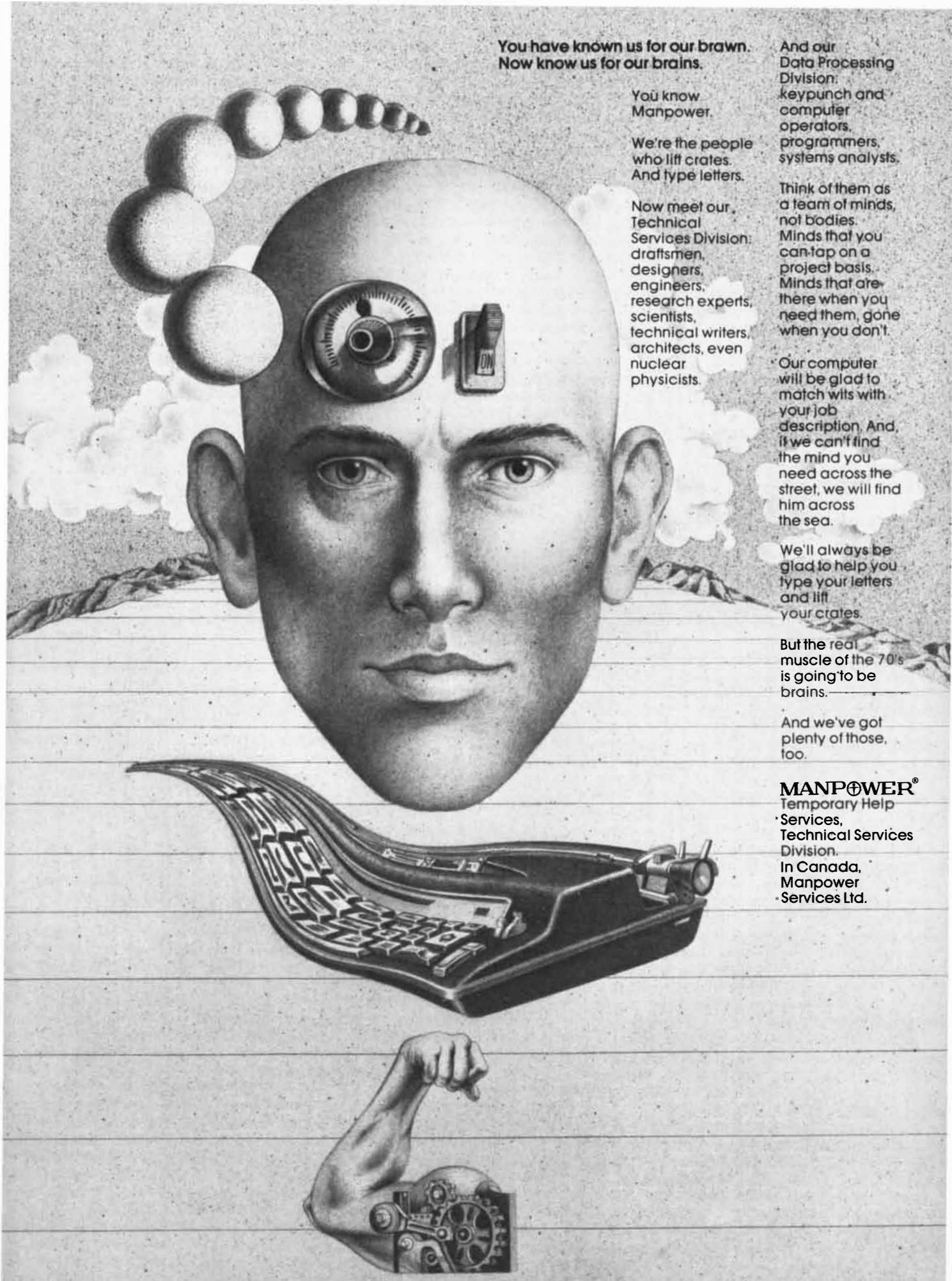
The Chinese painting on the cover, which is in the collection of the Smithsonian Institution's Freer Gallery of Art, bears on one of the great advances in the history of ships: the invention of the rudder. The oldest steering device is an oar held or mounted at the side of the stern. Rudders swinging on a post in the middle of the stern appeared in Europe early in the 13th century. The painting on the cover has been interpreted as suggesting that the Chinese had the rudder 1,000 years earlier. It is a painting of the 11th or 12th century based on a design by Ku Kai-Chih, a celebrated artist of the second half of the fourth century. At the stern of the ship is a long sweep, and above it is a flat structure that looks like a rudder that has been lifted out of the water. The nature of the steering device is unclear, but other evidence uncovered in 1958 shows unmistakably that the Chinese had the rudder even earlier. In that year a pottery model of a ship with a rudder was found in a Chinese tomb dating back to the first or second century. The evolution of steering devices and other nautical matters are discussed in the latest volume of Joseph Needham's *Science and Civilisation in China,* which is reviewed in this issue of SCIENTIFIC AMERICAN (page 113).

THE ILLUSTRATIONS

Cover courtesy of the Smithsonian Institution, Freer Gallery of Art, Washington, D.C.

Page	Source
14–15	Allen Beechel
16–17	Los Alamos Scientific Laboratory
18–19	Henry R. Myers
20–23	Allen Beechel
24	N. Sarkar, Institute for Medical Research, Camden, N.J.
25–33	Bunji Tagawa
34–35	Bernarda Bryson
36–42	Graphic Presentation Services, Inc.
54	J. Pawley, University of California at Berkeley
56–61	Dan Todd
64	James K. Koehler, University of Washington (*top left and middle*); Thomas E. Everhart and Thomas L. Hayes, University of California at Berkeley (*top right and bottom*)
65	L. McDonald, University of California at Berkeley
66–67	Hyman Hartman, University of California at Berkeley (*top*); E. R. Lewis and Y. Y. Zeevi, University of California at Berkeley (*bottom*)
69	Thomas E. Everhart and Thomas L. Hayes, University of California at Berkeley
71	Pacific Gas and Electric Company (*top*), *Geotimes* and National Electric Power Authority of Italy (*bottom*)
72	Resources and Transport Division, United Nations
73	R. M. Moxham, U.S. Geological Survey
74–77	Jim Egleson
78–79	Edward P. Ney, University of Minnesota
80–81	Alan D. Iselin
82	M. F. Ingham
83–85	Alan D. Iselin
86	Jeanne M. Riddle
88–92	Lorelle A. Raboni
93	Jeanne M. Riddle
95–100	Tom Prentiss
101	Robin Ingle
103	Tom Prentiss
104–107	Alan D. Iselin
109–111	Roger Hayward
114–115	Gabor Kiss

© 1971 SCIENTIFIC AMERICAN, INC



© 1971 SCIENTIFIC AMERICAN, INC

LETTERS

Sirs:

In your department "Science and the Citizen" for November you published a brief account of the report issued by a subcommittee of the Operations Research Society of America (ORSA), which appraised the performance of witnesses participating in the Congressional testimony about the ABM system.

The report of the ORSA subcommittee rightly deserves at least as much criticism as it makes of those it purports to investigate. The ABM decision involves a large number of disciplines of which operations research is only one. For a specialized society to take it upon itself to appraise the performance of witnesses working in all these fields is, to put it charitably, improper. We are dealing with a subject where judgment based on technical knowledge and experience, rather than pedantic calculations based on ill-founded assumptions, is essential.

The narrow approach taken by the ORSA committee is reflected in its report. All of the critical appraisal deals with that part of the ABM controversy which is the least important; this is the projection as to when and by what means the U.S. Minuteman force may become endangered by the qualitative and quantitative growth of Soviet missilery. Naturally, many informed people will differ on this judgment since a wide variety of assumptions is possible.

However, the important question—whether or not the Safeguard ABM system will buy any worthwhile protection in case the threat against Minuteman materializes—is not addressed in the ORSA report at all. It is this point that drew most of the valid criticism in Congress. The Safeguard system, even if it were to work fully as specified, can deal only with a very narrow band of threats; even this limited performance is bought at great cost. There is much ground for substantive criticism of the ORSA report; in particular it reflected enormous bias in attacking anti-ABM witnesses while omitting investigation of almost all pro-ABM witnesses, many of whom appeared not to be acquainted at all with the engineering features of the Safeguard system.

WOLFGANG K. H. PANOFSKY

Director
Stanford Linear Accelerator Center
Stanford, Calif.

Sirs:

James H. Brown's article on the desert pupfish [SCIENTIFIC AMERICAN, November, 1971] appears at a most opportune time. In June, 1971, Senator Alan Cranston introduced legislation proposing the creation of the Desert Pupfish National Monument in order to protect these unique fish from certain destruction if farming and mining activities in the Ash Meadows region of California and Nevada are not halted.

Recently I called the Senator's attention to a series of unique species of flowering plants that are restricted to the area outlined by his legislation for the National Monument. These plants are equally threatened by man's activities in Ash Meadows. Since 1970 three new species of plants have been described: *Astragalus phoenix, Machaeranthera ammophila* and *Grindelia fraxino-pratensis.* These join the endemics already discovered in the area: *Sisyrinchium funereum, Nitrophila mohavensis, Ivesia eremica, Mentzelia leucophylla* and the more widespread *Cordylanthus tecopensis.* Some plants in the area are still being studied, and at least two remain to be named and their taxonomy understood.

Cranston has brought the danger these flowering plants face to the attention of Congress, but he still needs active support for bill S. 2141 from the scientific community.

JAMES L. REVEAL

University of Maryland
College Park, Md.

Sirs:

My congratulations to Don L. Anderson on his lucid and well-illustrated article on the San Andreas fault [SCIENTIFIC AMERICAN, November, 1971]. I submit my own attempt at translating geophysics into geopoetry.

All aboard, you Californians,
for we're headed out to sea,
From foggy Point Arena to La Paz,
Forget those state and county lines
for they will cease to be,
And the statutes of the earth
will be our laws.

All aboard, you Californians,
for we're splitting from the rest,
Setting sail upon an ocean of basalt,
Our helmsman is the granite rock
on Inverness' crest,
And our bearing line
the San Andreas Fault.

All aboard, you Californians,
and bring your winter clothes,
For we're headed you can bet
for colder climes,
And the chaparral and palm trees
that the Angelino knows,
Will be left behind for fern
and forest times.

All aboard, you Californians,
for those who miss this trip,
Will be drooling at the Cliff House
on a bench,
As we wave while drifting past them
on our golden-stately ship,
Half a centimeter per month
on toward the trench.

But beware, you Californians,
for our voyage is surely doomed,
Jump across the Fault
to safety while you may,
'Neath the Arc of the Aleutians
our ship will be entombed,
To reappear some other geo-day.

STEPHEN M. ROWLAND

Cupertino, Calif.

Scientific American, January, 1972; Vol. 226, No. 1. Published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Gerard Piel, president; Dennis Flanagan, vice-president; Donald H. Miller, Jr., vice-president and secretary; George S. Conn, treasurer.

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

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

Offprint correspondence and orders should be addressed to W. H. Freeman and Company, 660 Market Street, San Francisco, Calif. 94104. For each offprint ordered please enclose 25 cents.

Change of address (or other subscription correspondence) should be addressed to Circulation Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017. Please notify us four weeks in advance of change. If possible, kindly furnish an address imprint from a recent issue. Send both old and new addresses or complete and mail Post Office Form 3578.

Please enter change of address ☐
Please enter subscription ☐
☐ 1 year, $10 ☐ 2 years, $18 ☐ 3 years, $25

NAME

NEW ADDRESS

OLD ADDRESS

© 1971 SCIENTIFIC AMERICAN, INC

Tough tank for a snow goer.

Our Marlex Marketing Team helped the manufacturers create it. How can our team help you?

Maybe you don't make snowmobiles. Or plastic fuel tanks. Neither do we.

But our Marlex Marketing Team can work with you to improve the products you do make, or develop the products you plan to make.

Just as we've done recently for snowmobiles.

One thing the industry needed was a fuel tank that was light but durable, and safe. One that offered substantial savings in fabrication, materials and shipping costs.

The industry got it—with the help of our Marlex Marketing Team: the field sales engineers and technicians, mold designers, production specialists, resin experts.

Working together, they brought to commercial reality a fuel tank that's rotationally molded from a new cross-linkable high-density polyethylene, Marlex® CL-100. The ideal resin for many products requiring superior impact strength at low temperatures and superior-stress crack resistance.

That's what Marlex Marketing Team performance is all about. And that's how it helps make Marlex resins worth more to you. For help, call:

Robert F. Lynch, Fairfield, N. J. 201 227-6600
Bill Shriver, Southfield, Mich. 313 353-6666
Ralph Porter, Villa Park, Ill. 312 834-6600
Harry House, Memphis, Tenn. 901 682-8317
David Day, Los Angeles, Calif. 213 685-6380

PHILLIPS 66

At Phillips 66 it's performance that counts®

© 1971 SCIENTIFIC AMERICAN, INC

Unusual tools, hard-to-find items.

Send for free catalog!

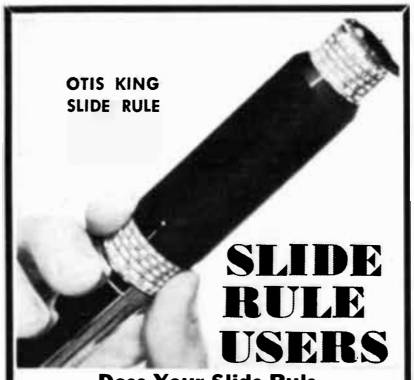
Diversified collection-

tools of all kinds
precision instruments
essential shop items

. . . convenient one-stop shopping from your desk.

NC Flasher

National Camera
Dept. JSB
Englewood, Colorado, 80110



SLIDE RULE USERS

Does Your Slide Rule Provide Enough Accuracy?

Or must you check important calculations? Remarkable cylindrical slide rule gives accuracy you need! Scales 66" long give **reliable results to 4 or even 5 figures.** Has C, D, & 5-place Log scales, each **over 6X longer** than usual. ■ Patented spiral construction packs 3 66" scales in **easily read** pocket size calculator. Only 6" closed, 10" open. **Simple to use.** Full instructions. Quickly learned. No seldom used scales getting in way. **Ideal for lab or field!** ■ Lifetime chromed steel body can't warp or break. No fragile glass cursor. **Precision crafted in England.** ■ Gives fast, precise answers. Stops errors.

Why Not Solve Your Calculating Problems Today?

Satisfaction Guaranteed! Only $24.95 ppd. U.S.

☐ Send me the **OTIS KING SLIDE RULE** for immediate delivery. I enclose $24.95. Refund within two weeks if not delighted. *(In Calif. add 5% tax.)*

Name_____

Address_____

_____ **Zip**_____

CALCULATOR COMPANY—Dept. 3A
Box 593, Lakewood, Calif. 90714

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

JANUARY, 1922: "Many methods of estimating the age of the earth have been put forward; scarcely any respectable geologist or physicist has failed to suggest one. They all depend upon mere arithmetic, to the extent that we observe some condition such as the salinity of the ocean or the thickness of the rocks or the amount of erosion in some deep gorge and then calculate how long it would take this condition to be brought about. They all are weak to the extent that they assume that the process, whatever it may be, has taken place in the past at the rate of its observed occurrence in the present. The student of radioactivity now puts forward a phenomenon which, in the very nature of things, must always have proceeded at the same rate, namely the disintegration of radioactive elements into other elements of lower atomic weight. On this basis an estimate of the world's age has been made which seems to be far more rational than any ever made before, and which we can well afford to accept as marking an epoch in the discussion of this problem."

"A recent writer in *The American Machinist*, with the success of two such diverse bits of apparatus as the player piano and the monotype machine before him, makes the suggestion that this principle is deserving of an effort to extend it more generally into machinery practice. He thinks that automatic lathes of all descriptions are a particularly promising field for the innovation, and he points out that, theoretically at least, there is not the slightest limit to what the tape can instruct the machine to do."

"Most unusual effects can be obtained by taking motion pictures at a high rate of speed and projecting them at the standard speed of 16 pictures, or 'frames,' per second. There are various special cameras for this kind of work, but until now their construction has been considered the deepest secret. A new high-speed camera for motion-picture work, designed to overcome the objectionable features of the makeshift devices heretofore used, has been invented by a Pacific Coast man. The chief characteristic of the new camera is a straight pull on the film. With a camera capable of taking 14 feet of film (224 separate frames) per second, it will be readily appreciated that there is a terrific strain on the film, to say nothing of similar destructive influences on the camera itself. With the film magazines mounted at the back, the new camera takes the film in a direct line back to the take-up magazine, thus providing an instrument that has only one turn for the film. With a turning mechanism that is geared very high, each turn of the handle exposes seven feet of film. The handle is turned at the same rate as with an ordinary camera: two turns per second. Hence to the operator there is no difference in manipulation. Of course, when such a film is projected at the ordinary rate of speed, objects move very slowly, making the instrument of incalculable value in scientific and industrial work."

"There are two ways, and only two, in which we can continue to meet our wood requirements, says Uncle Sam. One is to grow more wood; the other is to use more effectively the supplies now available. We must see that our remaining 137,000,000 acres of virgin forest are cut in such a way as to maintain the productivity of the land, and that our 81,000,000 acres of wholly idle, and our 235,000,000 acres of partially idle, forest lands are put to work. Furthermore, we must see to it that, of the 24 billion cubic feet removed from the forests each year, 70 to 75 per cent no longer goes to waste in the form of sawdust, shavings and other mill refuse."

SCIENTIFIC AMERICAN

JANUARY, 1872: "Of all the fearful diseases that scourge the human race, small pox ranks among those that are justly feared most. It would be strange did not some exaggeration creep into the statements of the progress of small pox which reach us from every direction, but there is no doubt an unusual prevalence of the disease in both England and America. So far as can be gathered, the disease is not remarkably malignant in type, and the average of deaths from it is not greater than is common. It cannot, however, be predicted that the disease will

© 1971 SCIENTIFIC AMERICAN, INC

not assume at any time the form of a malignant epidemic. In view of these facts, it is pertinent to inquire if the periodical spread of small pox cannot be prevented. We think there is no fact better established in medical science than that persistent revaccination will practically exterminate the disease. Had we a law compelling vaccination and revaccination, and could its enforcement be insured, we should soon cease to hear of the ravages of a scourge so dreadful. As it is, it is quite doubtful if such a law could be enforced if enacted. The best we can do at present is for the intelligent to protect themselves without law, against the neglect arising from the prejudices of the ignorant, and the disease will be mostly confined to the ignorant and the careless."

"The habits of the present generation are such as to give rise to more refuse matter and poisonous products than those of previous ages. The fuel we use, the articles we manufacture and the waste of sewage combine to create more impurities than were known to our forefathers, and if it were not for the fact that science has given us remedies nearly in proportion to the increased evil, our population would diminish under the high-pressure system which at present prevails. It is evident that the contruction of great chimneys to carry off foul gases, together with the immense loss to agriculture, could be avoided if we applied the remedy at the outset, and that would be by using the ounce of prevention and disinfecting all animal matter by dry earth, and never allowing it to pollute our waters. Whereas our water arrangements appear to us individually a great convenience, they are collectively the source of most of our diseases and ought to be regulated. In spite of all precautions much impurity finds its way into the sewers, but the worst evil could be stayed and disinfecting rendered substantially unnecessary. The true remedy is to stop filling the sewers with matter that no power can afterward cleanse."

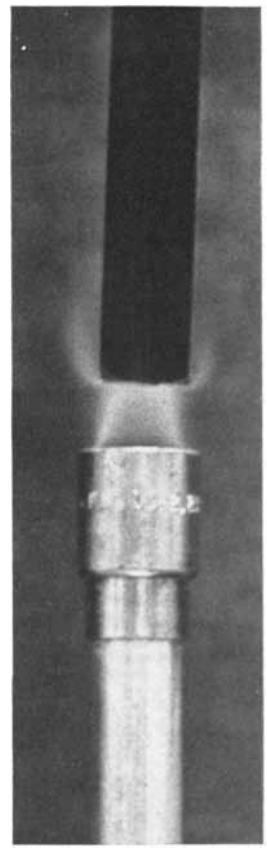
"So much has been said about the superiority of asphalt over granite as a road-making material that those who have not the opportunity to form opinions from actual observation will be glad to obtain as a basis of comparison some reliable statistics relating to the actual duration of granite roads under service. On the whole the comparison seems favorable to the asphalt roads, not only in point of cost but also in comfort."

PLENCO

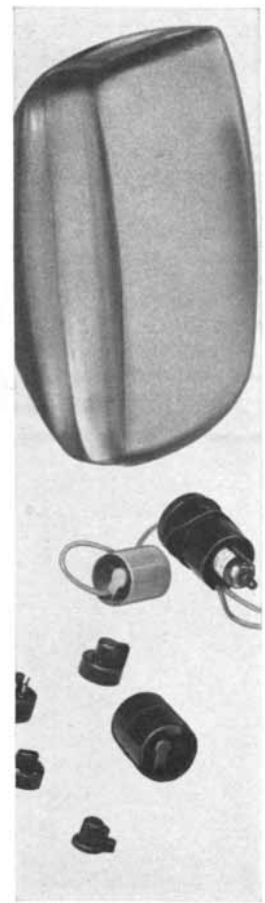
■ FLAME-RETARDANT
■ TRACK-RESISTANT
■ X-RAY SHIELDING

PHENOLIC, MELAMINE-PHENOLIC AND ALKYD MOLDING COMPOUNDS

Thermosets formulated to provide an extra factor of product safety. With greater range in choice of properties.



Improved arc-resistance, excellent resistance to tracking under the influence of conductive contaminants, X-radiation shielding, and humidity-resistant characteristics are among the qualities offered by Plenco in this series of compounds. All are recognized under the Component Recognition Program of Underwriters' Laboratories, and were developed to help answer current concern for increased safety in electronic and electrical products. Where design and reliability requirements are being upgraded, Plenco gives the designer a wide range of properties and cost. Permits him to optimize the best approach to his problem.



PLENCO
THERMOSET PLASTICS

PLASTICS ENGINEERING COMPANY
1607 Geele Ave., Sheboygan, Wis. 53081

Through Plenco research . . . a wide range of ready-made or custom-formulated phenolic, melamine, epoxy and alkyd thermoset molding compounds, and industrial resins.

© 1971 SCIENTIFIC AMERICAN, INC

The 80" (f/10-2000mm) focal length of this scope is accomplished by multiple passes within the short 16" tube. 25# total weight and 9x12x24" secured size. Complete with fork mount and clock drive.

✓ The Schmidt-Cassegrain optical system of the Celestron 8 assures the sharpest 40X telephoto images. Distortion-free and perfect color rendition. Full 35mm format.

✓ For the wife — a nature telescope to study the frolicking of a bobolink at 150 feet or a bee pollinating a flower at 30 feet. She'll be sold.

✓ For the kids — a real educational bonus. They'll study the wealth of detail of the Moon, rings of Saturn, twirling satellites and belt structure of Jupiter, and the filamentary detail of a remote nebula.

"Celestron Techniques" published quarterly. Current issue contains articles on Lunar, Solar, and terrestrial telephotography and many full-color Celestron 8 shots of these subjects. $2.00 for 4 issues.

Celestron 8 — price $850 — For details write to:

Celestron Pacific 2430 Amsler Torrance, Calif. 90505

TALL or BIG MEN Want New Fashions In Your Exact Size? TRY KING-SIZE

If you're 6'3" or taller: need sleeve lengths to 38", waists 32 to 60", shoes to 16 EEE— the famous KING-SIZE Co. can fit you with the latest fashion Dress and Sport Shirts, Double-Knit Slacks, Sport Coats and Jackets — all proportioned for Tall and Big Men only! Send for our FREE 144-page Color Catalog today. No obligation to buy. No salesman will call. Everything Fully Guaranteed.

SEND FOR FREE CATALOG

Please rush me your FREE Catalog of Apparel and Footwear for Tall and Big Men.

Name _____

Address _____

City _____ State _____ Zip _____

The KING-SIZE Co. 552 KING-SIZE BLDG. BROCKTON, MASS.

THE AUTHORS

HENRY R. MYERS ("Extending the Nuclear-Test Ban") is a consultant on arms control and other subjects; he writes that for the past four years he has worked on "interdisciplinary matters involving the environment, arms control and military spending, earthquake effects, technical assistance to developing countries, the economic utility of nuclear explosives in nonmilitary projects, the trans-Alaska pipeline and the SST." Myers received his degrees in physics: a bachelor's degree from the Massachusetts Institute of Technology in 1954 and a Ph.D. from the California Institute of Technology in 1960. Before taking up his present work he spent three years as an experimental physicist at the Cambridge Electron Accelerator and six years as a technical adviser with the U.S. Arms Control and Disarmament Agency.

HOWARD M. TEMIN ("RNA-directed DNA Synthesis") is professor of oncology and of cancer research at the University of Wisconsin. Born in Philadelphia, he received his bachelor's degree at Swarthmore College in 1955 and his Ph.D. from the California Institute of Technology in 1959. He was a postdoctoral fellow at Cal Tech for a year before going to Wisconsin.

JEAN LIPMAN-BLUMEN ("How Ideology Shapes Women's Lives") is a postdoctoral fellow in the department of sociology at Stanford University. She writes: "In addition to my empirical work on women, I am particularly interested in testing a theoretical framework for analyzing how roles change within society. This applies not simply to sex roles (although my original ideas about it stemmed from my thinking about sex roles) but to all categories of social roles, including occupational roles, generational roles, social class roles and so forth. Outside of my work interests I am always involved in a wide range of things stemming from my children's interests and activities. I have learned a great deal about dinosaurs, space exploration and Abraham Lincoln in this way. When I have spare time, I play the piano (mostly Bach, Beethoven and Mozart) and write poetry. I also am a tennis buff and try to play five or six times a week." After graduation from Wellesley College in 1954 with a degree in English literature she obtained her master's degree in sociology at Wellesley and her Ph.D. in sociology from Harvard University. She did the research for the study that she describes under a grant from the Sperry-Hutchinson Fund while she was project director of the Life Plans Study at the Radcliffe Institute.

THOMAS E. EVERHART and THOMAS L. HAYES ("The Scanning Electron Microscope") are at the University of California at Berkeley; Everhart is professor of electrical engineering and Hayes is a biophysicist at the Donner Laboratory and a lecturer in the department of medical physics. Everhart obtained his bachelor's degree at Harvard College in 1953, his master's degree at the University of California at Los Angeles in 1955 and his Ph.D. from the University of Cambridge in 1958. He writes that his avocations include photography and "making sure all appliances used by a family of six stay in good repair," adding that his "outside activities include five years in Y-Indian guides with two sons and service one year as superchief of the Tamalpais nation." Hayes received his degrees at Berkeley—his A.B. in physics in 1949 and his Ph.D. in biophysics in 1955. He writes of his "interest in applying the techniques of art in a fundamental way" to the study of small biological systems.

JOSEPH BARNEA ("Geothermal Power") is director of the Resources and Transport Division of the United Nations. After completing work at a technically oriented high school in Germany he studied economics in Germany, Britain and Switzerland, where he obtained his Ph.D. He worked in Palestine and then Israel, specializing in natural resources of the Middle East, and joined the UN in 1951. He writes of his interest in "new concepts of exploring for natural resources" and in "interdisciplinary and multipurpose" approaches to both exploration and development.

M. F. INGHAM ("The Spectrum of the Airglow") is a member of the staff of the Sedbergh School in England, having taken up that work in the current academic year after serving as a lecturer in the department of astrophysics at the University of Oxford. He was graduated from the University of Cambridge and went to Oxford in 1960. "My recreations are gardening and woodwork," he writes. "I gave up mountaineering when I acquired dependents."

HORACE W. DAVENPORT ("Why the Stomach Does Not Digest Itself") is professor of physiology and chairman of

© 1971 SCIENTIFIC AMERICAN, INC

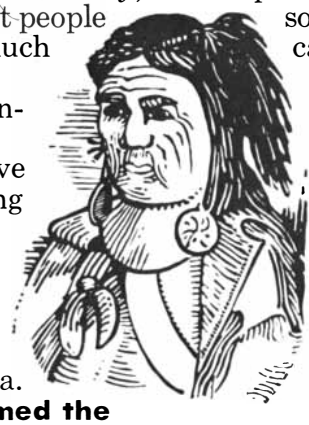
the department of physiology at the University of Michigan Medical School. He holds degrees from the California Institute of Technology and the University of Oxford, including two doctorates: a Ph.D. from Cal Tech in 1939 and a D.Sc. from Oxford in 1961. Before going to Michigan he taught at the University of Pennsylvania, the Harvard Medical School and the University of Utah. Davenport is the author of two textbooks, *The Physiology of the Digestive Tract* and *ABC of Acid-Base Chemistry.* He writes that his hobbies are cooking and fencing and adds: "I don't engage in any good works or promote any good causes, thinking that the one talent which is death to hide has not been lodged with me useless. I try to remain a cultivated man and to transmit to the young something of the light of human experience, which may illuminate their current problems."

ROBERT B. ECKHARDT ("Population Genetics and Human Origins") is assistant professor of anthropology at Pennsylvania State University. He entered anthropology after majoring in preveterinary medicine at Rutgers University; in graduate work at the University of Michigan he obtained a master's degree in human genetics in 1966, a master's degree in anthropology in 1967 and a Ph.D. in both subjects in 1971. He writes: "Outside of professional activities I must confess that I am addicted to none of the socially acceptable leisure-time activities such as golf, sailing and bridge; instead I enjoy a quite unfashionable but abiding interest in the field of automobile safety, which has become specifically focused on modifying our ancient but faithful 1955 Chevrolet sedan to improve its crashworthiness and overall safety. Since becoming involved in this task I've met a fascinating variety of nonacademic types—automobile wreckers, automobile racers, rallyists, manufacturers of technical equipment—most of whom are convinced that I'm a trifle odd to be more interested in stopping quickly than in going swiftly." Eckhardt says his roll bar and shoulder harness appear to make young drivers think he is interested in drag racing.

N. SIVIN, who in this issue reviews *Science and Civilisation in China. Volume 4: Physics and Physical Technology; Part III: Civil Engineering and Nautics,* by Joseph Needham, with the collaboration of Wang Ling and Lu Gwei-Djen, is associate professor of humanities at the Massachusetts Institute of Technology.

Write down all you know about Indiana here.

You probably noticed that the space we left for writing about Indiana is exceedingly small. Frankly, that's because most people just don't know much about us. Alright, maybe we have "underwhelmed" you with our progressive image. But, starting now, we're out to dispel the rumor that Chief Potawatomi is alive and on the warpath in Indiana.



We tamed the wilderness years ago.

Oh, we still have our Indian uprisings. Every summer in Indianapolis. But they happen on a baseball diamond. And we take pennants, not scalps. There's also ample opportunity to enjoy football, basketball, racing and golf. And if it's culture you want, we've got the theatre, the symphony, the opera, and one of the best art museums you'll find anywhere. Who'd have believed it.

Indiana.
Where the sun also rises.
And you can see it.

Fortunately, there are a lot of things Indiana doesn't offer. The smog doesn't block out the sun. Our kids don't hack and cough from the filth in the air. You don't need a bodyguard to stay out after sunset. And we don't live packed on top of one another like so many sardines in a can. We like Indiana the way it is. You will, too.

You don't have to sacrifice your profits to stretch your legs.

Your industry will grow in Indiana. First, because you're close to the raw materials your business needs. Second, because we've got a labor force that gets things done fast. Without losing quality. They're happy where they work, because they like where they live. Third, because we've got the transportation systems to get your finished product to market with no delays, overland or across the ocean.

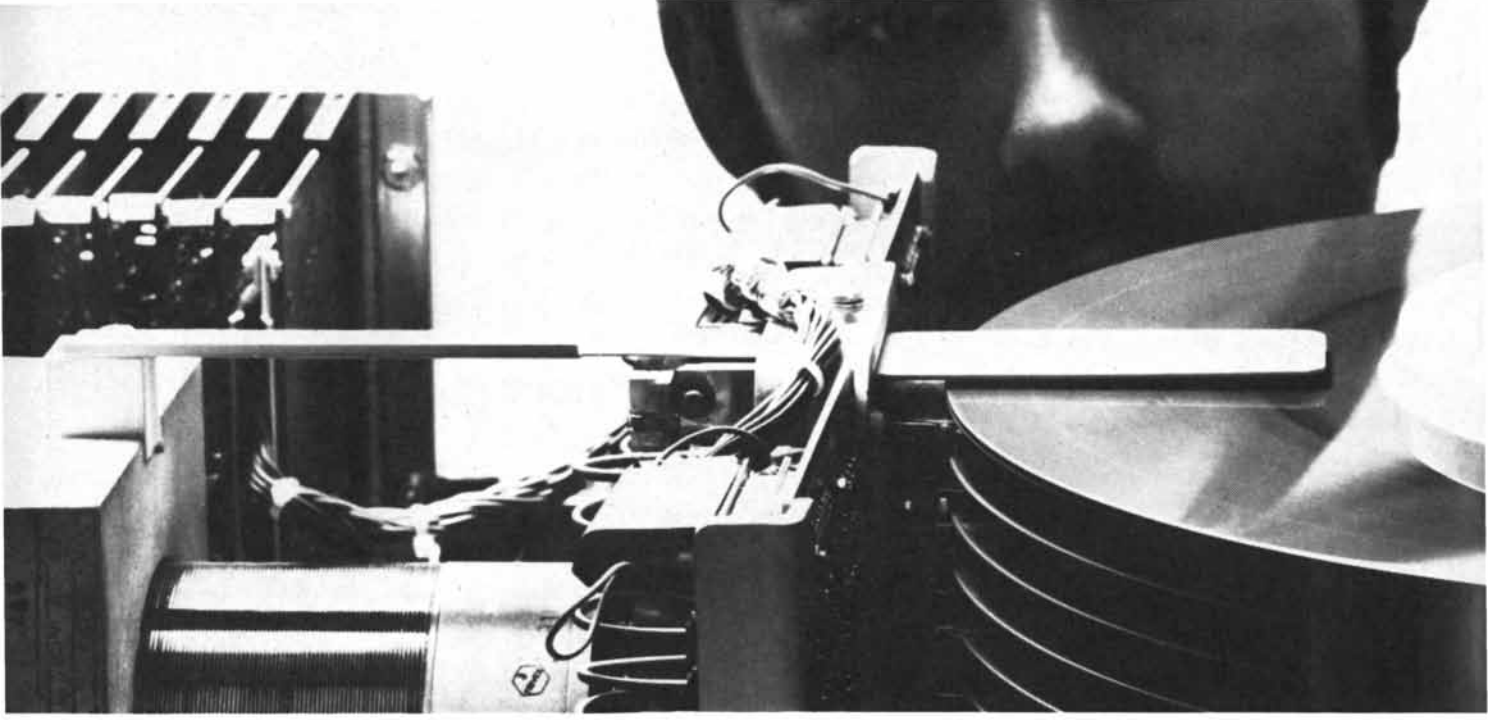
We'll prove it to you.

A lot of people talk about what they can do for you. Let us prove what we say. All it takes is a note to start things moving. We've got choice urban and rural sites waiting for you all over our state. Write today. You'll find Indiana is a lot more than you expected.

Indiana Department of Commerce
Lt. Gov. Richard E. Folz, Director
Room 3322, State House
Indianapolis, Indiana 46204

Indiana
...a lot more than you expected

© 1971 SCIENTIFIC AMERICAN, INC



Look at a CalComp disk drive. It's faster than IBM's.

Now look at the leases.

A CalComp lease assumes that CalComp is going to be improving their equipment all the time.

An IBM lease assumes the same thing. Except, they won't write you a 3 or 4-year lease.

We will.

If you want to add equipment we'll do it without lengthening your lease. Or as your performance requirements grow, we'll re-write the lease with no penalty.

IBM won't.

And CalComp is faster than IBM. The start and stop time is faster. And the access time is faster. So you'd get more for your money, even if CalComp equipment cost as much as IBM.

Except, it doesn't.

CalComp is less to lease. And less to own.

So look at a CalComp disk drive, then look at IBM's. Then compare leases.

Then write us at California Computer Products, Dept. SA-M1-72 2411 West La Palma Avenue, Anaheim, California 92801. Or call (714) 821-2011.

CALCOMP

© 1971 SCIENTIFIC AMERICAN, INC

Established 1845 SCIENTIFIC AMERICAN January 1972 Volume 226 Number 1

Extending the Nuclear-Test Ban

Recent advances in seismological techniques for distinguishing underground nuclear explosions from earthquakes have spurred a renewal of interest in negotiating a comprehensive test ban

by Henry R. Myers

Treaties to ban the testing of nuclear weapons have been a subject of widespread interest since the middle 1950's. The history of the matter is marked by turgid international negotiations, conferences of scientific experts, acrimonious domestic debates, changes in position by one side or the other, a three-year informal moratorium on testing suddenly terminated by a series of Russian tests in the atmosphere, "breakthroughs" in the technology of detecting clandestine tests, supposed "breakthroughs" in the technology of conducting clandestine tests, proposals that were sincere and proposals that were not so sincere, a limited treaty banning all but underground tests, accusations of minor violations of that treaty and—most recently—renewed talk of extending the limited treaty to cover underground tests, stimulated primarily by reports of significant new advances in seismic techniques for distinguishing underground explosions from earthquakes.

Advocates of a test ban have over the years pointed out that such a treaty would be a major step toward terminating the strategic-arms race, that it would bring an end to radioactive pollution by nuclear explosions, that it would make the proliferation of nuclear weapons less likely and that it would increase barriers against the use of nuclear weapons in warfare. Those opposed to a test ban have tended to argue that a treaty would address the symptoms rather than the disease, and that whatever advantages there might be were outweighed by certain military and political disadvantages.

Throughout this meandering history there has been little consensus on answers to a series of related questions: How important are the advantages of treaties that limit the testing of nuclear weapons? Is it really a good idea to accept such limitations even if there could be confidence that other countries had done the same? How much assurance could there be that others would live up to the agreements? What developments might stem from clandestine bomb-testing by others?

Questions related to the problem of detecting violations have dominated the test-ban dialogue out of all proportion to their significance. The reasons for this anomaly are interesting in themselves. One is that the verification question is the one aspect of an otherwise complex issue that appears amenable to quantitative analysis. (Actually it is much less so than appearances suggest.) A second reason is that there seems to be a widely held obsession with the *possibility* of violations rather than with their probability or their significance. A third is that opponents of limitations on nuclear testing have exploited this obsession by encouraging fears that have little basis in fact.

Following the established tradition, much of the present discussion will be concerned with the problem of detecting violations. This is done, however, not to emphasize that aspect of the matter but to place it in proper perspective, so that the political and military implications of continued nuclear testing or a complete cessation of it may receive the greater attention they deserve.

Some people have feared that the U.S. Government, in its eagerness to make arms-control agreements, has not been concerned enough about its ability to detect violations. Such fears have never been well founded. An often-stated precondition for the U.S. to enter into a treaty limiting underground nuclear testing has been that there be adequate assurance that the national security would not be placed in jeopardy by developments growing out of clandestine testing by other countries. Indeed, the failure to arrive at a ban on underground tests has in part reflected the inability to negotiate a treaty that provided the U.S. with the assurance of compliance with the terms of the treaty that many officials and citizens believe is necessary.

Such assurance can be derived from several sources: (1) seismological records, intelligence observations or on-site inspections indicating that particular events were earthquakes and not explosions; (2) awareness that any significant clandestine program would involve a large enough number of tests so that eventually a mistake would be made that would reveal the prohibited activity; (3) knowledge of the relative insensitivity of the balance of strategic power to further developments in nuclear-weapons technology; (4) the belief that a po-

© 1971 SCIENTIFIC AMERICAN, INC

tential violator would probably find an excuse for withdrawing from a treaty rather than run the risks inherent in violating it.

On the last point Herbert Scoville, Jr., a former deputy director of the Central Intelligence Agency and assistant director of the Arms Control and Disarmament Agency, has suggested that consideration be given to the experience of the 1958–1961 moratorium on all nuclear testing. At that time the ability to acquire information concerning violations was much inferior to what it is today. In addition, nuclear-weapons technology was less advanced, and there was more incentive to cheat. Yet even then the U.S. was willing to rely for assurance on its own verification systems. The moratorium ended not because of mutual suspicions about cheating (although these did exist) but because of the initiation by the U.S.S.R. of a massive atmospheric-testing program.

In discussing the various interacting factors that influence assurance I shall begin with the area in which the significant new developments have emerged, namely seismology.

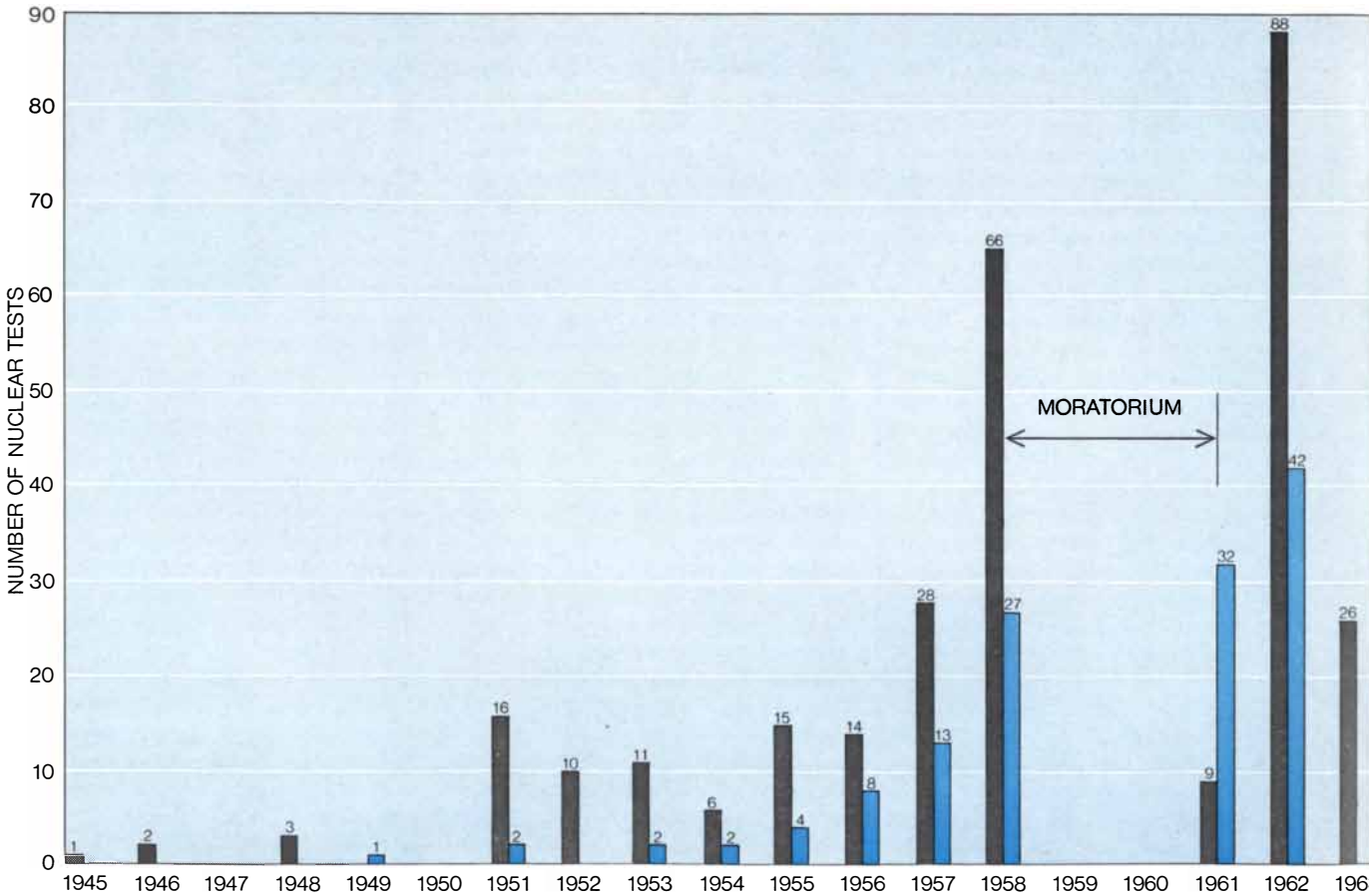
Both earthquakes and explosions cause elastic vibrations in the ground that propagate as seismic waves through the earth (body waves) and along its surface (surface waves). Seismometers are used to detect such waves, and the data obtained from a number of instruments, separated by hundreds or thousands of miles, can be used to determine where the event took place, to define its seismic magnitude and to learn something about the nature of the source mechanism.

Seismic magnitude is an approximate indication of the energy released by an earthquake or an explosion. It is a logarithmic quantity based on the average of measurements of the amplitude of certain portions of seismic waves. For some purposes magnitude is based on measurements of the body wave; in other circumstances it is based on the surface wave.

The largest earthquakes have a body-wave magnitude slightly in excess of 7 and a surface-wave magnitude slightly in excess of 8.5. On the average there are one or two such earthquakes a year. Underground explosions in the megaton range can have a body-wave magnitude of 6.5 to 7. Explosions in the low-kiloton range typically have a body-wave magnitude of 4. The seismic magnitude associated with underground explosions depends not only on the explosive yield but also on the material in which the detonation is set off [*see top illustration on page 20*].

Determining the location of seismic events eliminates the large majority of earthquakes from the "probably suspicious" category. For example, unless there are other grounds for suspicion, it is possible to immediately eliminate events under the ocean floor or in countries where there is no reason to believe a nuclear test might have been conducted.

If an event cannot be discarded on the

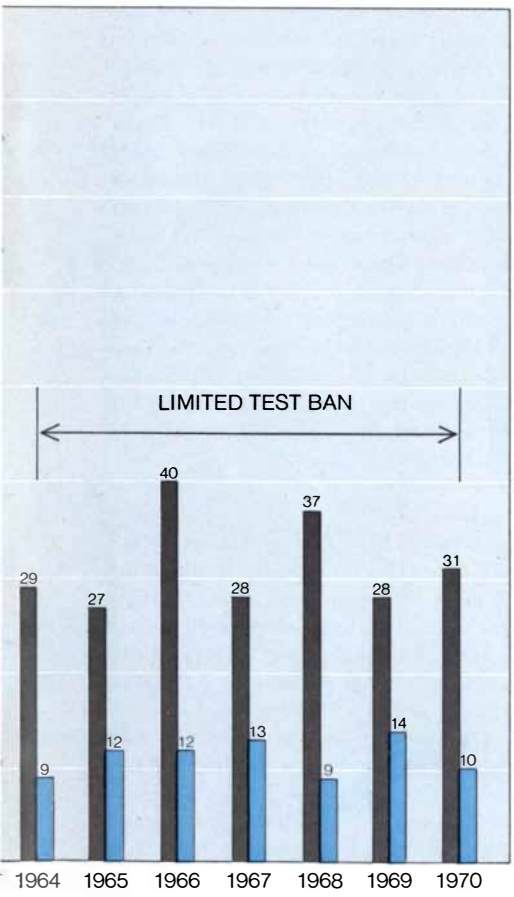


CHRONOLOGY OF NUCLEAR TESTING since 1945 is presented in this bar chart, which indicates the approximate numbers of nuclear-weapons tests conducted by the U.S. (*gray bars*) and the U.S.S.R. (*colored bars*). The estimates were compiled by the Stockholm International Peace Research Institute (SIPRI) from data provided by the U.S. Atomic Energy Commission and the Swedish Research Institute for National Defense. Totals before the limited test-ban treaty went into effect in 1963 include tests conducted in the atmosphere, in space, underwater and underground; these numbers are probably more complete than those available after the 1963 test ban. Although the totals indicated for the latter period may understate the actual rate of underground testing since 1963,

© 1971 SCIENTIFIC AMERICAN, INC

basis of its location, analysts can exploit the fact that in terms of the spectral distribution and the partition of energy among wave modes the seismic signals generated by explosions are quite different from those generated by earthquakes. At distances within 1,000 kilometers of an event, explosions in excess of a few kilotons can easily be differentiated from earthquakes. The effects of the paths along which the waves travel, however, are such that as the waves propagate they increasingly assume a form that is determined more by the characteristics of the path than by the characteristics of the source. Moreover, the seismic wave is attenuated with distance, and diagnostic features contained in it (and for small events the entire signal) can be obscured at the point of detection by seismic noise generated by winds, nearby road traffic or distant ocean waves.

An enduring tenet of the test-ban dialogue has been that the negotiability and viability of a prohibition on underground testing were related to the amount of verification information that could be extracted from the seismological data. More specifically, such a ban was dependent on the ability to understand and reduce the effects of travel paths and seismic noise. As a result the U.S. undertook in 1959 to sponsor a research program intended to lead to improved capabilities for detecting seismic events, for determining their location and for establishing whether the events were earthquakes or explosions. The ultimate objective was to develop monitoring procedures that could provide sufficient assurance in a treaty situation while calling for a minimum of provisions that would be difficult to negotiate, namely on-site inspections or internationally supervised seismological observations inside the U.S.S.R.



it is likely that they give a correct indication of the relative numbers of such tests conducted by the U.S. and the U.S.S.R. According to SIPRI there have been in addition, through 1970, a total of 21 tests conducted by Britain, 46 by France and 11 by China.

Since 1959 more than $300 million has been spent on the U.S. seismological research effort. A lesser effort was mounted by the United Kingdom. As a result there has been a considerable advance in the understanding of earthquake mechanisms and of the structure of the earth. Other significant developments have come in instrumentation, in data-handling and analysis procedures, in the ability to separate signals from noise and in the ability to determine the location of events and to isolate information that might help in determining whether the event was an earthquake or an explosion. On the other hand, the rate at which this progress has been translated into a more negotiable position at the nuclear-test-ban discussions at Geneva has not been as great as might have been hoped. Nevertheless, there have been two notable advances that have led to a substantial improvement in the ability to collect and abstract useful information from the seismological data.

The first major "breakthrough," announced in 1962, amounted to a reduction in the annual number of earthquakes in the U.S.S.R. that might be confused with explosions. Part of the basis for this reduction was a thorough study of earthquake and explosion data that resolved the existing uncertainties about the rate at which earthquakes occur and the relation between earthquake magnitude and explosive yield. In addition, it was determined that the depth of seismic events could be better inferred from the data than had been believed. It became possible to conclude that many more events were too deep to be explosions and therefore were earthquakes. These results were the basis for dropping from the U.S. negotiating position the requirement that there be manned, internationally supervised seismological observatories in the territory of the U.S.S.R.

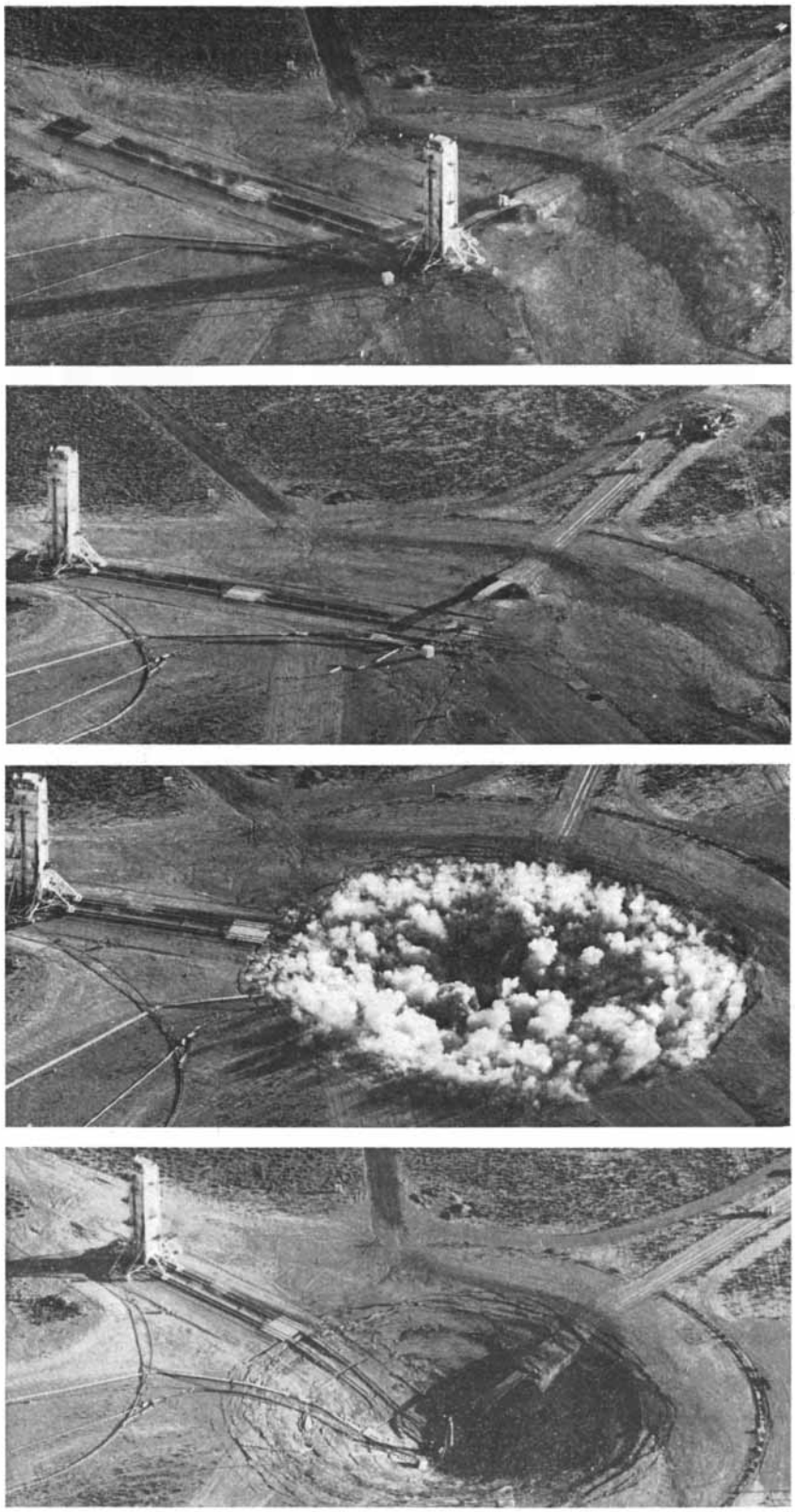
The second such major advance has emerged in the past few years from a series of studies initiated by Jack F. Evernden, a seismologist formerly employed by the Department of Defense and now working at the Arms Control and Disarmament Agency. The most significant result of this work, which has been carried on by groups at several universities, was to verify the hypothesis that for a given total release of seismic energy, explosions were much less efficient at generating long-period surface waves than small, shallow earthquakes were [*see illustration on pages 18 and 19*]. Although there were indications before 1962 that this was a worthwhile line of investigation, the early analyses were based on limited data and were not pursued to the point where results would have had a strong impact on the U.S. position in the test-ban negotiations. Moreover, it had been suggested that as the total seismic energy released became smaller, the differences between explosions and earthquakes would tend to diminish to the point where they would not be observable.

It has now been demonstrated that the difference persists for the smallest events of any interest, and that the average ratio of surface-wave magnitude to body-wave magnitude for explosions is less than the average ratio for earthquakes at least as far down as a body-wave magnitude of 4, and probably lower [*see bottom illustration on page 20*]. For the comparatively small number of events that cannot be eliminated on the basis of location, depth or surface-wave criteria there are other diagnostic methods that can be applied to the data. These criteria include the relative complexity of the seismic wave (earthquakes generally produce more complex signals than explosions do) and the ratio of compressional-wave energy to shear-wave energy (earthquakes usually produce larger shear waves than explosions do).

Other research has demonstrated that through the use of seismometer arrays, which have the effect of enhancing seismic signals with respect to the background noise, it is possible to detect both surface waves and body waves at distances of thousands of miles from events with a body-wave magnitude of 4.

As a result of these studies it is now known that it is possible to build a network of seismological observatories, located many thousands of miles from seis-

© 1971 SCIENTIFIC AMERICAN, INC



UNDERGROUND NUCLEAR EXPLOSION is chronicled in this sequence of photographs made from a helicopter circling above ground zero at the Nevada Test Site of the Los Alamos Scientific Laboratory. The first photograph, made about a minute after the nuclear explosive was detonated, shows dust rising around ground zero caused by the explosion's shock wave. At this point a 100-foot-high experimental tower began to move by rail away from ground zero to a safe position outside the anticipated crater area (*second photograph*). Some 23 minutes later the material above the temporary underground cavity formed by the explosion collapsed (*third photograph*), leaving a subsidence crater (*fourth photograph*).

mic sources of interest, that will acquire data that can be used to identify "virtually all" earthquakes in the U.S.S.R. having a body-wave magnitude in excess of 4. It also means that in the absence of active evasion procedures there would be a substantial probability of detecting and identifying explosions in the U.S.S.R. with yields greater than one or two kilotons. At such yields a nuclear-test program would have little significance. The significance of these conclusions of course depends on the meaning of "virtually all." Some seismologists suggest it means that each year there will be as many as 25 earthquakes in the U.S.S.R. with a body-wave magnitude above 4 that could not be differentiated from explosions on the basis of the seismic data. Others believe that the number will be 10 or less. These estimates are to be compared with the average expected number of unidentified events, ranging from 75 to 150 events above a body-wave magnitude of 4, that would have been anticipated before the surface-wave techniques were developed.

The spread in present and past estimates is partly a reflection of natural fluctuations in the rate of occurrence of earthquakes. Mostly, however, it results from different interpretations of the data and different estimates of the likely outcome of ongoing studies. The data themselves are sometimes subject to diverse interpretations because the great variability in the source mechanisms and the transmission properties of the earth creates a corresponding variability in the observations; therefore questions are sometimes raised concerning the validity of conclusions based on a particular set of measurements. Also contributing to the diversity in estimates is the fact that there are very different views concerning the elaborateness of the seismological network that might exist at some future time. This uncertainty is important because the degree of differentiation that is possible depends directly on the number, distribution and quality of the seismometer stations providing the raw data.

The residual number of events that cannot be identified on the basis of the seismic data would be further reduced by the analysis of data supplied by intelligence systems. In particular it has been hypothesized that satellite photography should be capable of providing information that might alleviate doubts about specific occurrences. For example, a seismic event might be detected and its location determined on the basis of seismological data but without there being enough information to provide confi-

© 1971 SCIENTIFIC AMERICAN, INC

dence that the event was an earthquake. Photography from space might indicate that in the area where the event took place there was no sign of activity of the kind associated with a nuclear test. On the other hand, if the photography were to reveal activity such as mining, it would not serve to decrease suspicion. It is therefore probable that satellite photography would serve much more to reduce suspicions about earthquakes than to provide evidence that a prohibited explosion had occurred.

Nevertheless, a country deciding to undertake clandestine tests would be constrained to do so in a way that would not be susceptible to observation. The possibility that violations might be observed from space would thus have the effect of imposing substantial restrictions on a potential violator. This would tend to deter attempts to carry out secret tests.

On-site inspections have been another omnipresent element in the test-ban dialogue. In the beginning there was a generally held belief that seismological data would be much less useful and inspections more useful than either are now known to be. By supplementing information provided by seismic instruments and intelligence systems, inspections were considered a mechanism that would make a substantial contribution to the verification process. As a result of the recent developments in seismology, inspections would apply only to the very small number of events that are large enough to be detected but that produce signals of insufficient signal-to-noise ratio to allow the application of seismic identification techniques. As usually envisioned, inspections would be conducted after the presentation of ambiguous seismic data and under the auspices of an international body. They would involve visits to the area of origin of seismic events that could not be identified after analysis of the seismic and intelligence data. The inspection area would extend over some several hundred square kilometers, an area reflecting the expected error in determining the point at which the seismic event had occurred.

The procedure for deciding that a seismic event could not be identified, and therefore should be subject to an inspection, is not easily articulated. As a practical matter it would be exceedingly difficult to specify in a treaty, in a manner that would be acceptable to all parties, the data that would need to be provided in substantiation of a claim that an inspection was necessary.

The ability of on-site inspections to uncover evidence of clandestine testing, should there be such testing, would depend on a chain of probabilities: the probability that an explosion would be selected as a candidate for inspection, the probability that the inspection area would encompass the point of detonation and the probability that the inspectors would actually uncover positive evidence should the inspection area include the shot point. The foregoing factors would in turn depend on the quality of

SUBSIDENCE CRATERS caused by underground nuclear explosions pock the desert at the Nevada Test Site. Viewed from space by satellites such craters would be telltale evidence that a test had occurred. Hence a country planning to violate a test-ban agreement would be compelled to detonate its nuclear explosives at sufficient depths to have confidence that there would be no such crater.

© 1971 SCIENTIFIC AMERICAN, INC

the evidence on which the inspection decision was based, the number of inspections that might be conducted in a given time period, the accuracy of locating the source point of seismic events, the size of the area covered by an inspection, the duration of the inspection, the number and training of the inspectors and the elaborateness of the inspection procedures.

Since a country that had violated the treaty would be unlikely to allow access to inspectors who might uncover the clandestine activity, inspections would probably be undertaken only in areas where earthquakes had actually occurred. Because the seismic data would be adequate for identifying large earthquakes, inspections would be directed toward smaller events that typically would produce no ground effects that might be found by inspectors. Moreover, Defense Department experts now say their research has indicated that careful planning and execution of a clandestine test would preclude the existence of surface manifestations that might be found by inspectors. Therefore on-site inspections could not be expected to turn up any evidence of either explosions or earthquakes. Finally, if an evader should be successful in employing active evasion techiques, there would be no seismic evidence on which a request for an inspection might be based.

If this is the case, do inspections serve any purpose even in theory? Yes, according to some. The lack of evidence of an explosion would supposedly provide a limited amount of assurance (skeptics say false assurance) that there had been no explosion. In addition, the prospect of an on-site inspection and the possibility that one might miscalculate or make a mistake would be at least a slight deterrent to a potential violator.

Whatever the theoretical utility of inspections, it is questionable that a treaty that provided for meaningful inspections was ever a realistic possibility. The Russians have always had a negative attitude toward inspections on the grounds that they are not needed and would be used as a pretext for espionage. During periods when the Russians did seem willing to consider the possibility of some small number of inspections (allegedly for the purpose of diminishing some American domestic opposition to a treaty), it was not clear that the nature of these inspections was even close to the elaborate procedures envisioned by the U.S.

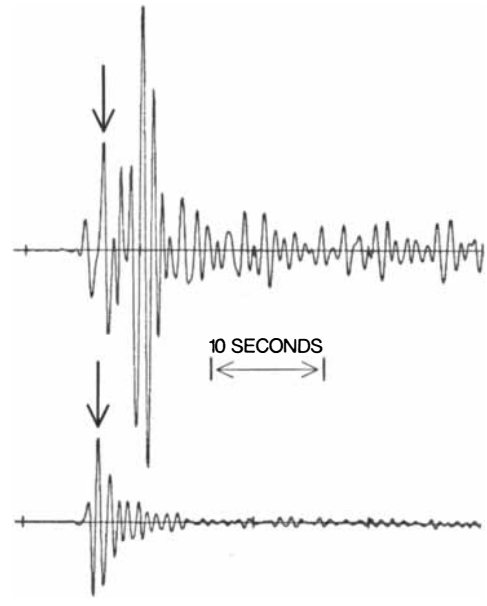
With the recent advances in seismological research, doubts concerning the usefulness of inspections have been intensified. To continue to maintain that a ban on underground tests is important but that inspections are still necessary implies the acceptance of the following dubious propositions: (1) that significant tests might escape the various monitoring systems operated by the U.S., (2) that inspections would play a substantial role in deterring violations or in alleviating suspicions that might arise and (3) that the Russians would agree to inspection procedures of the kind that would be required if they were to serve their purpose.

To put the matter in perspective it should be noted that new ground would not be broken if the demand for inspections were dropped because they were deemed unnecessary. The U.S. is already a party to a series of arms-control agreements that contain either no provision for inspections or provision only for inspections of a kind that would yield virtually no assurance in addition to the assurance acquired by other means. These agreements include the ban on nuclear explosions in the atmosphere or outer space, the ban on placing nuclear weapons in orbit around the earth, on the moon or on other celestial bodies, the ban on placing weapons of mass destruction on the ocean bottom and the ban on nuclear weapons in the Antarctic. In each case unilateral means are considered adequate to provide the necessary confidence that the impact of possible undetected violations would be so minuscule as to make them not worth worrying about.

In effect the data provided by seismological and intelligence systems and, to a much lesser extent, by on-site inspections would serve to reduce the background "noise" against which the "signal," or positive evidence, of clandestine test activity would be detected. This "noise" is a factor that a potential violator would take into account in making a decision about conducting a clandestine test. Would he assume that his activity would be lost in the "noise," that it would be discounted as a spurious earthquake in the residuum of unidentified seismic events? Or would he believe that he should play a more active role in attempting to conceal his tests?

Since a complete test ban became a serious possibility several such active schemes for clandestine evasions have been proposed. These include the muffling of the explosion so that the resulting seismic signals would be unlikely to be detected, modification of the character of the seismic waves so that even though they might be detected they would not be characteristic of waves associated with explosions, and testing in the aftermath of a large earthquake so that the explosion signals would be concealed by the ground motion caused by the earthquake.

The muffling of seismic signals might be achieved by conducting the test in an underground material where a comparatively small fraction of the energy of the explosion would appear in a propagating seismic wave. For example, a 20-kiloton explosion in a dry desert alluvial deposit might produce a seismic signal as small as the one produced by a one-kiloton explosion in hard rock. If a country were to seek to test in such an alluvial deposit, it would first be confronted by the fact that a deposit deep enough and dry enough is very uncommon. There is, in fact, reason to believe these deposits do not exist in the U.S.S.R. in depths that would allow tests of more than one or two kilotons. Since dry alluvial deposits, or other dry materials that might be useful for the same purpose, are usually found in areas that are relatively free of clouds and lacking in human activity, any attempt to conduct clandestine tests in these circumstances would be particularly vulnerable to observation from space. Further restrictions arise from the fact that the alluvium must be deep enough not only to contain the immediate effects of the explosion but also to make it possible to set off the explosion at a depth that



SEISMIC RECORDS produced by an earthquake (*top traces*) and an underground nuclear explosion (*bottom traces*) each located at approximately the same distance (80 degrees) and the same azimuth (350 degrees)

© 1971 SCIENTIFIC AMERICAN, INC

would not leave a collapse crater. This is an indirect effect in which the cavity produced by the nuclear explosion fills with material from above. The result is a propagating void that ultimately appears at the surface as a characteristic subsidence crater. Such craters are common at the U.S. testing ground in Nevada [*see illustration on page 17*]. If such a crater were to be observed from space, it would be telltale evidence of a nuclear explosion.

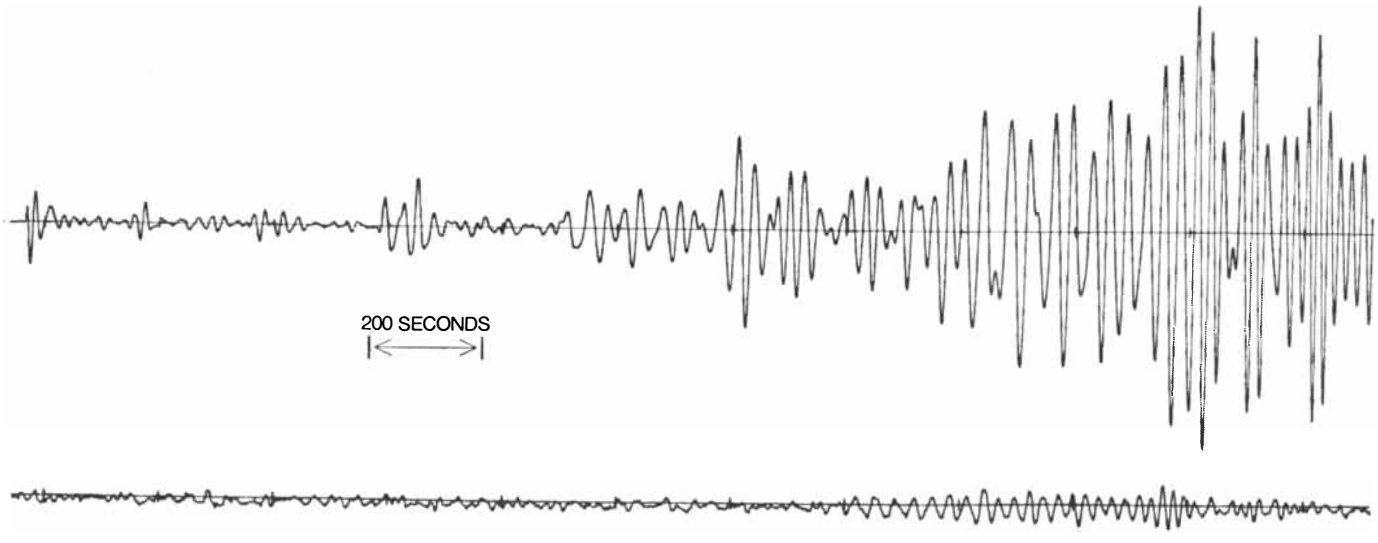
A more elaborate muffling technique calls for detonating the nuclear device in a spherical cavity. It is predicted that with a cavity of the right size the seismic signal would be about 100 times weaker than it would be without the cavity. The diameter of the required cavity increases with the cube root of the explosive yield. A 10-kiloton explosion would call for a cavity 300 feet in diameter 3,000 feet below the surface. It is questionable whether the cost of excavating such a cavity would be commensurate with the gains that might be made from a clandestine nuclear test conducted in it.

For a time it seemed that such muffling, or decoupling, schemes had fallen out of the domain of plausible evasion possibilities. In recent Congressional testimony, however, decoupling was resurrected by Defense Department experts, who implied that decoupled detonations with yields of 50 to 100 kilotons were feasible and could be carried out at costs that were within reasonable bounds. The assertion raised the question of why, if the possibility of decoupling was considered to be significant enough to bring up, there had been only one small experiment to check out the predictions. This test had involved a 300-ton shot in a cavity created by a 5,000-ton nuclear explosion in a salt deposit in Mississippi. Because the observed decoupling was less by a factor of two than had been forecast, and because there are serious questions concerning the validity of extrapolating the results of this experiment from a yield of 300 tons to one of 50,000 or 100,000 tons, it would seem important to have conducted experiments with much larger explosions. The willingness to extrapolate results here (where the extrapolated conclusion is consistent with the desires of many Defense Department experts to avoid a complete ban on nuclear-weapons tests) conflicts with the Defense Department's usual position on nuclear-weapons development. For example, the five-megaton Cannikin test conducted on Amchitka in November was supposed to be necessary because it was not possible to base conclusions on data obtained from a one-megaton explosion.

At various times it has been suggested that a series of nuclear detonations, appropriately located and timed, would produce a composite seismic signal that was more typical of earthquakes than explosions. Although this procedure might generate waves that in some respects were similar to waves made by earthquakes, it would be difficult for a treaty violator to be confident that the similarity was sufficient to avoid creating suspicion.

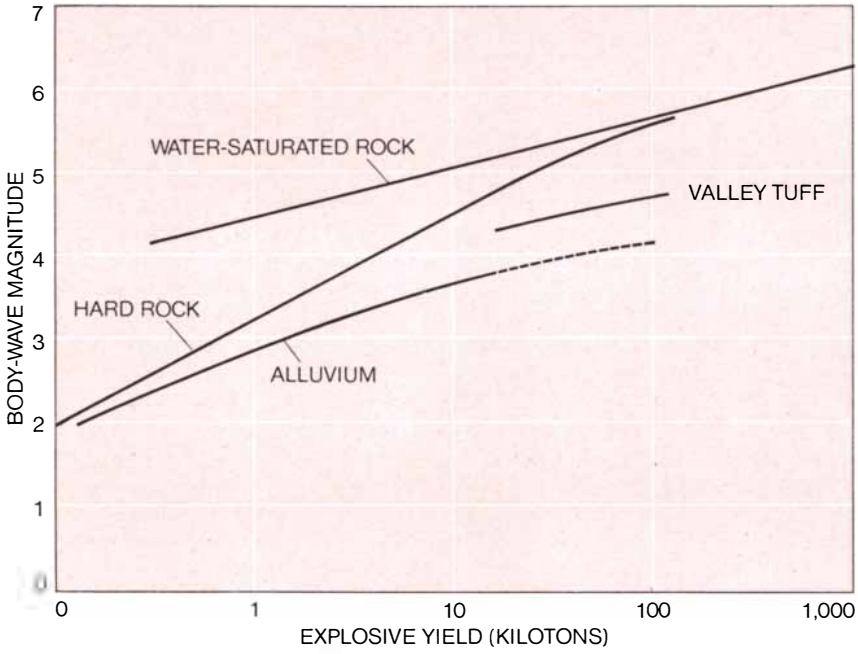
A violator might also try to conceal illicit test activity by setting off an explosion almost immediately after an earthquake. If this were done properly, the seismic signals associated with the explosion would be masked by those generated by the earthquake. One disadvantage of testing in this fashion would be the need to maintain the nuclear device and the apparatus associated with the experiment in a constant state of readiness for a period that would range from months to years, depending on the size of the test detonation. It would also be necessary to determine rather quickly that the earthquake was suitable for purposes of masking. That would require an interconnected network of seismometers from which data would be fed to a computer. The computer would immediately calculate the location and size of the earthquake, and it would give a "go" signal when the earthquake met the preset conditions. As with other imagined evasion techniques, the practicality of this method would depend on how badly a country wanted to test, on the size of the test and on how the country evaluated its chances of being caught.

The usefulness of on-site inspections in deterring such efforts would be con-

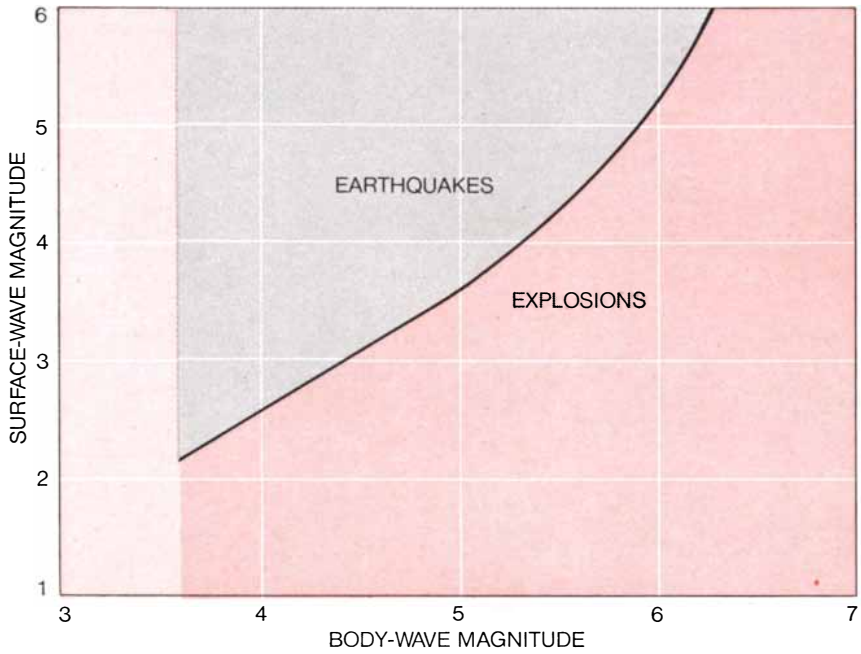


from the detection point are compared here. The two seismograms at left were produced by body waves: shock waves that traveled through the earth from each event to the seismic station. The two traces at right were produced by surface waves: shock waves that traveled along the surface of the earth. The body-wave magnitude of each event is determined from a measurement of the maximum amplitude in the first five cycles of the wave (*arrows*). In this case the body-wave magnitudes are approximately equal for the two events. As expected, however, the magnitude of the surface wave is much smaller for the explosion than it is for the earthquake.

© 1971 SCIENTIFIC AMERICAN, INC



SEISMIC MAGNITUDE recorded for a particular underground explosion depends not only on the explosive yield but also on the material in which the explosion takes place. These curves illustrating that relationship are based primarily on measurements made in the eastern U.S. of explosions in Nevada. For explosions in excess of 20 kilotons the alluvial deposits at the Nevada Test Site are not deep enough to contain the immediate effects of the explosion. Hence for larger explosions it is necessary to drill beyond the alluvium into the valley tuff underlying it. This accounts for the upward shift of the right end of the alluvium curve. The fluctuations about these curves are such that any body-wave magnitude inferred from them is uncertain by .3 unit and any explosive yield is uncertain by a factor of 2.



DIFFERENTIATION of underground explosions and shallow earthquakes by means of the ratio of their body-wave magnitude to their surface-wave magnitude appears to be feasible down to a body-wave magnitude of approximately 4. In this graph of that ratio, inferred from a large quantity of seismic data, virtually all explosions lie below the curve, whereas most shallow earthquakes lie above the curve. (Deep earthquakes can be identified on the basis of other properties of the seismic data.) It is generally agreed that a small number of shallow earthquakes will be within the explosion zone. Some seismologists believe there will be each year as many as 25 such events with a body-wave magnitude greater than 4 in the U.S.S.R.; others believe the number of such ambiguous seismic events will be 10 or less.

tingent on procedures that had been established for submitting requests to conduct an inspection. If it were necessary to base the request on specific kinds of seismic data indicating that an unidentified event had occurred, there would be much less deterrent value than if it were possible to base the request on unrevealed intelligence information with only scant reference to publicly available seismic data.

Scoville notes that an easy, although seldom discussed, way for a country to cheat a weapons-test ban would be for it to mount a program of setting off nuclear devices for peaceful purposes, such as the U.S. Plowshare program. One part of Plowshare is the development of nuclear explosive devices for use in nonmilitary projects. Since these devices are in many ways identical with weapons, it would be virtually impossible to determine the kind of development activity being undertaken without being given access to the experimental apparatus. Such access is not likely to be granted by either the Russians or ourselves. Therefore if there is going to be a complete ban on nuclear-weapons tests, it will be necessary either to live with the possibility of some weapons development under the shelter of Plowshare-like programs or to forgo the yet to be established economic benefits of this type of activity.

A further consideration affecting a country's assurance that it could detect clandestine tests (or evade detection) is the possibility that there might be a mistake or an accident in such a test. An unexpected collapse crater or the venting of radioactive debris could lead to the discovery of prohibited activities. There is a significant chance of unintentional venting and a significant probability that such venting, if it occurs, will be detected. In the past radioactive debris has been unintentionally emitted from underground explosions conducted by both the U.S. and the U.S.S.R. The possibility of such accidents would tend to deter violations.

A useful device for summarizing the concepts that enter into evaluations of the likelihood of detecting clandestine test activity or of escaping detection is represented by the graph on the opposite page. The abscissa represents the product of the annual number of tests and their average yield; for instance, a value of 10 could mean 10 tests with an average yield of one kiloton or two tests with an average yield of five kilotons. The ordinate is a measure of the confidence that a test or series of tests with

© 1971 SCIENTIFIC AMERICAN, INC

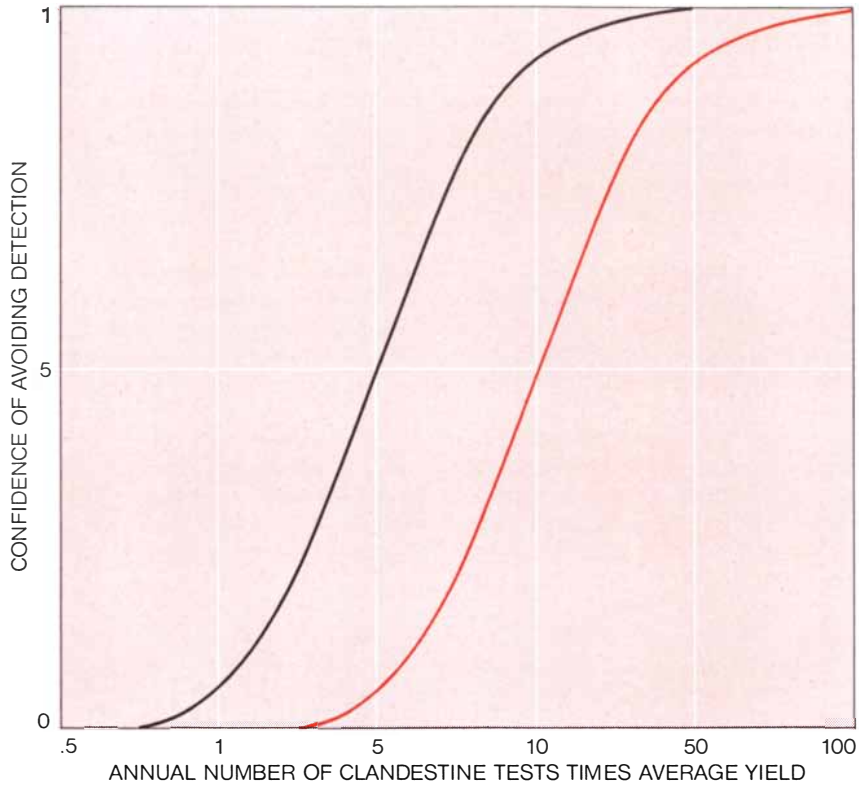
the specified yield-number product would be detected.

The "deterrence" curve represents the situation as it might be viewed by a country undertaking a clandestine test program. The "assurance" curve represents the situation as it might be perceived by countries concerned about their ability to detect violations by others. The deterrence curve invariably lies to the left of the assurance curve because the country concerned about violations by others will, in comparison with the evaluations of the potential violator who presumably places a high premium on not being caught, tend to underestimate the chances of a violation's being detected.

For example, the great variability in seismic data may cause a violator to "play it safe." He would assume, as actually happens, that his explosions would produce at some observation points seismic signals substantially greater than the signals suggested by the averages implied by the curves in the top illustration on the opposite page. The violator would have to conservatively evaluate his chances of being observed from space or having an accident that resulted in the venting of radioactive debris. On the other hand, countries seeking assurance might well assume that explosions would produce seismic signals of an average size implied by the average curve, that the violator would make sure that there was no substantial deviation from the average, that there would be no basis for counting on anomalously large signals to reveal a clandestine test, and that the test would be conducted in a manner so as not to produce a collapse crater or other evidence that might be collected by intelligence systems.

Such curves or their conceptual equivalent are an aid in evaluating the risks associated with a nuclear-test ban even though they are necessarily imprecise because of the large number of intangibles and uncertainties they encompass. The inability of anyone to develop more quantitative estimates has caused consternation among presidents, members of Congress and interested citizens. Adding to the uncertainty is the fact that the detection of clandestine tests must be considered an integral part of the broader question of how significant undetected violations would actually be.

It is fortunate that the recent advances in verification technology have simplified the issue. Many informed observers now conclude that the U.S. need no longer be concerned about developments that might result from *clandestine* nuclear-test activities, since such developments would not affect the strategic position of the U.S. with respect to its adversaries. It is thus possible to focus on the role of *any* nuclear tests at the present juncture of the arms-development cycle.



DETERRENCE AND ASSURANCE CURVES indicate the degree of confidence either party to an underground-test-ban agreement might have that at least one of a certain annual number of clandestine tests with a certain average explosive yield would be detected. The deterrence curve (*black*) represents the situation as viewed by a potential violator of the agreement; the assurance curve (*color*) represents the situation as viewed by a country concerned about detecting the violations of others. The deterrence curve lies to the left of the assurance curve because the country concerned about the violations of others will, as compared with the evaluations of a potential violator, tend to underestimate its chances of detecting violations. Although the locations of such curves are necessarily imprecise, they do serve as a useful device for evaluating the risks associated with a nuclear-test ban.

In discussing nuclear testing it is necessary to consider global questions, including what the intentions of adversaries are, how nuclear testing would influence the probability of a nuclear war's starting and what effect nuclear testing would have on the outcome of the war if it should start. It is clear that such complex questions cannot be answered conclusively, but certain facts, opinions and speculations can be stated, and inferences can be drawn from them.

The public record contains very few details of the U.S. nuclear-weapons program. Information about what goes on in the U.S.S.R. is even scantier. Nevertheless, it is possible to get an idea of both the objectives and the costs of U.S. activities from official testimony before committees of Congress and from committee reports based on that testimony. From these sources it appears that during each of the past six years the U.S. has spent approximately $450 million on the development and testing of nuclear weapons. The objectives are to develop weapons that have greater explosive yield for a fixed weight, to develop weapons in which certain nuclear effects are either increased or decreased with respect to other effects, to verify that the performance of existing weapons does not decline with the passage of time and to gain confidence in safety against accidental nuclear detonations. Some nuclear explosions are used as sources of radiation and blast in experiments to determine their effects on weapons systems and their components.

Nuclear detonations for these purposes can be justified to some extent on the grounds that in the long run they will make it possible to save money. For example, it may be that a missile warhead with a certain explosive power and

© 1971 SCIENTIFIC AMERICAN, INC

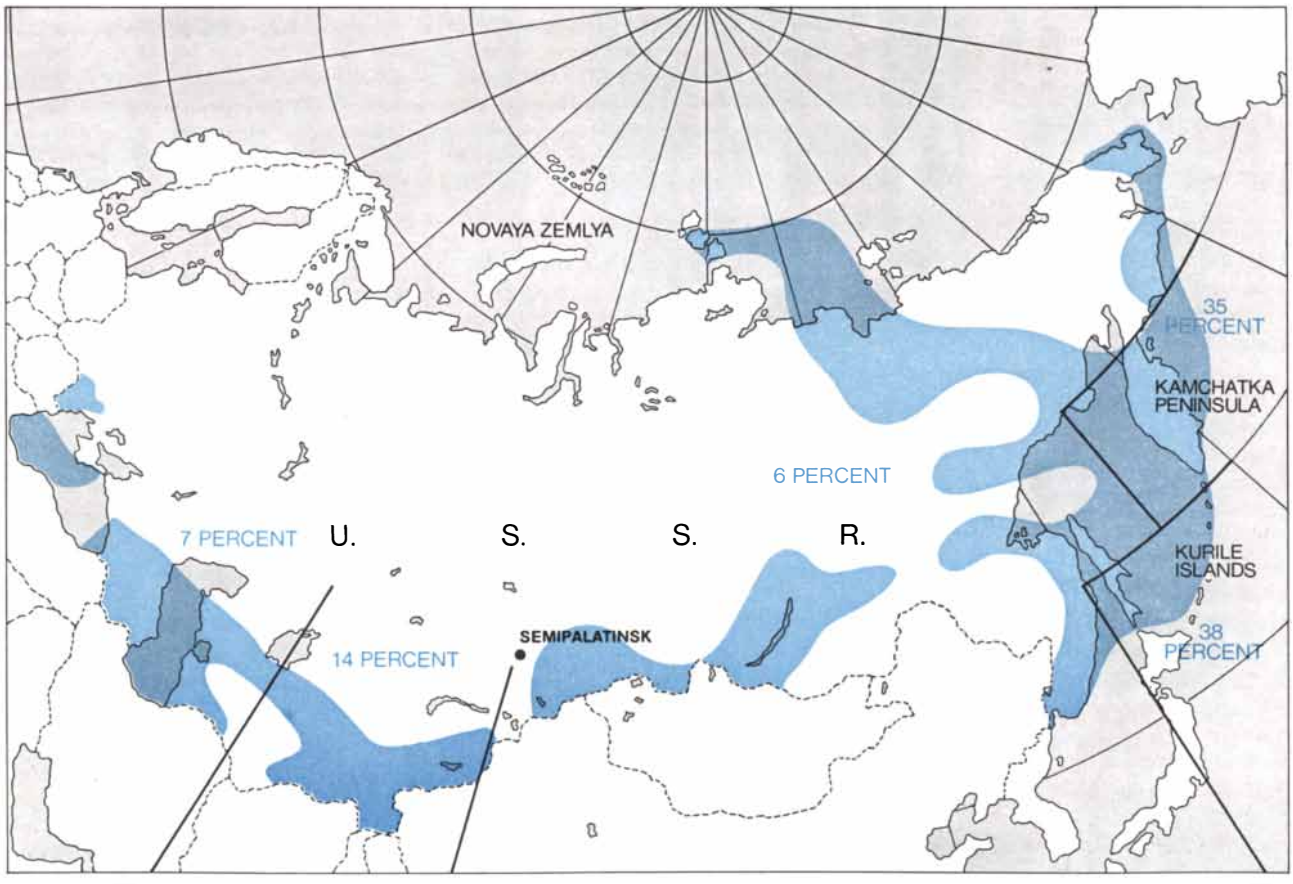
a specified weight could be developed only through nuclear testing. If tests were not allowed, the same destructive power would have to be put into, say, three missiles instead of two, thus preventing a saving in the cost of producing a given amount of damage.

Similarly, there could be uncertainties about the performance of stockpiled weapons or the effect of nuclear explosions on military hardware that could be resolved only by subjecting the hardware to the effects of a nuclear explosion. If such uncertainties could not be eliminated by nuclear tests, it would be necessary to eliminate them through the more costly route of conservative design. In "National Security and the Nuclear-Test Ban," an article that appeared in the October 1964 issue of *Scientific American*, Jerome B. Wiesner and Herbert F. York addressed the uncertainty question at some length. They pointed out that whatever uncertainties might be resolved by nuclear testing would be very small compared with those that abound in military planning. The fact that some of these uncertainties might be removed by nuclear tests would, in their view, "contribute virtually nothing more to management of the real military and political problems, even though it would produce neater graphs."

Some have argued that nuclear testing must continue because of the possibility of surprise findings: either findings that lead to important new developments or, what is more likely, findings that expose a problem where none had been expected. Others argue that tests are necessary to help ensure that the Russians do not acquire knowledge not in the possession of American experts. It is of course not possible to demonstrate that surprise results will not be obtained. Nor is it possible to prove that the Russians do not know things of which American experts are unaware. As for the likelihood of there being a surprise of the kind that suggests a new line of weapons, one may take note of York's belief that the rapid technological advances of recent decades are neither typical of the past nor predictive of the future.

It is sometimes suggested that the U.S. could not forgo testing as long as China or France was continuing its tests. This contention has little validity. For the indefinite future whatever dangers might arise from the fact of Chinese or French nuclear-weapons developments are most unlikely to be ameliorated by the products of U.S. tests. Such dangers must be countered by other means.

More generally, the need for continued testing should be considered in the light of the fact that in this 27th year of the nuclear age some thousands of nuclear weapons of proved design are fully deployed. No foreseeable weapons developments based on further nuclear tests, conducted either openly in the absence of a complete test ban or clandestinely in the presence of one, would affect the ability of either the U.S. or the U.S.S.R. to deliver a substantial portion of its strategic stockpile to an opponent's cities. McGeorge Bundy has remarked on the implications of even a small fraction of these weapons being detonated over urban targets: "In the real world of real political leaders—whether here or in the Soviet Union—a decision that would bring even one hydrogen bomb on one city of one's own country would be recognized in advance as a cata-



SEISMIC MAP OF U.S.S.R. indicates the principal regions of seismic activity (*colored areas*) and the approximate percentage of the total number of earthquakes that occur in each area (*colored type*). As far as is known, underground nuclear-weapons tests have been conducted by the U.S.S.R. on the arctic island of Novaya Zemlya and near the city of Semipalatinsk in south-central Siberia.

© 1971 SCIENTIFIC AMERICAN, INC

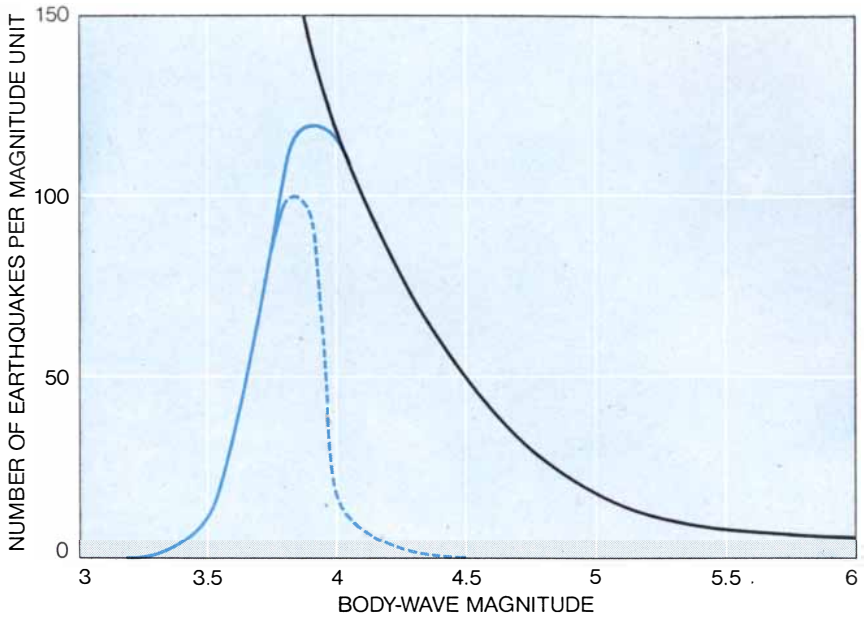
strophic blunder; 10 bombs on 10 cities would be a disaster beyond history, and 100 bombs on 100 cities unthinkable."

In brief, the advanced status of nuclear-weapons development, the existence of very large weapons stockpiles and the deterrent value of the ability to destroy even one city suggest that neither the continuation of nuclear-weapons tests nor the cessation of such tests will affect the strategic position of one superpower with respect to the other. If this is the case, it should mean that continued testing need not necessarily be a source of international tension and a motive force in the arms race.

Yet in spite of the decline of the substantive importance of nuclear tests, testing has developed a life of its own and, as was evident in the Cannikin episode, does exacerbate international discord. If only because there are such strong feelings concerning the need for either an indefinite continuation of testing or its elimination, the matter retains a large symbolic importance. Accordingly there is much to be said for a treaty that removes nuclear testing as an issue that obscures the real problem: how to avoid nuclear war.

The ban on atmospheric testing grew out of the fears generated by the Cuban missile crisis of 1962 and the increasing radiation levels caused by fallout. The superficial reason the treaty did not encompass underground tests was the inability of the signatories to reach agreement on verification procedures. Underlying the omission of such tests was the opposition of persons in and out of the U.S. Government who believed it was essential that this country continue its test activities. At the same time the importance of a prohibition on underground tests was recognized in the preamble of the Limited Test-Ban Treaty of 1963 and later in the Nonproliferation Treaty of 1970, both of which incorporated a commitment to continue negotiations toward a ban on all nuclear-weapons tests.

Not only has little been done to fulfill this commitment but also it is debatable that the atmospheric ban has had much effect on the development of nuclear weapons. In order to secure the endorsement of the U.S. Joint Chiefs of Staff for the Limited Test-Ban Treaty the Kennedy Administration agreed to implement a program of safeguards that included "the conduct of comprehensive, aggressive and continued nuclear-test programs designed to add to our knowledge and improve our weapons in all areas of significance to our military posture for the future." This was to counter the Joint Chiefs' concern that the treaty might presage a "euphoria in the West which will eventually reduce our vigilance and willingness of our country and of our allies to expend continued effort on our collective security." The result of these safeguards has been what many believe is a nuclear-weapons development effort that is indistinguishable in scope—if it is not even larger—from the one that had been under way before the treaty.



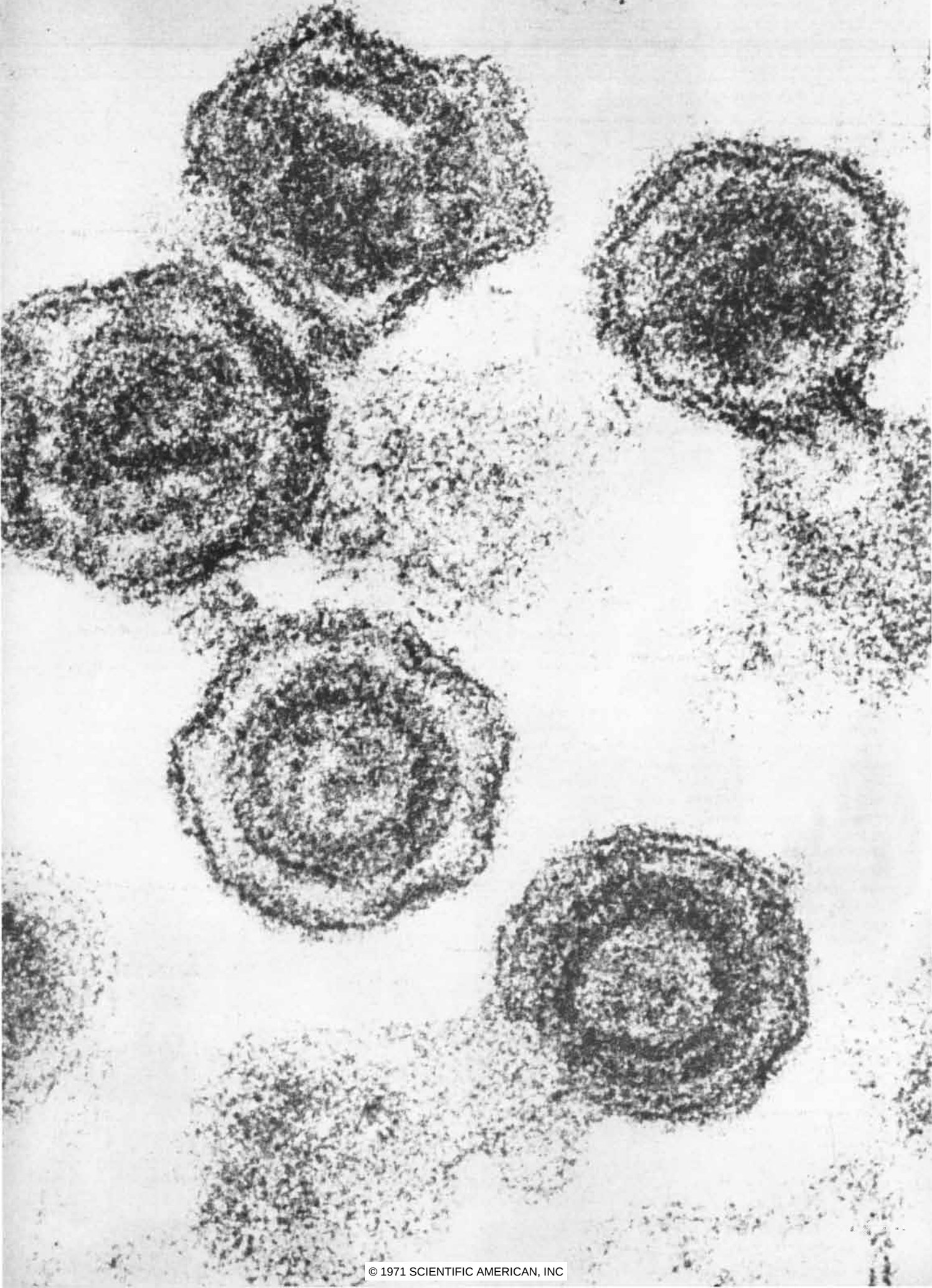
AVERAGE NUMBER OF SHALLOW EARTHQUAKES in the U.S.S.R. per magnitude unit per year is indicated approximately by the black curve in this graph. The estimated average annual number of shallow earthquakes per magnitude unit that would be detected by an appropriate network of seismological observatories located outside the U.S.S.R. is indicated by the solid colored curve; the estimated average number of shallow earthquakes per magnitude unit that would be detected but not identified by such a seismic network is indicated by the broken colored curve. The solid colored curve diverges from the black curve at a point where seismic signals become comparable in amplitude to irreducible seismic noise. If noise-suppression techniques should be further improved, the divergence of these two curves would occur at a lower magnitude. Improvements in identification techniques would have the effect of moving the broken colored curve to the left and reducing it in amplitude. The quantities represented by these curves fluctuate about the average by substantial amounts. Moreover, the curves are dependent on certain "ground rules" for considering earthquakes to be of "shallow" depth and for including certain seismic events in the waters off the Kurile Islands and the Kamchatka coast (*see map on opposite page*).

Since 1963 there has been little serious discussion of extending the Limited Test-Ban Treaty to include underground tests. The ostensible obstacle remains: the U.S.S.R. continues to maintain that on-site inspections are not needed; the U.S. asserts the contrary. The verification issue is still a pretext for not achieving agreement. As the foregoing discussion suggests, however, the plausibility of the pretext has dwindled to the vanishing point.

The failure to prohibit underground tests plays into the hands of those in the non-nuclear nations who press for the acquisition of nuclear forces; it tends to reinforce the contention that the statements of the superpowers are not matched by actions, and it undermines the credibility of the U.S. Government, which has maintained for more than a decade that unresolved verification problems are the obstacle to a ban on underground tests.

The attainment of a treaty banning underground tests would demonstrate to the non-nuclear countries that the major nuclear powers would accept substantial restrictions on their own nuclear activities, a demonstration that would strengthen arguments against the acquisition of nuclear weapons by other countries. As a result of its symbolic value, a cessation of underground testing, particularly in the absence of a major agreement at the Strategic Arms Limitation Talks, would more than any other likely step signify an almost irrevocable commitment to seek security through arms-control agreements rather than through the never ending cycle of weapons and counterweapons that has characterized the period since World War II.

© 1971 SCIENTIFIC AMERICAN, INC



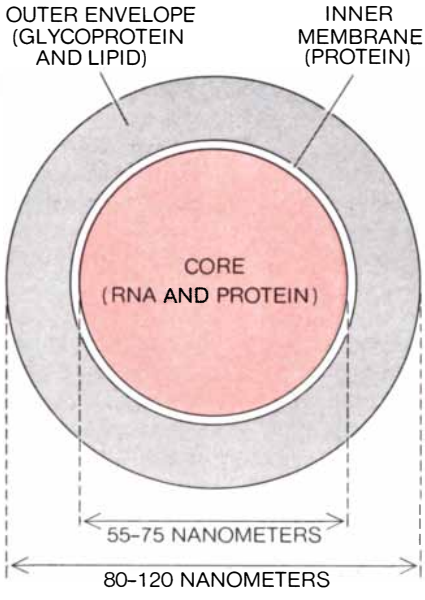
© 1971 SCIENTIFIC AMERICAN, INC

RNA-DIRECTED DNA SYNTHESIS

The discovery that in certain cancer-causing animal viruses genetic information flows "in reverse"—from RNA to DNA—has important implications for studies of cancer in humans

by Howard M. Temin

A major goal of present-day biology is to learn how information is coded in molecular structures and how it is transmitted from molecule to molecule in biological systems. Discovery of the rules governing this transmission is an integral part of understanding how embryonic cells differentiate into the hundreds of distinct types of cell observed in plants and animals and how normal healthy cells become cancerous.



VIRIONS, or individual particles, of an "RNA-DNA virus," an animal-tumor virus that transfers genetic information from RNA (ribonucleic acid) to DNA (deoxyribonucleic acid) in addition to the normal modes of information transfer used by cells and other viruses, are enlarged about 700,000 diameters in the electron micrograph on the opposite page. The particular RNA-DNA virions shown in thin section in the micrograph cause leukemia in mice; they are similar in structure and function to the Rous sarcoma virions discussed in this article. The electron micrograph was made by N. Sarkar of the Institute for Medical Research in Camden, N.J. A diagram of structure of a virion of this type is given above.

It has now been known for nearly 20 years that the genetic information in all living cells is encoded in molecules of deoxyribonucleic acid (DNA) consisting of two long strands of DNA wound in a double helix. The genetic information for each organism is written in a four-letter alphabet, the "letters" being the four different chemical units called bases. In the normal cell short passages of the genetic message (individual genes) are transcribed from DNA into the closely related single-strand molecule ribonucleic acid (RNA). A length of RNA representing a gene is then translated into a particular protein, a molecule constructed with a 20-letter alphabet, the 20 amino acids. When a cell divides, the information contained in each of the two strands of DNA is replicated, thereby equipping the daughter cell with the full genetic blueprint of the parent.

Francis Crick, one of the codiscoverers of the helical structure of DNA, originally proposed that information can be transferred from nucleic acid to nucleic acid and from nucleic acid to protein, but that "once information has passed into protein it cannot get out again," that is, information cannot be transferred from protein to protein or from protein to nucleic acid. These concepts were simplified into what came to be known as the "central dogma" of molecular biology, which held that information is sequentially transferred from DNA to RNA to protein [*see illustration on next page*]. Although Crick's original formulation contained no proscription against a "reverse" flow of information from RNA to DNA, organisms seemed to have no need for such a flow, and many molecular biologists came to believe that if it were discovered, it would violate the central dogma.

I shall describe experiments that originally hinted at a flow of information from RNA to DNA and that since have provided strong evidence that the "reverse" flow of information not only takes place but also accounts for the puzzling behavior of a sizable group of animal viruses whose genetic information is encoded in RNA rather than in DNA. Many of these viruses also produce cancer in animals. Although they have not yet been linked to cancer in man, their ability to transmit information from RNA to DNA inside the living cell makes it attractive to unify two hypotheses of the cause of human cancer that had previously seemed separate: the genetic hypothesis and the viral hypothesis.

There are two broad classes of viruses: viruses whose genome, or complete set of genes, consists of DNA and viruses whose genome consists of RNA. In the cells that they infect the DNA viruses replicate their DNA into new DNA and transmit information from DNA to RNA and thence into protein. Most RNA viruses, such as the viruses that cause poliomyelitis, the common cold and influenza, replicate RNA directly into new copies of RNA and translate information from RNA into protein; no DNA is directly involved in their replication.

In the past few years it has become apparent that a group of viruses, variously called the RNA tumor viruses, the leukoviruses or the rousviruses (after their discoverer, Peyton Rous), replicate by another mode of information transfer. The rousviruses use information transfer from RNA to DNA in addition to the modes of information transfer (DNA to DNA, DNA to RNA and RNA

© 1971 SCIENTIFIC AMERICAN, INC

to protein) that are found in cells and in DNA viruses. The rousviruses do not transfer information from RNA to RNA, as other RNA viruses do. The existence of the RNA-to-DNA mode of information transfer in the replication of rousviruses has led some to suggest that there should be three major classes of viruses: DNA viruses, RNA viruses and RNA-DNA viruses [*see top illustration on opposite page*].

The prototype RNA tumor virus, the Rous chicken sarcoma virus, was discovered by Rous 61 years ago at the Rockefeller Institute for Medical Research. An RNA tumor virus had actually been found earlier by V. Ellerman and O. Bang of Copenhagen, but their virus was little studied because it caused leukemia in chickens and was harder to work with than Rous's virus. Rous was studying a transplantable tumor of the barred Plymouth Rock hen. Originally he observed that he could transfer the tumor by the transfer of cells. In 1911 he found that the tumor could also be transferred by means of fluid from which the cells had been filtered. Demonstration that a disease can be transmitted by a cell-free filtrate is commonly accepted as evidence that it is caused by a virus. Descendants of the virus originally discovered by Rous are still being worked on in laboratories all over the world. At the time, however, Rous's discovery was met with disbelief, and after 10 years Rous himself stopped working with the tumor. It was not until nearly 30 years later, when Ludwik Gross of the Veterans Administration Hospital in the Bronx discovered that RNA tumor viruses cause leukemia in mice, that the study of rousviruses became popular.

It is now known that viruses in the same group as the virus originally discovered by Rous, or closely related to it, can cause tumors not only in chickens and mice but also in rats, hamsters, monkeys and many other species of animals. Moreover, viruses of the same group have been isolated from nonmammalian species, including snakes. As yet no bona fide human rousvirus has been discovered. It also appears that some members of this group, for example some of the "associated viruses," do not produce cancer.

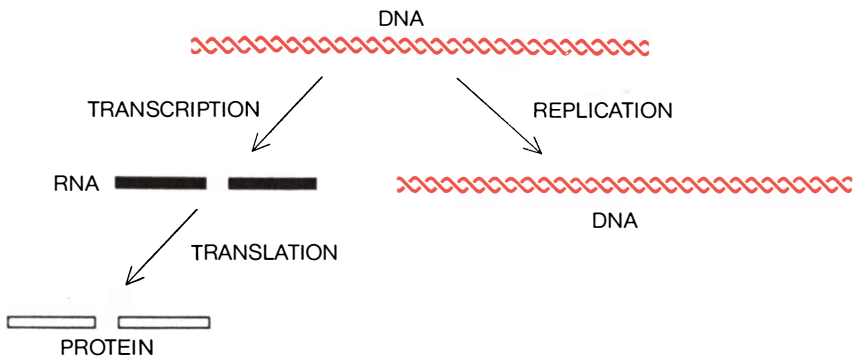
In the 1950's, with the beginning of the application of cell-culture methods to animal virology, a tissue-culture assay for the Rous sarcoma virus was developed, first by Robert A. Manaker and Vincent Groupé at Rutgers University and subsequently by Harry Rubin and me at the California Institute of Technology. The assay involves adding suspensions of the virus to sparse cultures of cells taken from the body wall of chicken embryos. The Rous sarcoma virus infects some cells and transforms them into tumor cells. The transformed tumor cells differ in morphology and in growth properties from normal cells and therefore create a focus of altered cells. Assays of the same type have been developed for infections that the Rous sarcoma virus causes in cells taken from turkeys, ducks, quail and rats. Similar assays have also been developed for other transforming rousviruses.

The number of foci of transformed cells is proportional to the number of infectious units of the virus added to the cell culture and provides a rapid and reproducible assay for the Rous sarcoma virus. The use of this assay led to the discovery that the Rous sarcoma virus differs from the other viruses that had been studied up to that time in the way it interacts with the cell. The replication of most viruses is incompatible with cell division; in other words, the virus causes the infected cells to die. Chicken cells infected with the Rous sarcoma virus not only survive but also continue to divide and produce new virus particles [*see middle illustration on opposite page*]. When the Rous sarcoma virus infects rat cells, there is a slightly different interaction of the cell and the virus. The rat cells are transformed into cancer cells, which divide, but the transformed cells do not produce the Rous sarcoma virus even though the genome (DNA) of the virus can be shown to be present. Production of the Rous sarcoma virus can be induced if the transformed rat cells are fused with normal chicken cells.

In the early 1960's the antibiotic actinomycin D was found to be very useful in unraveling the flow of genetic information in cells infected with RNA viruses. The antibiotic inhibits the synthesis of RNA made on a DNA template but not the synthesis of RNA made on an RNA template. The antibiotic therefore stops all RNA synthesis in cells infected by RNA viruses except for RNA specifically related to the viral genome. With this new tool it became easy to determine which RNA's were specific for the viruses.

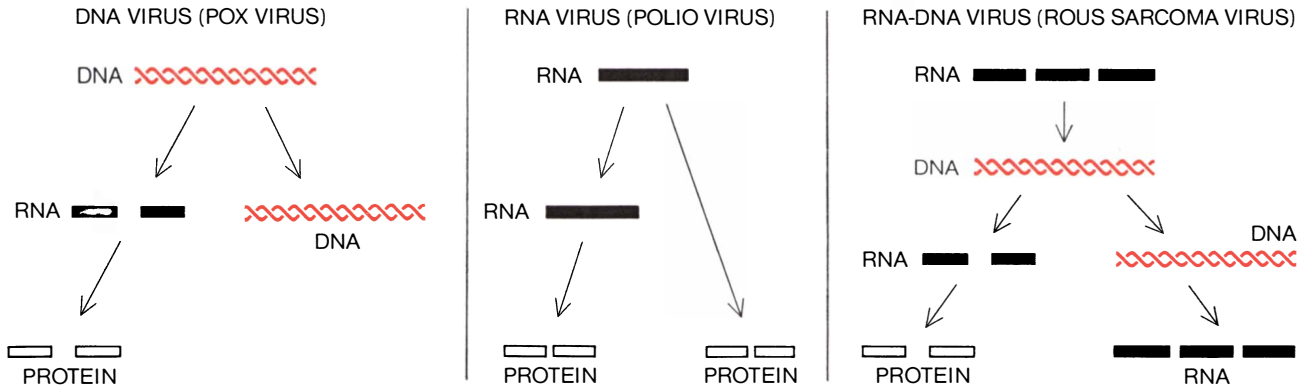
When I added actinomycin D to cultures of cells producing Rous sarcoma virus, however, I found that the antibiotic inhibited the production of *all* RNA. One would have expected the replication of RNA on the template of an RNA viral genome to continue without hindrance [*see bottom illustration on opposite page*]. This result was the first direct evidence that the molecular biology of the replication of Rous sarcoma virus was different from that of other RNA viruses. Since that observation was made the inhibition of the replication of rousviruses by actinomycin D has been recognized as one of their defining characteristics. The actinomycin D experiments suggested to me that the Rous sarcoma virus might replicate through a DNA intermediate. This hypothesis is called the DNA provirus hypothesis.

Further experiments, carried out by me and by John P. Bader at the National Cancer Institute, demonstrated that if one inhibits the synthesis of DNA in cells immediately after they have been inoculated with Rous sarcoma virus, one can protect the cells from infection. Here the inhibitors were amethopterin, fluorodeoxyuridine and cytosine arabinoside. These experiments appeared to support the idea that infection requires the synthesis of new viral DNA pro-

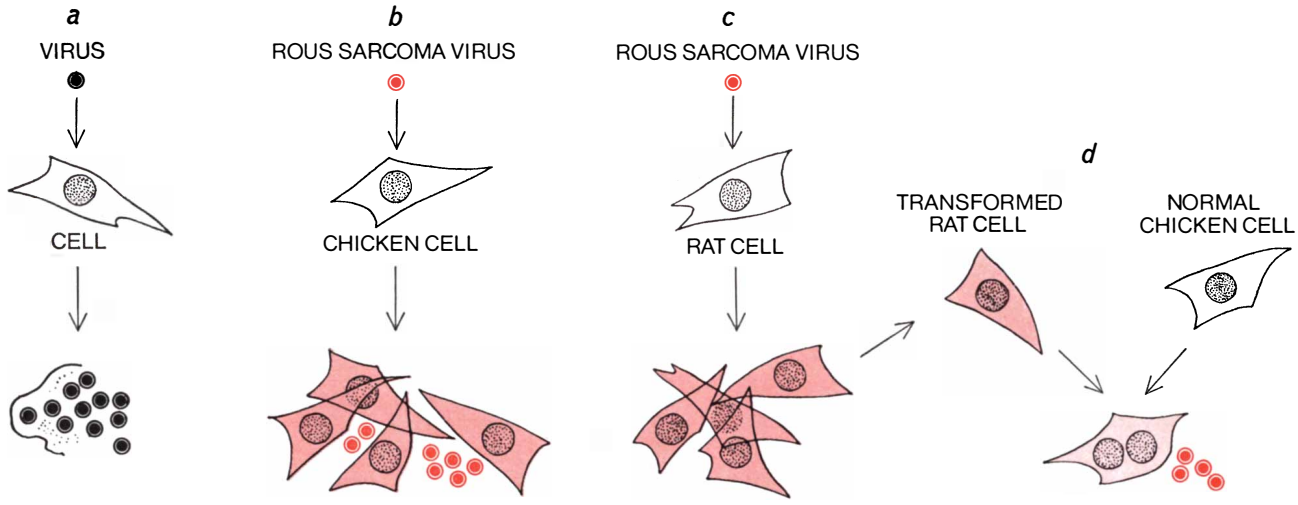


"CENTRAL DOGMA" of molecular biology, originally formulated by Francis Crick, states that within an organism genetic information can be transferred from DNA to DNA or from DNA to RNA to protein, but that it cannot be transferred from protein to protein or from protein to either DNA or RNA. Although a "reverse" flow of genetic information from RNA to DNA was not proscribed in Crick's original formulation, many molecular biologists came to believe that if such a flow were ever discovered, it would violate the central dogma.

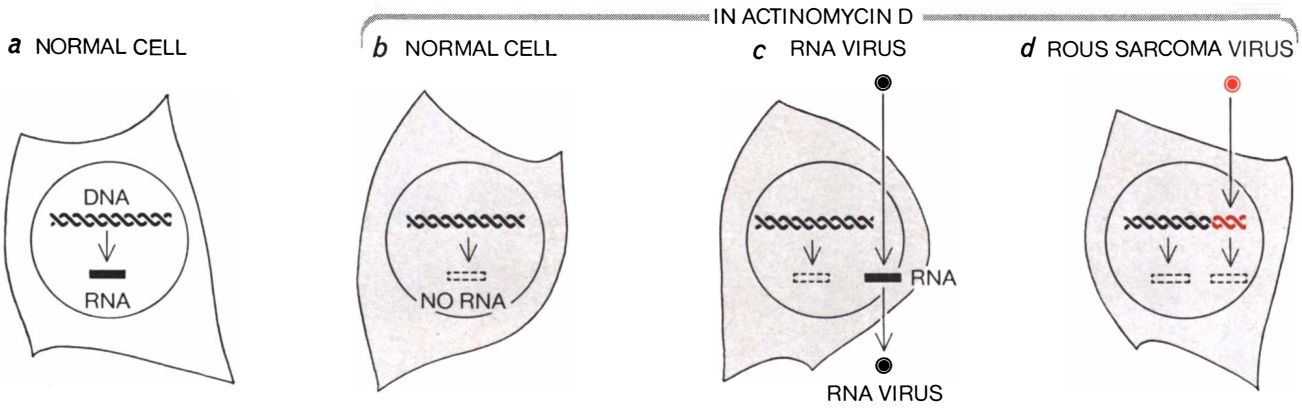
© 1971 SCIENTIFIC AMERICAN, INC



VIRUSES CAN BE GROUPED into three major classes: DNA viruses (*left*), whose genome, or complete set of genes, consists of DNA; RNA viruses (*middle*), whose genome consists of RNA, and RNA-DNA viruses (*right*), the most recently discovered group, whose genome consists alternately of RNA and DNA. A prototype tumor virus in each class is indicated in parentheses next to the class name. The diagrams illustrate the mode of information transfer that characterizes the replication of viruses in each class.



VIRUS-CELL INTERACTION usually leads to the death of the infected cell (*a*), since the replication of most viruses is incompatible with cell division. The Rous sarcoma virus, however, interacts with cells in a different way. Chicken cells infected with the Rous sarcoma virus (*b*) not only survive but also are transformed into cancer cells, which continue to divide and produce new virions. Rat cells infected with the Rous sarcoma virus (*c*) are transformed into cancer cells, which divide but do not produce new virions. By fusing the transformed rat cells with normal chicken cells the production of Rous sarcoma virions can be induced (*d*).



SYNTHESIS OF RNA on a DNA template in normal cells (*a*) is inhibited by the addition of the antibiotic actinomycin D (*b*). Since the antibiotic does not affect the synthesis of RNA made on an RNA template, however, it does not stop RNA synthesis specifically related to the viral genome in cells infected by most RNA viruses (*c*). The finding that actinomycin D inhibited the production of *all* RNA in cells producing Rous sarcoma virus (*d*) was the first direct evidence that the molecular biology of the replication of Rous sarcoma virus was different from that of other RNA viruses. The actinomycin D experiments led the author to propose the DNA provirus hypothesis, which holds that rousviruses such as the Rous sarcoma virus replicate through a DNA intermediate.

© 1971 SCIENTIFIC AMERICAN, INC

duced on an RNA template. This interpretation was not unequivocal, however, because successful production of Rous sarcoma virus requires that the cells divide normally after infection. Therefore the inhibition of DNA synthesis after infection could inhibit production of Rous sarcoma virus not only by blocking possible new viral DNA synthesis but also by preventing normal cell division.

To get around this problem I introduced the idea of infecting cultures of stationary, or nondividing, cells with Rous sarcoma virus. Cells in culture usually require specific factors in blood serum to support their multiplication. If the serum is removed from the medium of the cell cultures, the cells stop dividing. If they are then exposed to Rous sarcoma virus, they become infected but there is no virus production or morphological transformation until serum is added back and the cells divide once again. When such stationary cells are exposed to inhibitors of DNA synthesis, the cells are not killed because they are not making DNA. When the stationary cells are exposed simultaneously to Rous sarcoma virus and to inhibitors of DNA synthesis, the cells are not killed but neither are they infected [*see illustration at right*].

If one now removes the inhibitor of DNA synthesis and adds serum, enabling the cells to divide once more, one finds that the cells remain free of infection. They do not become transformed and they do not produce virus. These experiments supported the hypothesis that after cells are infected by the Rous sarcoma virus new viral DNA is synthesized at a time different from the cell's normal synthesis of DNA. The new viral DNA is evidently synthesized on a template of viral RNA.

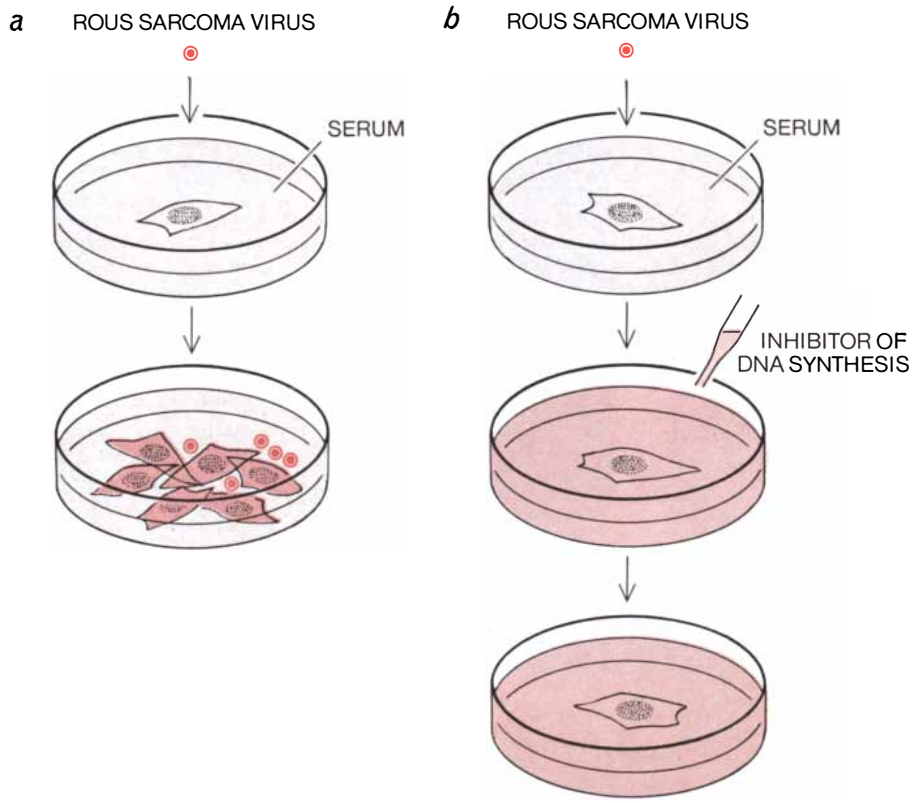
A further extension of this approach to understanding the replication of Rous sarcoma virus was carried out by one of my students, David E. Boettiger, and independently by Piero Balduzzi and Herbert R. Morgan at the University of Rochester School of Medicine and Dentistry. It had been found by others that if 5-bromodeoxyuridine, an analogue of the DNA constituent thymidine, is incorporated into DNA, the DNA becomes sensitized so that it can be inactivated by light. Under the same conditions normal thymidine-containing DNA is not affected by light. Boettiger therefore exposed stationary cells to Rous sarcoma virus in the presence of bromodeoxyuridine and then exposed the cells to light. Although the cells were not killed, the treatment prevented their being infected by the virus. When serum was again added to enable the cells to divide, they did not become transformed and did not produce virus [*see illustration on page 30*].

In a related experiment Boettiger showed that the rate of inactivation of the infection by Rous sarcoma virus was dependent on the number of viruses infecting a cell. As he raised the number of viruses infecting each cell, he found that the infection became increasingly resistant to inactivation by light. We interpreted these experiments as showing that each infecting virus makes a new specific DNA, and that the more viruses that infect a cell, the more molecules of new viral DNA that are produced. The experiment seemed to effectively rule out the alternative hypothesis, which was that the infecting virus provokes a new synthesis of some preexisting cellular DNA.

Unfortunately no one has yet been able to unequivocally demonstrate the existence of newly synthesized viral DNA in cells infected with the Rous sarcoma virus. The available techniques are evidently too crude to detect the tiny amounts of new viral DNA expected to be present. Certain results have been reported, however, with transformed cells. One approach has been to bring DNA from infected cells together with labeled viral RNA to see if single strands of the two molecules would coalesce into a double-strand hybrid molecule. Such hybrids are readily created when the base sequences in the DNA are complementary to the base sequences in the RNA, indicating that both carry the same genetic message and hence that each could arise from the transcription of the other.

The hybridization experiments reported thus far have aroused a great deal of controversy. Although some experiments, notably those of Marcel A. Baluda and Debi P. Nayak of the University of California at Los Angeles, have seemed to demonstrate the presence in infected cells of DNA complementary to viral RNA, the results have not been universally accepted. The finding of an intermediate viral DNA is an essential link in the chain of evidence that is still needed to establish firmly the DNA provirus hypothesis.

Meanwhile strong support for the hy-



EXPERIMENTS carried out by the author and by John P. Bader at the National Cancer Institute supported the hypothesis that the infection of cells with Rous sarcoma virus requires the synthesis of new viral DNA produced on an RNA template. When the virus is added to cultures of normally dividing cells (*a*), the cells are transformed into cancer cells, which divide and produce new Rous sarcoma virus. By adding a substance that inhibits the synthesis of DNA in the cells immediately after they have been inoculated with Rous sar-

© 1971 SCIENTIFIC AMERICAN, INC

pothesis has come from experiments of a different kind. In 1969 Satoshi Mizutani, who had written his doctoral thesis on bacterial viruses, came to my laboratory for postdoctoral training. We decided to ask the question: What is the origin of the enzyme (a protein) responsible for forming proviral DNA using the viral RNA as template? When Mizutani exposed stationary cells to Rous sarcoma virus in the presence of inhibitors of protein synthesis, he found that the cells still became infected. We interpreted this experiment to mean that the enzyme that synthesizes DNA from the viral RNA template is already in existence before the infection.

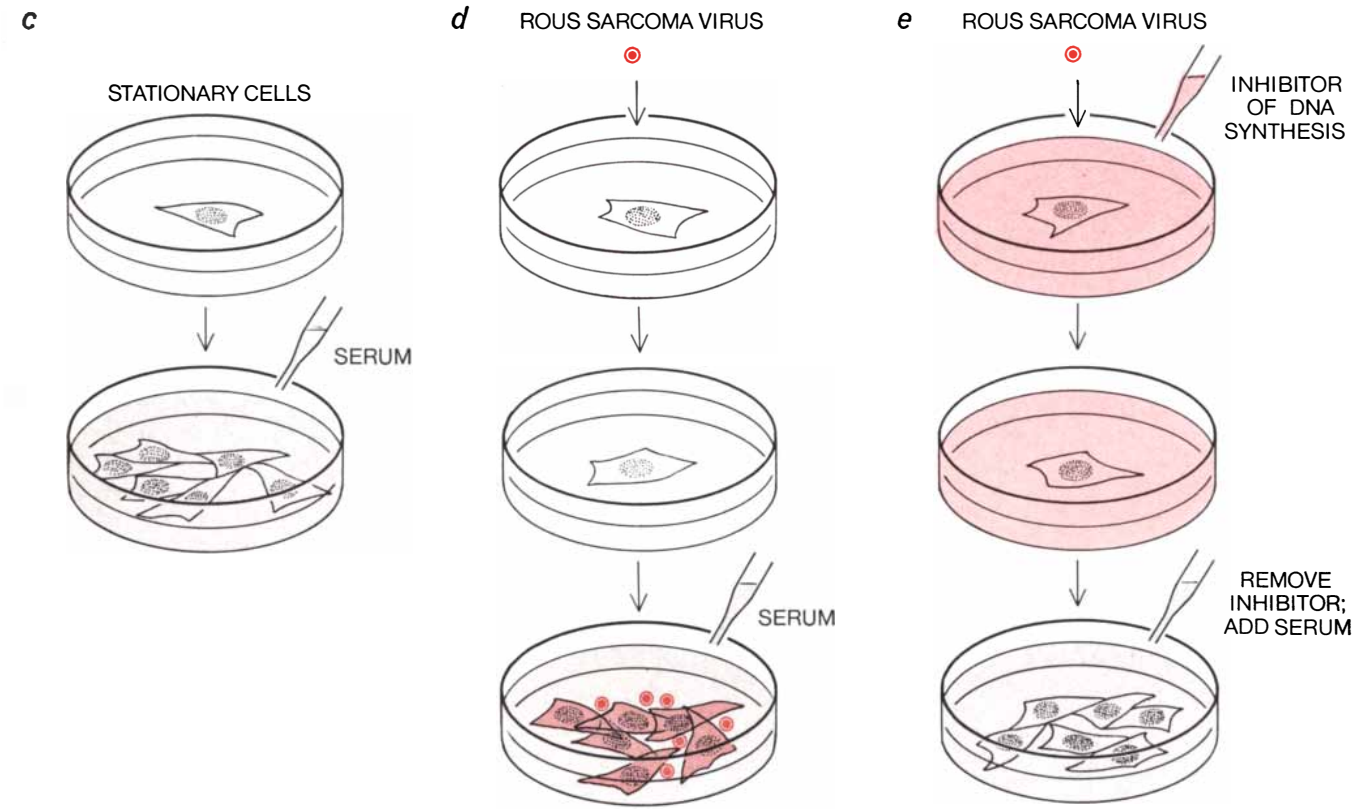
Somewhat earlier other workers had fractionated virions—the actual virus particles as distinct from the forms assumed by the virus inside cells—and had found RNA polymerases, enzymes that catalyze the synthesis of RNA from its building blocks: four different ribonucleoside triphosphates. In 1967 Joseph Kates and B. R. McAuslan of Princeton University and William Munyon, E. Paoletti and J. T. Grace, Jr., of the Roswell Park Institute had found RNA polymerases in a poxvirus, a large DNA virus. Other workers had found another RNA polymerase in a reovirus, a double-strand RNA virus. Therefore we decided to look in the virions of Rous sarcoma virus for a DNA polymerase capable of using the viral RNA as a template. After several months of preliminary experiments we succeeded in showing the existence of a DNA polymerase in purified virions of Rous sarcoma virus.

Before discussing this result I should digress briefly to describe the structure of the Rous sarcoma virus [*see illustration on page 24*]. The virion of the Rous sarcoma virus has a diameter of about 100 nanometers, which makes it larger than the particles of the viruses that cause poliomyelitis and smaller than the particles of the viruses that cause smallpox. The virion of the Rous sarcoma virus consists of a lipid-containing envelope (derived by budding from the cell membrane), an inner membrane and a nucleoid, or core, that contains the viral RNA and certain proteins.

In order to demonstrate that the Rous sarcoma virus contains a polymerase capable of producing DNA on an RNA template, we first treated the virion with a detergent to disrupt its lipid-containing envelope. We then added to the disrupted virus the four deoxyribonucleoside triphosphates that are the building blocks of DNA. One of the deoxyribonucleoside triphosphates was radioactively labeled.

When the mixture was incubated at 40 degrees Celsius, it incorporated the radioactive label into an acid-insoluble substance that met the usual tests for DNA. The substance was stable in the presence of alkali and the enzyme ribonuclease, treatments that are known to destroy RNA, whereas it was attacked and fragmented by an enzyme that destroys DNA. When we repeated the experiment with disrupted virions pretreated with ribonuclease, an enzyme that destroys RNA, little or no DNA was produced, indicating that intact viral RNA was needed as the template for the synthesis of DNA [*see top illustration on page 31*].

After we had announced these results at the Tenth International Cancer Congress in Houston in May, 1970, we learned that David Baltimore of the



coma virus (*b*), one can protect the cells from infection. In subsequent experiments by the author cultures of stationary, or nondividing, cells were used; when blood serum is added to such cultures (*c*), they divide normally. If such stationary cells are first exposed to Rous sarcoma virus (*d*), however, they become infected but there is no virus production or morphological transformation until serum is added back and the cells divide once again. When the stationary cells are exposed simultaneously to Rous sarcoma virus and to an inhibitor of DNA synthesis (*e*), the cells are not killed but neither are they infected; when the inhibitor of DNA synthesis is removed and serum is added, cells divide normally, are not infected, do not become transformed and do not produce virus.

© 1971 SCIENTIFIC AMERICAN, INC

Massachusetts Institute of Technology had independently made similar observations with the virion of a mouse leukemia virus. The two papers describing these findings were published together in the June 27, 1970, issue of *Nature,* the British scientific weekly. The two publications stimulated an enormous amount of work whose peak is not yet in sight.

In our early papers we called the new viral enzyme RNA-dependent DNA polymerase because the template was RNA and the product was DNA. Subsequently we and others found that the enzyme could also use DNA as a template for DNA synthesis. We therefore decided to change the word "dependent" to "directed," so that we now refer to the enzyme as RNA-directed DNA polymerase. The revised name makes no statement about the origin of the enzyme or its relation to other DNA polymerases. Independently *Nature* began referring to the enzyme as "reverse transcriptase," a name that I do not like because of its ambiguity but that has gained wide currency.

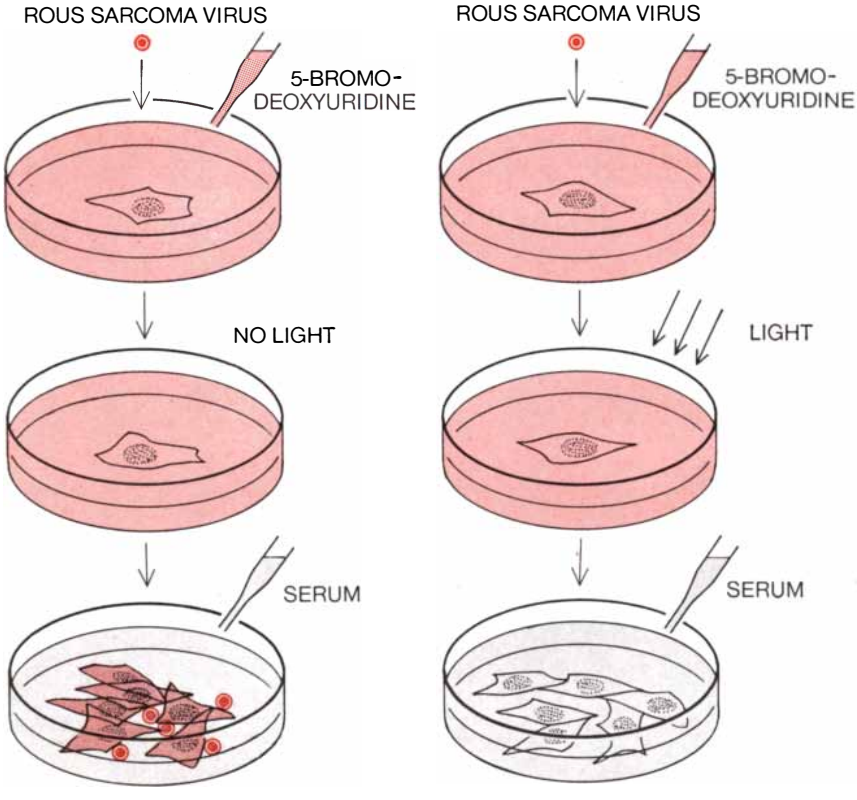
All the later studies confirm the original finding that the virions of RNA tumor viruses contain a DNA polymerase system that is activated by treating the virion with a detergent and that is sensitive to ribonuclease. Moreover, the virion enzyme functions only as a DNA polymerase; it will not act as an RNA polymerase. As I have mentioned, however, other unrelated RNA viruses do contain an RNA polymerase.

If the DNA produced by the RNA-directed DNA polymerase is isolated free of protein, the size of its molecule can be estimated by spinning it at high speed in a sucrose gradient in an ultracentrifuge. The molecule is surprisingly small: less than a tenth as long as one would expect a copy of the complete viral RNA to be. The reason for the small size is still elusive. If the isolated DNA product is centrifuged in a cesium sulfate density gradient, which separates RNA from DNA on the basis of their different densities, one finds that the product has the density of DNA [*see bottom illustration on opposite page*]. Further characterization, for example by treatment with enzymes that specifically attack either single- or double-strand DNA, shows that the product of the DNA polymerase system is a double strand. From such studies one can conclude that the DNA polymerase system of the virion makes short pieces of double-strand DNA.

Many workers have demonstrated that the DNA product of the RNA-directed DNA polymerase system has a base sequence complementary to the viral RNA [*see top illustration on page 32*]. This conclusion is drawn from annealing, or molecular hybridization, experiments. Labeled DNA from the virion polymerase reaction is treated so that the strands of the DNA dissociate. The single-strand DNA is added to unlabeled viral RNA, and the mixture is incubated so that complementary strands can form a hybrid combination. The mixture is then centrifuged in a cesium sulfate density gradient. About half of the product DNA forms a band at a density characteristic of RNA or of hybrid RNA-DNA molecules rather than at a density characteristic of DNA. The test is quite specific and indicates that the DNA polymerase of the virion copies the sequence of the bases of the viral RNA into DNA. This experiment, however, still does not demonstrate that such a copying process takes place in cells infected by Rous sarcoma virus.

The viral DNA polymerase was shown to be present in the core of the virion by the following experiment carried out by George Todaro's group at the National Cancer Institute and by John M. Coffin in my laboratory at the University of Wisconsin. Rousvirus virions were treated with a detergent to disrupt the envelope. Then the disrupted virus was centrifuged in a sucrose density gradient. Most of the viral RNA, about 20 percent of the protein and most of the RNA-directed DNA polymerase activity were found to sediment together in "cores," a term given to structures that are denser than whole virions [*see bottom illustration on page 32*]. Further studies showed that with more extensive disruption of the virion the viral DNA polymerase can be freed from the viral RNA and then purified. The purified enzyme is capable of directing the synthesis of DNA on a variety of templates: synthetic and natural DNA, RNA and RNA-DNA hybrids.

The general conclusion from studies in a number of laboratories is that the

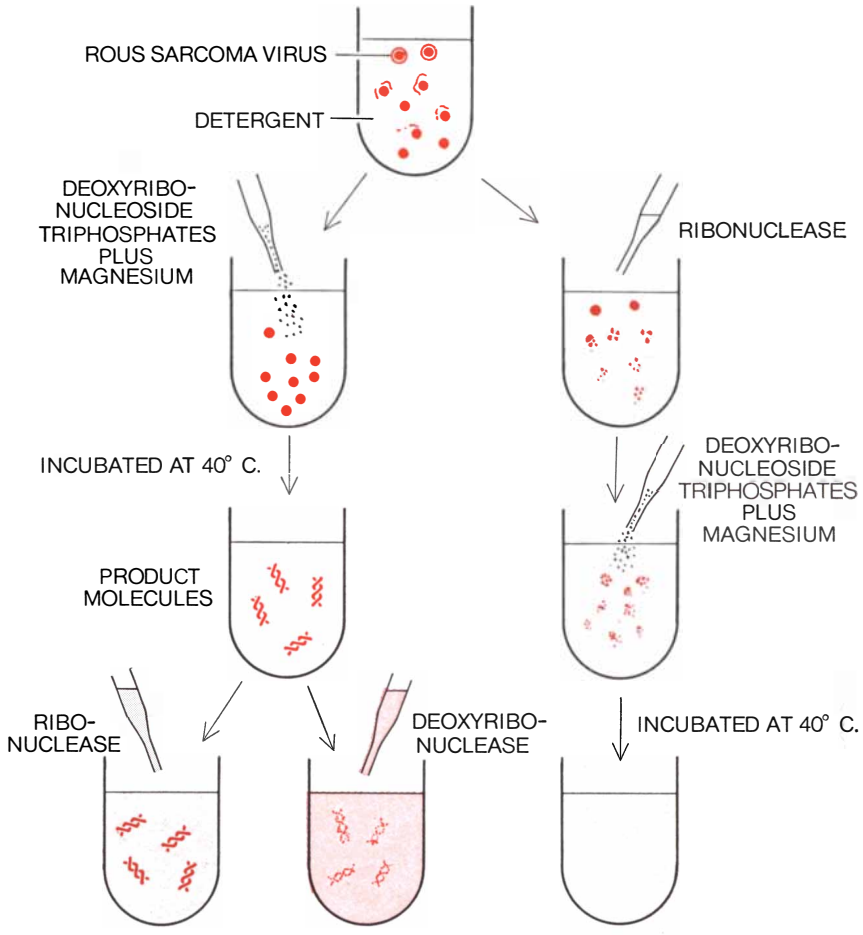


FURTHER EXPERIMENTS, carried out by one of the author's students, David E. Boettiger, and independently by Piero Balduzzi and Herbert R. Morgan at the University of Rochester, involved exposing stationary cells to Rous sarcoma virus in the presence of 5-bromodeoxyuridine, an analogue of the DNA constituent thymidine that, when incorporated into DNA, sensitizes the DNA to inactivation by light. As a control some of the treated cells were first not exposed to light (*left*); after serum was added to these cells to enable them to divide they were transformed into cancer cells and began to produce virus. When another culture of treated cells was exposed to light (*right*), the cells were not killed, but the treatment prevented their infection by the virus. When serum was again added to enable these cells to divide, they did not become transformed and did not produce virus.

© 1971 SCIENTIFIC AMERICAN, INC

rousvirus DNA polymerase closely resembles the other DNA polymerases described above that are present in more familiar biological systems and that catalyze the synthesis of DNA on a DNA template. In other words, it is not a unique property of the rousvirus DNA polymerase to be able to use RNA as a template for DNA synthesis. (This was first proposed several years ago by Sylvia Lee Huang and Liebe F. Cavalieri of the Sloan-Kettering Institute.) What is unique so far is the apparent biological role of RNA-directed DNA synthesis in the replication of rousviruses.

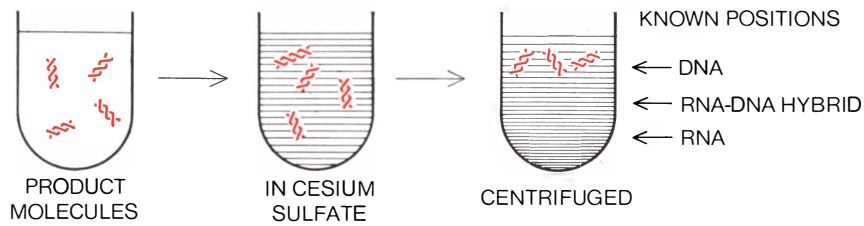
Further work in my laboratory has shown that preparations of purified virions of the Rous sarcoma virus contain other enzymes related to DNA replication. The most unusual of them is an enzyme that is named polynucleotide ligase, which repairs breaks in DNA molecules. It is an attractive hypothesis that the function of the ligase is to join the viral DNA to the chromosomal DNA of the host cell, thus integrating the viral genome with the cell genome. After this integration the genetic information of the virus would be replicated with that of the host and passed from the parent cell to the daughter cell. The Rous sarcoma virus virion also contains many other enzymes whose role is completely unknown. We do not know whether they participate in the life cycle of the virus or whether they are merely accidental contaminants picked up in the formation of the virion.



EXISTENCE OF A POLYMERASE capable of producing DNA on an RNA template in RNA tumor viruses was demonstrated by the author and his colleague Satoshi Mizutani (and also independently by David Baltimore of the Massachusetts Institute of Technology). In the experiment conducted by Mizutani and the author purified virions of Rous sarcoma virus were first treated with a detergent to disrupt their lipid-containing envelope. Four deoxyribonucleoside triphosphates, the "building blocks" of DNA, were then added to the disrupted virions. When the mixture was incubated, it incorporated the radioactive label associated with one of the building blocks into an acid-insoluble substance that was stable in the presence of ribonuclease (an enzyme known to destroy RNA), whereas it was fragmented by deoxyribonuclease (an enzyme that destroys DNA). When the experiment was repeated with disrupted virions pretreated with ribonuclease, little or no DNA was produced, indicating that intact viral RNA was needed as template for synthesis of DNA.

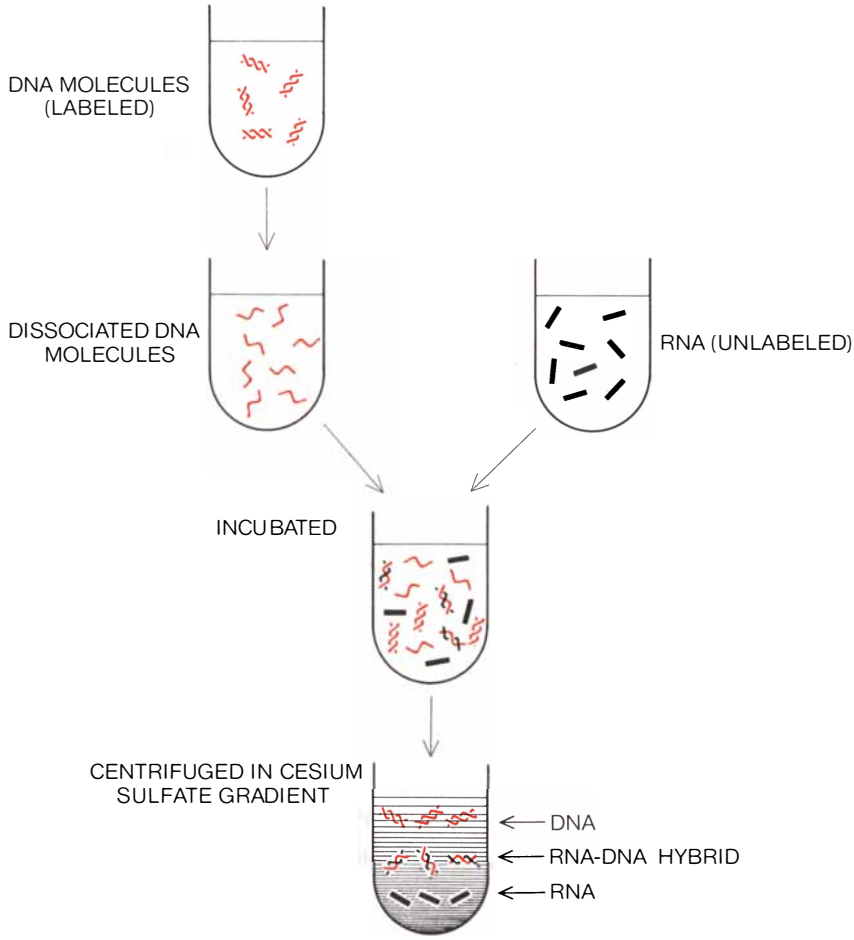
After the first discovery of a DNA polymerase in the virions of RNA tumor viruses, a great many other RNA viruses were examined to see if they contain a similar DNA polymerase system. First it was found that all the viruses previously classified in the RNA tumor virus group contain such an enzyme system. This group of RNA viruses includes both the rousviruses that cause tumors and those that do not cause tumors. Even more interesting, it was found that two types of virus that had not been classified in the same group with RNA tumor viruses also contain a DNA polymerase system. One of these viruses is Visna virus, which causes a slowly developing neurological disease in sheep. After the demonstration of a DNA polymerase in virions of the Visna virus, Kenneth Kaname Takemoto and L. B. Stone at the National Institutes of Health showed that the same virus could cause cancerous transformation of mouse cells in culture. Therefore Visna virus can now be considered a transforming rousvirus. The other type of virus that has been found to have a DNA polymerase system is the "foamy," or syncytium-forming, viruses. These viruses, isolated from monkeys and cats, have not been connected with any particular disease but are common contaminants of cell cultures. They have not yet been shown to cause tumors or cancerous transformation.

The DNA polymerase present in RNA tumor viruses may not only explain how these viruses produce stable cancerous transformations in the cells they infect but also account for some viral latency,

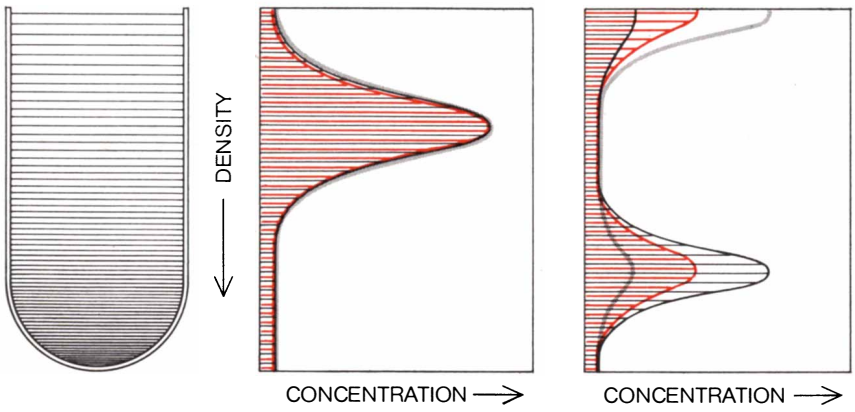


CENTRIFUGATION of the isolated DNA product of the RNA-directed DNA polymerase system in a cesium sulfate density gradient (which separates RNA from DNA on the basis of their different densities) resulted in the finding that the product has the density of DNA. In combination with other findings this result led to the conclusion that the DNA polymerase system of the Rous sarcoma virus virion makes short pieces of double-strand DNA.

© 1971 SCIENTIFIC AMERICAN, INC



MOLECULAR-HYBRIDIZATION EXPERIMENTS demonstrated that the DNA product of the RNA-directed DNA polymerase system within the virion copies the sequence of bases of the viral RNA into DNA. Labeled DNA from the virion polymerase reaction was first treated so that strands of the DNA dissociated. The single-strand DNA was then added to unlabeled viral RNA, and the mixture was incubated at high temperature so that complementary strands could form a hybrid combination. When the resulting "annealed" mixture was centrifuged in a cesium sulfate density gradient, about half of the product DNA was observed to form a band at a density characteristic of hybrid RNA-DNA molecules.



PRESENCE OF VIRAL DNA POLYMERASE in the cores of the Rous sarcoma virus virions was demonstrated by John M. Coffin in the author's laboratory. The curves at center show the density distribution of various radioactively labeled constitutents of the whole virions as determined by centrifugation in a sucrose density gradient (*left*). The curves at right show the density distribution of the same constituents determined by centrifugation after the virions were treated with a detergent to disrupt their envelopes. Most of the viral RNA (*black curves*), about 20 percent of the protein (*gray curves*) and most of the RNA-directed DNA polymerase activity (*colored curves*) of the disrupted virions were found to sediment together at a higher density than the corresponding constituents of the whole virions, indicating that these constituents are concentrated in the cores of the virions.

the phenomenon in which a virus disappears after infecting an organism only to reappear months or years later. Once an RNA virus has transferred its genetic information to DNA, it would be able to remain latent in a cell and be replicated by the cellular enzyme systems that replicate and repair the cell DNA. After some later activation the virus could appear again as infectious virus particles [*see top illustration on opposite page*].

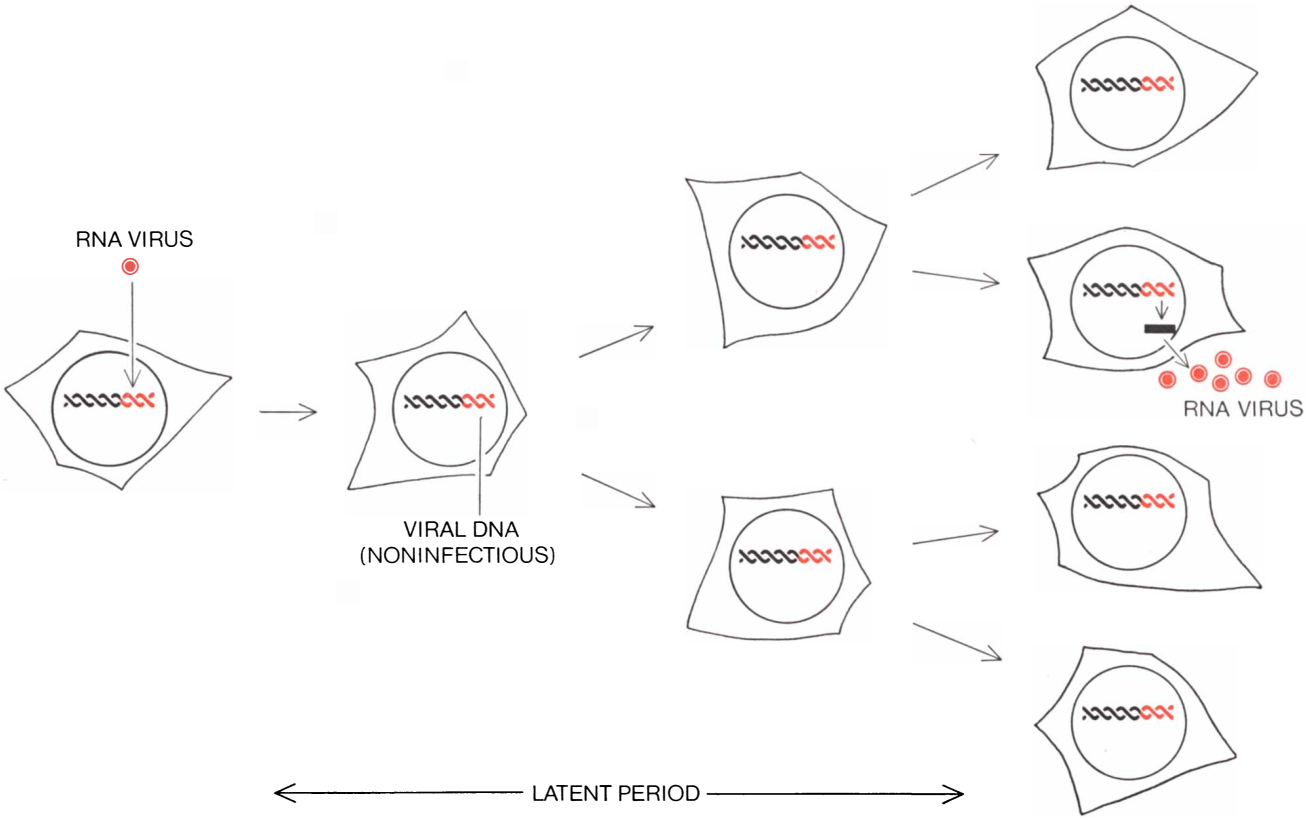
About a year ago considerable public excitement was generated by the reported discovery of "RNA-dependent DNA polymerases" in human tumor cells. The general conclusion I would draw now from most of this work has been stated above: All DNA polymerases are capable, under the appropriate conditions, of transcribing information from RNA into DNA. At present we lack generally accepted criteria for determining whether or not such syntheses have any biological role or any relation to rousviruses.

In my laboratory we have taken a slightly different approach to the question of RNA-directed DNA synthesis in cells. We have used detergent activation and ribonuclease sensitivity as criteria in a broad search for DNA polymerase systems in a variety of animal cells. That is, we have looked in cells for a DNA polymerase system similar to viral "cores." Coffin has found such a DNA polymerase system in normal, uninfected rat embryo cells. So far we do not know the full significance of this discovery, but it suggests that ribonuclease-sensitive DNA polymerase systems are present in cells other than tumor cells or virus-infected cells.

For many years I have favored the idea that RNA-directed DNA synthesis may be important in normal cellular processes, particularly those involved in the embryonic differentiation of cells. This idea has been expanded in the form of the protovirus hypothesis [*see bottom illustration on opposite page*]. The general idea is that in normal cells there are regions of DNA that serve as templates for the synthesis of RNA, and that this RNA serves in turn as a template for the synthesis of DNA that subsequently becomes integrated with the cellular DNA. By this means certain regions of DNA can be amplified. With additional processes that introduce changes in the DNA, the DNA of different cells can be made different. This difference might serve as a means of distinguishing different cells.

What, then, are the general implications of this work for the prevention or treatment of human cancer? We can

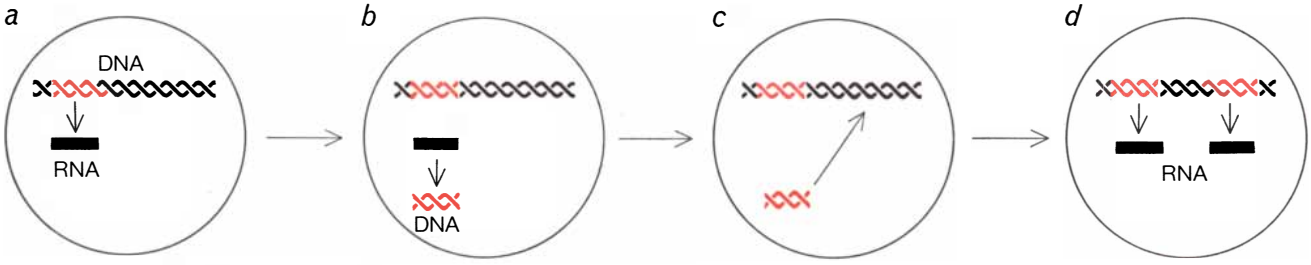
© 1971 SCIENTIFIC AMERICAN, INC



LATENCY OF RNA VIRUSES after infecting an organism may be attributable to the DNA polymerase system present in the cores of such viruses. After transferring its genetic information to DNA in the cell nucleus (*left*), the RNA virus would disappear, remaining latent in the cell by virtue of its replication by the cellular enzyme systems that replicate and repair the cell's DNA (*left center, right center*). Months or years later some form of activation could then cause the infectious RNA virions to appear again (*upper right*).

conclude only that some biological systems utilize a previously undescribed mode of information transfer: from RNA to DNA. It is an interesting coincidence that this new mode of information transfer was first discovered in tumor-causing viruses. We cannot say, however, that RNA-directed DNA synthesis is an exclusive property of such viruses. What the discovery of RNA-directed DNA synthesis does mean is that we now have some simple biochemical tests to determine whether or not newly discovered human viruses are members of the same group as the RNA viruses that produce tumors and cancerous transformations in animal cells, and to look for information related to these viruses in human cancers. We cannot now say that inhibitors of RNA-directed DNA synthesis would have any effect on human cancer. In rousvirus-induced tumors in animals the synthesis of new viral DNA appears to be important only at the initial stage of cancerous transformation, not thereafter.

Probably the most important implication of this discovery for the understanding of cancer in man has been the removal of the dichotomy between viral and genetic theories of the origin of cancer. At a time when genes were thought to consist of DNA alterable only by mutation, and when most of the known cancer-causing animal viruses were of the RNA type, it was hard to imagine common features of genetic and viral theories. Now that we have uncovered evidence that cancer-causing RNA viruses can produce a DNA transcript of the viral RNA, one can readily formulate hypotheses in which elements related to viral RNA are attached to the genome of the cell and transmitted genetically to become activated at some future time and cause "spontaneous" cancer. Experiments designed to test this idea are now in progress in a number of laboratories around the world.



PROTOVIRUS HYPOTHESIS, put forward by the author, embodies the idea that RNA-directed DNA synthesis may be important in normal cellular processes. According to this view, there are regions of DNA in normal cells that serve as templates for the synthesis of RNA (*a*). This RNA serves in turn as a template for the synthesis of DNA (*b*), which later becomes integrated with the cellular DNA (*c*). The amplification of certain regions of DNA resulting from the repetition of the process (*d*) may, in conjunction with additional processes that introduce changes in the DNA, play an important role in the embryonic differentiation of cells.

© 1971 SCIENTIFIC AMERICAN, INC

How Ideology Shapes Women's Lives

Data from a survey of college women reveal that a woman's life goals, particularly her educational and occupational aspirations, are guided by the type of sex-role ideology acquired in childhood

by Jean Lipman-Blumen

The "women's liberation" movement has brought to the fore the age-old question of what kinds of behavior are socially appropriate for women and men. It is often said that women tend to act on the basis of feelings and emotions rather than on the basis of reason. The motivating force behind female behavior is rarely seen as being ideology. It is becoming increasingly apparent, however, that certain ideologies can predict the values and behavior of women with remarkable accuracy. These powerful systems of beliefs, which shape the destiny of women in ways never imagined by Freud, are transmitted implicitly rather than explicitly; they usually guide the behavior of women silently and without their being consciously aware of it.

Such ideologies are largely based on a woman's concept of what kinds of behavior are appropriate to her role as a female. In the study I shall describe here female-role ideology referred primarily to a woman's system of beliefs regarding the appropriate behavior of women with

"CONTEMPORARY" VIEW of the female role, as defined in the survey, is based on the belief that women should be as free as men to pursue educational and occupational goals, that men and women should equally share responsibilities inside and outside the home.

© 1971 SCIENTIFIC AMERICAN, INC

respect to men. The study, which involved an extensive survey of the life plans of married women, was conducted under the auspices of the Radcliffe Institute in Cambridge, Mass., in 1968 (before the women's liberation movement had had a major impact). A detailed questionnaire was mailed to the wives of graduate students in the Boston area. Out of the 1,868 responses a subsample of 1,012 wives who had attended college was selected for analysis. The questionnaire inquired into early childhood experiences, academic achievements and plans, past and present family situations, personal values and life goals.

The age of the women who responded ranged from 18 to 54, with 23.4 years as the median. Forty-two percent of the women had been married one year or less, another 43 percent had been married from two to five years and 15 percent had been married more than five years. The mean number of years married was 3.2. Sixty percent of the women had no children, 21 percent had one child, 12 percent had two children, 5 percent had three children and 2 percent had four or more children. Since the women selected for analysis were all married to graduate students, their socioeconomic status was fairly homogeneous. Their original family backgrounds, however, were quite varied and in their diversity were presumably not unlike the backgrounds of the larger population of women who have attended college.

"TRADITIONAL" VIEW of the female role involves the belief that a woman's primary responsibilities are homemaking and child-rearing, that men are responsible for financial support of the family and that women with children should not expect to have a career.

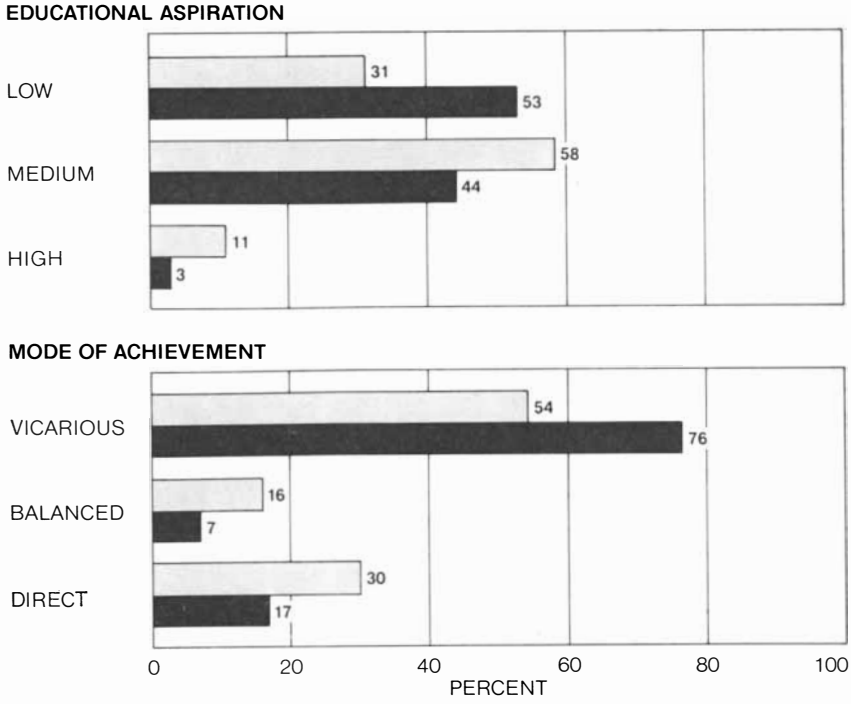
An index of female-role ideology was developed to encompass two major dimensions of the adult female role: an internal dimension, based on issues of task-sharing between husband and wife, and an external dimension, related to patterns of appropriate female behavior outside the home. Responses to a six-item scale were summed to obtain a female-role-ideology score. Although sex-role ideologies form a continuum, we grouped the respondents into two polar categories, which we labeled "traditional" and "contemporary." An oversimplified version of the traditional ideology is the belief that under ordinary circumstances women belong in the home, caring for children and carrying out domestic duties, whereas men are responsible for the financial support of the family. The contemporary ideology holds that the relationships between men and women are ideally egalitarian and that husbands and wives may share domestic, child-rearing and financial responsibilities. In our study sample 27 percent of the women adhered to the traditional ideology and 73 percent held the contemporary view.

Studies by Ruth E. Hartley and others in the early 1960's have shown that by the age of five children have developed a sex-role ideology with well-defined notions of appropriate behavior for men and women. Because girls are presumably socialized at an early age to either the traditional or the contemporary sex-role ideology, the belief system is likely to shape much of their life pattern. Here a number of important questions arise. What kind of family do women with the contemporary or traditional ideologies come from? What adolescent relationships did they have with their parents? How is female-role ideology related to a woman's present life-style? How does it affect her life choices, particularly her educational and occupational aspirations? Are the contemporary and traditional ideologies associated with different hierarchies of values? Are married women with the contemporary viewpoint as happy with their life as women with the traditional viewpoint?

In America college education is regarded as a major avenue to career opportunities and a desirable style of life. Therefore it is not surprising that there is a strong relation between the educational aspirations of a college woman and her plans to pursue a professional career. What link is there, if any, between a woman's concept of her sex role and her educational plans?

In our study educational aspiration was measured by the response to the question: "What is the highest level of academic training that you expect to obtain?" As I have noted, all the women in the study sample had completed at

© 1971 SCIENTIFIC AMERICAN, INC



SEX-ROLE IDEOLOGY affects a woman's educational aspirations and how she seeks to satisfy her need for achievement outside the home. Women with a contemporary view (*gray bars*) have higher educational goals and are more likely to satisfy their achievement needs through their own efforts than women who hold a traditional view (*black bars*).

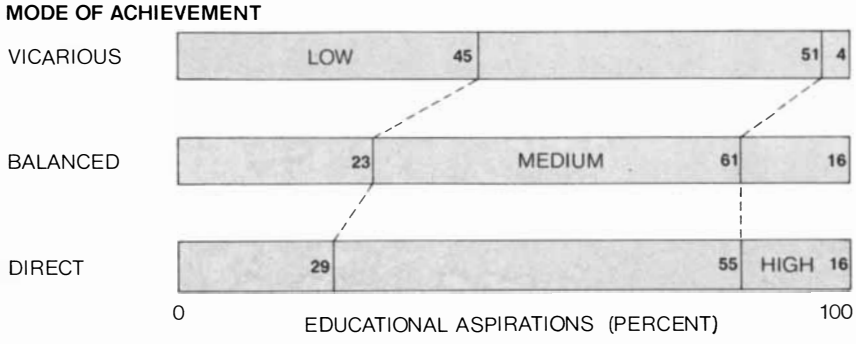
least part of a college program. The responses were grouped into three categories: low aspiration, medium aspiration and high aspiration. Women with low aspiration did not plan to seek a degree beyond the bachelor level; those with medium aspiration planned academic work up to and including a master's degree; those with high aspiration planned doctoral or postdoctoral studies.

Analysis of the data revealed that there was a strong interaction between a woman's concept of the female role and her educational aspiration. More than half of the women who held the traditional view of the female role did not plan to seek a degree beyond the bachelor's level, whereas a majority of the women with the contemporary viewpoint aspired to graduate studies, with 58 percent having medium aspiration and 11 percent having high aspiration. Only 3 percent of the women with the traditional viewpoint had high educational aspiration [*see illustration above*].

In order to gain a more precise understanding of how sex-role ideology is related to educational aspiration we examined a possible linking factor: mode of achievement. Mode of achievement is a measure of how a woman seeks to satisfy her need for achievement outside the home. There were three categories for mode of achievement: direct, balanced and vicarious. Women who preferred the direct mode felt it necessary to satisfy their achievement needs completely or predominantly through their own efforts; those who chose the balanced mode placed equal weight on their husband's accomplishments and on their own; those who selected the vicarious mode fulfilled their achievement needs either completely or predominantly through the accomplishments of their husband.

Since girls often are socialized in early childhood to satisfy their achievement needs passively by identification with the accomplishments of their father or their brothers, it is not surprising that they transfer this vicarious mode of achievement to their husband when they marry. The survey data show the strength of such socialization: a majority of all the women in the sample, both in the contemporary and in the traditional categories, sought to satisfy their achievement needs vicariously. As might be expected, those with the traditional ideology adhered much more to the vicarious mode, with 76 percent selecting this means of satisfying their achievement needs compared with 54 percent of those with the contemporary viewpoint. Almost a third of the college women with the contemporary ideology, but only a sixth of those with the traditional one, felt that they had to satisfy their achievement needs primarily through their own accomplishments. Relatively few seemed to prefer the balanced mode in which the accomplishments of husband and wife had equal weight.

There was a clear connection between mode of achievement and educational aspiration. Women who were passive, who sought vicarious satisfaction of their achievement needs, were more likely to express low educational aspiration, whereas balanced and direct achievers had higher educational aspirations. The relation between sex-role ideology and educational aspiration changed somewhat for vicarious achievers. Among those who held the contemporary view of female roles and were also vicarious achievers, educational aspiration was reduced so that they could not be distinguished from women in the traditional group in terms of expectations for doctoral or postdoctoral studies. For balanced and direct achievers there was an even stronger relation between sex-role ideology and educational aspiration, with the contemporary-ideology women 19 percent more likely to have plans for doctoral studies than the traditional-



MODE OF ACHIEVEMENT and educational aspirations were closely linked. Vicarious achievers tended to have lower educational goals than did the balanced or direct achievers.

© 1971 SCIENTIFIC AMERICAN, INC

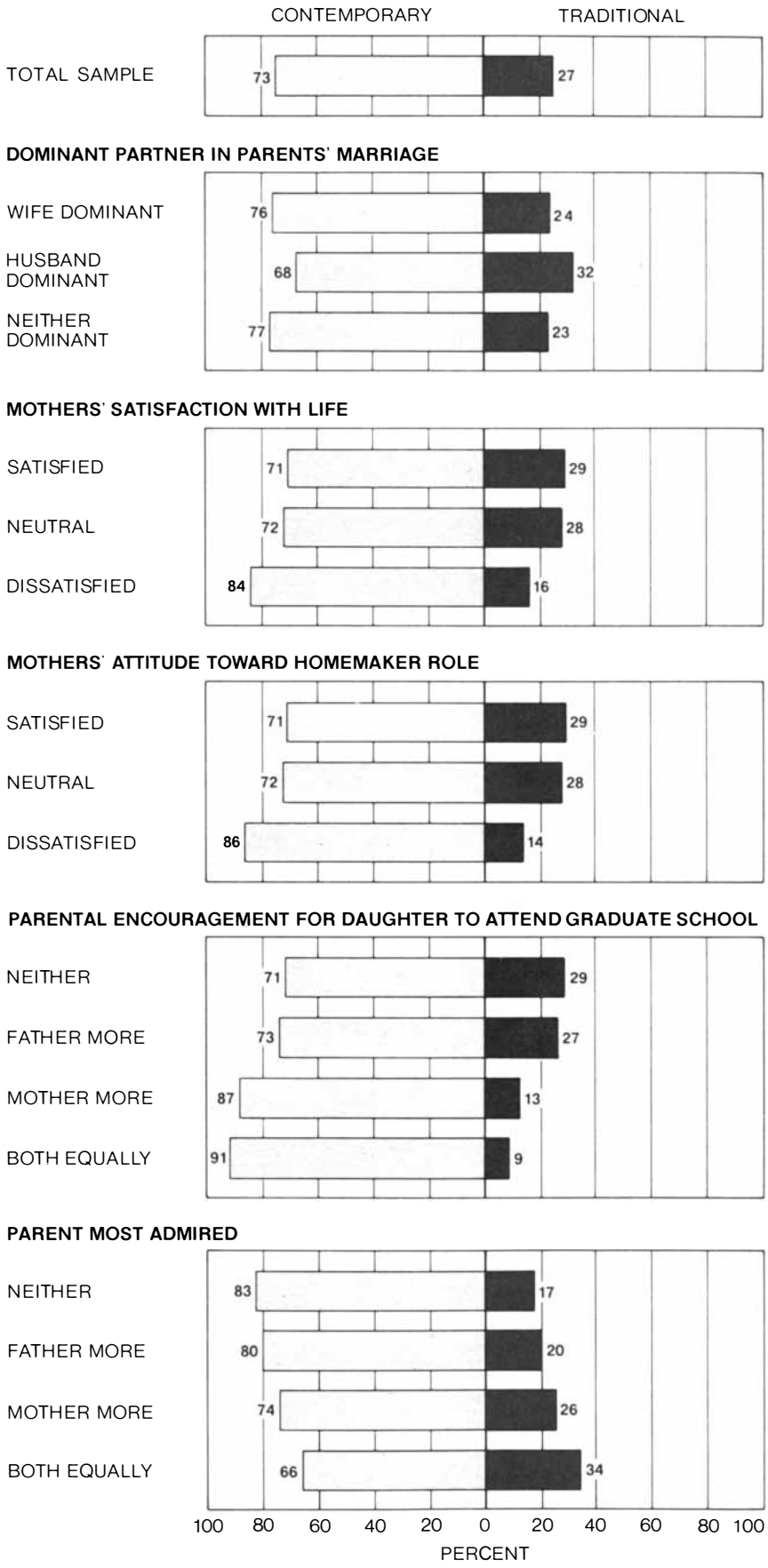
ideology women, who strongly tended not to have plans for further education.

In short, the relation between female-role ideology and educational aspiration is contingent on the way a woman has been socialized to meet her achievement needs outside the home. It is quite clear that the vicarious mode and the traditional view are linked and tend to predispose a woman to limit her educational goals. The balanced and direct modes of achievement are linked to the contemporary view of sex roles and tend to encourage high educational aspirations.

Since it appears that women are socialized at an early age to a female-role ideology, let us examine the family background of the respondents. Obvious socioeconomic indexes, such as parents' income, education or occupation, surprisingly had no bearing on the daughter's sex-role ideology. In fact, none of the usual socioeconomic characteristics of the family was related to female-role ideology.* Women with the contemporary ideology were just as likely to come from homes with incomes of less than $3,500 as those with the traditional view were. Parents with a college education, including a doctoral degree, were just as likely to produce daughters with the traditional sex-role ideology as daughters with the contemporary one.

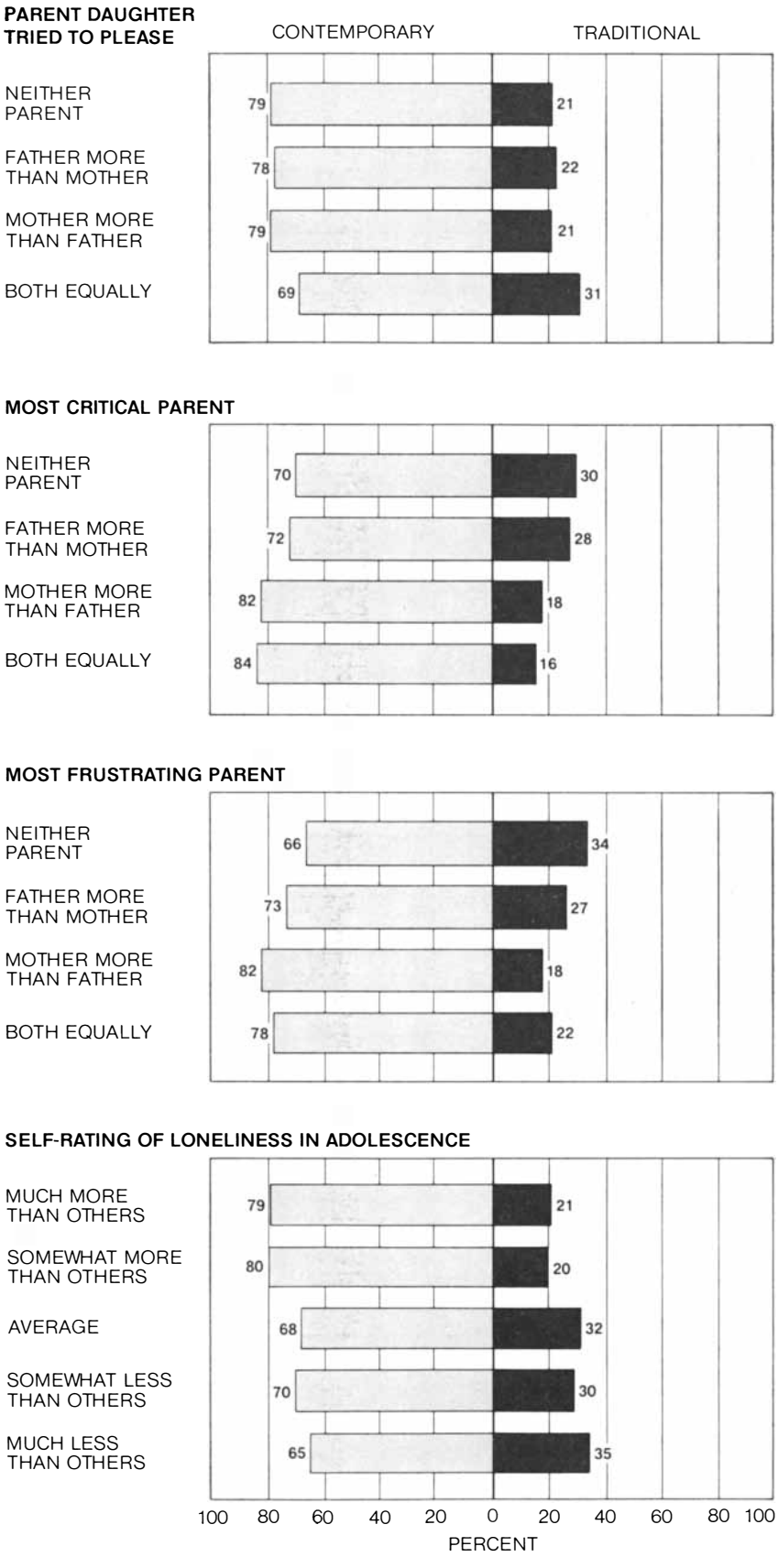
Another usually important factor, childhood religion, failed to show a statistically significant influence on sex-role ideology, although a Catholic upbringing tended to produce slightly more women with the traditional viewpoint than a Protestant or a Jewish upbringing did. And in spite of the common opinion that city dwellers are less bound by traditional attitudes than people raised in rural areas, women who came from rural homes were as likely to hold contemporary sex-role attitudes as women from urban and suburban areas were.

It is worth noting that homes disrupted by divorce, separation or death did not differ from intact homes in the proportion of traditional-ideology and contemporary-ideology women they produced. This holds true regardless of the age of the daughter when the home was disrupted. Moreover, having sisters or brothers, either younger or older, had no effect on sex-role ideology. Working mothers have been both praised and damned for the effects of their outside commitment on their children. In our sample whether or not a mother worked had no perceptible impact on her daughter's sex-role ideology. Nonworking mothers were just as likely as employed



PARENTAL INFLUENCES on daughter's sex-role ideology (*gray and black bars*) were not expected variables such as income, education or religion but rather qualitative factors such as dominant parent, mothers' attitudes, encouraging parent and most admired parent.

© 1971 SCIENTIFIC AMERICAN, INC



OTHER PARENTAL FACTORS that influenced the daughter's sex-role ideology were parent daughter tried to please most, most critical parent and most frustrating parent. Women with a contemporary view tended to be lonelier than their peers in adolescence.

mothers to have daughters with the contemporary view of the female role.

In the light of these negative demographic findings, do childhood family variables contribute anything to the ideology of college women? If there is such an influence, we must seek the answer in more subtle factors such as the marital relationship between the parents. Women with a contemporary sex-role ideology tended to come from families in which most of the time neither parent was dominant or from families in which the mother was dominant. When the father was more dominant in the marriage, the tendency of the daughter to develop the traditional concept of the female role was enhanced.

An important predictor of a woman's female-role ideology is her perception of her mother's overall satisfaction with life. Dissatisfied mothers were more likely than satisfied ones to rear daughters with the contemporary view of the role of women, and satisfied mothers were more likely to have daughters with the traditional orientation. In traditional homes where the father was dominant, dissatisfied mothers were even more likely to raise daughters with the contemporary ideology. In homes where neither parent was dominant, dissatisfied mothers were not significantly more likely than satisfied ones to raise daughters with the contemporary view of female roles. Interestingly enough, daughters from contemporary homes who felt that their mothers were dissatisfied with their life did not turn to the traditional sex-role ideology as a means of avoiding dissatisfaction with their own life.

The homemaking responsibilities of a woman are perhaps among the most traditional features of married life. Does a mother's attitude toward homemaking influence her daughter's sex-role attitudes? The answer is yes. Mothers who were dissatisfied with homemaking had a greater tendency than satisfied homemakers to raise daughters with the contemporary view [*see illustration on preceding page*]. Daughters of mothers who were satisfied with household tasks or had a neutral attitude toward them had the greatest likelihood of holding the traditional view.

Were mothers who were dissatisfied with household work less satisfied with their life? Again the answer is a definite yes. Half of the mothers who were dissatisfied with homemaking tasks were dissatisfied in general with their life, but only 7 percent of mothers satisfied with homemaking were dissatisfied

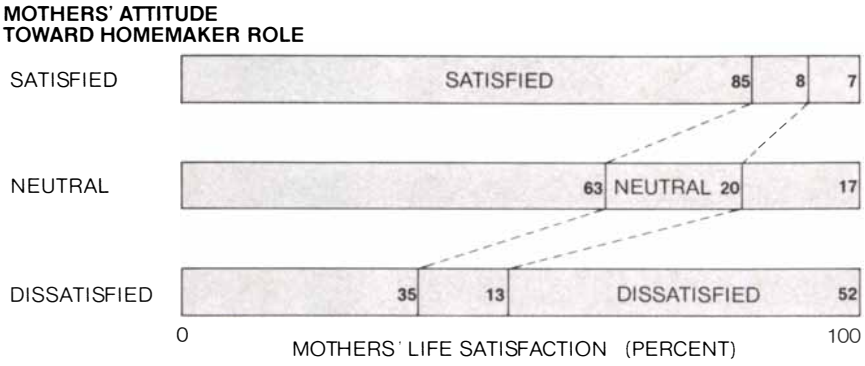
© 1971 SCIENTIFIC AMERICAN, INC

with their life. In addition, 85 percent of the mothers who enjoyed their homemaking role were satisfied with their life. Only 35 percent of the mothers who rejected homemaking found general satisfaction in their life. It may be that a mother who does not find household tasks satisfying and whose overall life satisfaction is low may encourage her daughter to seek a different life pattern. Higher education, particularly graduate school, represents a different way for a woman to allocate her time and energy. Is there, then, a connection between a mother's encouraging her daughter to go to graduate school and the daughter's concept of the female role? Apparently there is. A mother's encouragement enhanced the tendency toward the contemporary viewpoint more than a father's encouragement alone. When both parents equally encouraged their daughter to go to graduate school, however, the daughter was most likely to have the contemporary sex-role ideology.

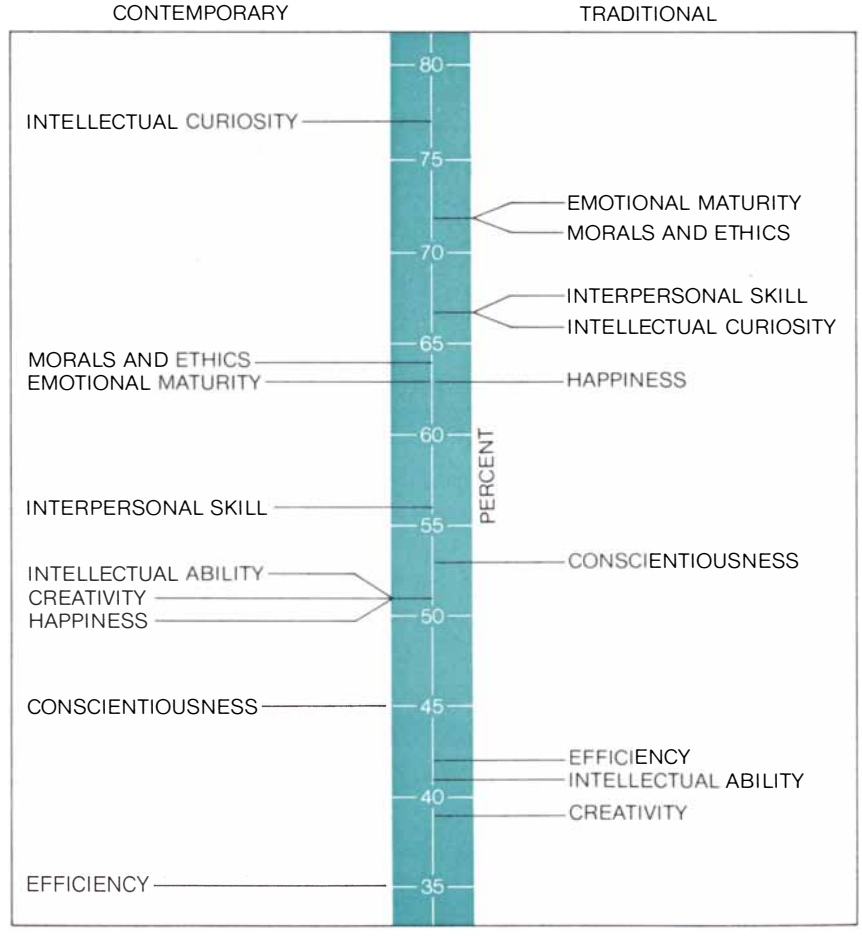
Although sex-role ideology may be developed in early childhood, it is usually not until adolescence that a girl begins to apply her system of beliefs to her life pattern. At this stage her interaction with her parents presumably would help her to test and adjust the viability of the attitudes and beliefs that had been guiding her since childhood. Do daughters with the contemporary ideology have a pattern of development in adolescence that is different from the pattern of daughters with the traditional view of the female role? In a general way it appears to be so. Women who emerge with the contemporary ideology tend in adolescence to achieve a certain psychological distance from their family, to evolve a sense of separateness as individuals.

Admiration for a parent may be regarded as an index of willingness to remain an integral part of the family. In describing their relationships with their parents during adolescence, women with the contemporary ideology were the most likely to have rejected both parents as objects of admiration: 83 percent of the women who reported that they admired neither parent adhered to the contemporary female-role ideology. There was also a slight tendency for women who admired their fathers more than their mothers to hold the contemporary view, whereas women who admired their mothers more tended to favor traditional sex-role attitudes.

Efforts to please a parent provide another clue to a daughter's willingness to maintain a close sense of identification with that parent. Women who reported that they did not try to please either parent, who sought to keep their distance, were more likely to hold the contemporary view than women who tried to please both parents [*see illustration on opposite page*]. The amount of criticism from a parent is another indicator of the kind of rapport between daughter and parent. Of the women who reported that they were constantly criticized by both parents, 84 percent held the contemporary view. Women with the contemporary viewpoint tended to have a critical mother, whereas women with the traditional viewpoint recalled having a critical father. Women with the contemporary ideology were frustrated most by their mother; those with the

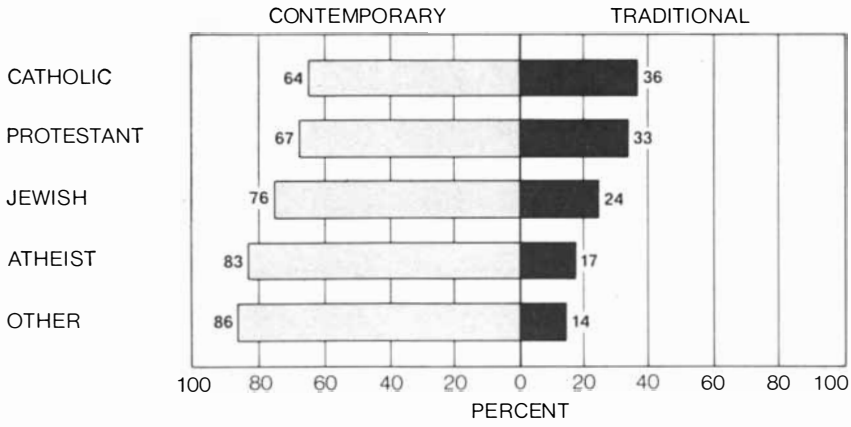


MOTHERS' SATISFACTION with life and homemaking were clearly related. Mothers perceived by daughters as satisfied homemakers were likely to be perceived as satisfied with life. Dissatisfied homemakers tended to be seen as being dissatisfied with life.

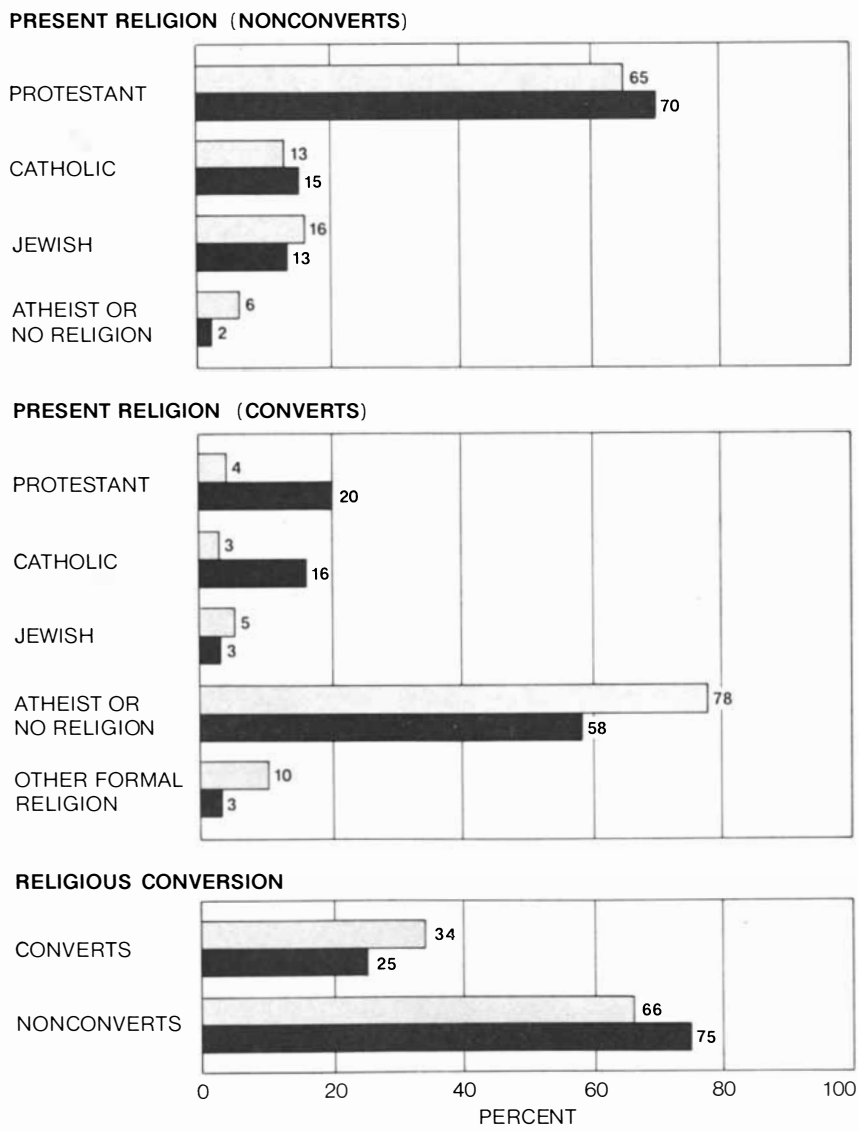


HIERARCHY OF VALUES of women with contemporary sex-role ideology differed from that of women with traditional views. Ranking is based on values rated as most important.

© 1971 SCIENTIFIC AMERICAN, INC



PRESENT RELIGION (*above*) was related to sex-role ideology but childhood religion was not. Women with non-Christian beliefs favored the contemporary view of sex roles.



RELIGIOUS CONVERSION was a linking factor between childhood religion, present religion and sex-role ideology. For women who maintained their childhood religion (*top graph*) there was no statistically significant link between religion and sex-role ideology. For converts (*middle graph*) there was a strong association between religious views and sex-role views: converts with contemporary view (*gray*) moved away from Christianity more than converts with traditional (*black*) view of sex roles. Women who held a contemporary view were more likely to have converted from their childhood religion (*bottom graph*).

traditional viewpoint were more likely to say that neither parent frustrated them.

To feel that one's parent has been successful in life is to recognize and approve of the parent's overall life pattern. Adolescents who regard their parents as being successful are better able to accept their parents as role models and to pattern their own lives in a similar fashion. Rejection of the parental life pattern may force the adolescent to seek new approaches to the conduct of life. This is borne out in the data from the study: women with the contemporary ideology tended to regard both parents as being unsuccessful, and those with the traditional viewpoint tended to see both parents as being successful. The desire of an admiring daughter to please her parents and to tolerate their criticism and her inclination not to regard parental demands as being frustrating are consistent with the daughter's acceptance of her parents' life-style. It is not surprising that such daughters tend to see both parents as being successful.

Approval of parents by an adolescent may be important to family living in traditional terms, but a relatively unquestioning attitude toward familial values and life-styles may make it more difficult for a woman to break through to an alternative life-style as an adult. Reluctance to break with familiar patterns and values may have implications for a woman's ability to adapt to new and perhaps enriching experiences. Although the traditional viewpoint may ensure continuity and security, it may also lead in later life to dissatisfaction, particularly where a changing external environment may require flexibility in adapting to new conditions.

The reluctance to please parents, coupled with a decreased admiration for parents and a tendency to see them as frustrating and critical, are closely associated with the contemporary sex-role ideology. They may be the first steps in a woman's developing a sense of distance from the family in which she grew up. Gaining a sense of oneself as an individual is a necessary step toward enlarging one's sense of self-fulfillment. In a rapidly changing environment self-fulfillment may take on a new meaning that requires a new life-style for women, including new roles such as student and worker as well as the customary ones of wife and mother.

All psychological growth has certain costs, and it is relevant to ask at what price a contemporary ideology has been

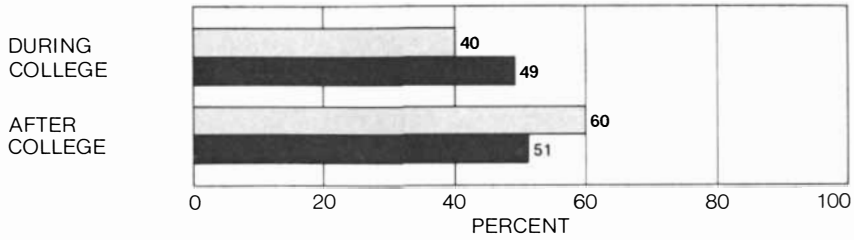
© 1971 SCIENTIFIC AMERICAN, INC

acquired. A sense of loneliness in adolescence may be considered one kind of short-term psychological price paid for developing new patterns of behavior. Women who held contemporary sex-role beliefs reported that they were lonelier than their peers in adolescence; women with the traditional viewpoint were less likely to be lonely than their peers. If the short-term effects of the contemporary sex-role ideology are painful, what about the long-term consequences? It appears that adult women with the contemporary sex-role ideology are just as happy with their present life as are women with the traditional viewpoint. Furthermore, there does not appear to be any difference in the anxiety levels of the two groups of women.

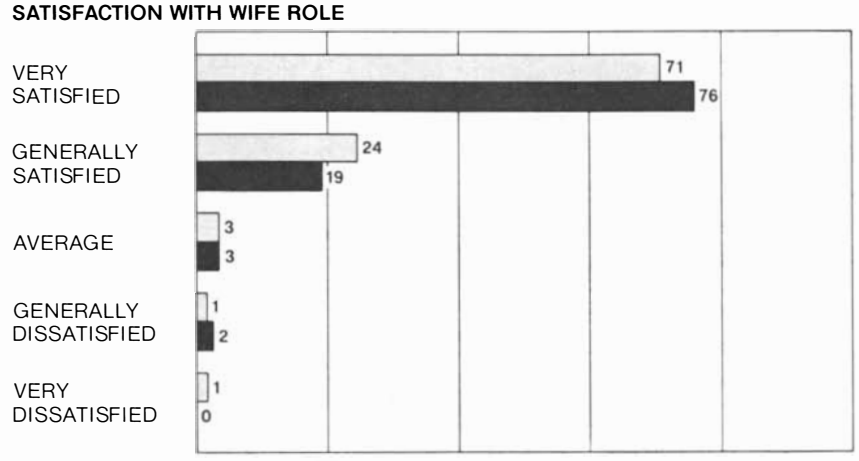
Differences in sex-role ideology might be expected to produce differences in value systems, and the hierarchy of values of the women in the contemporary category indeed differed from that of the women in the traditional group. The contemporary position placed the highest value on intellectual curiosity; the traditional one put emotional maturity, morals and ethics first [*see bottom illustration on page 39*]. Moreover, women with the traditional set of beliefs attached special meaning to interpersonal skills, happiness and conscientiousness.

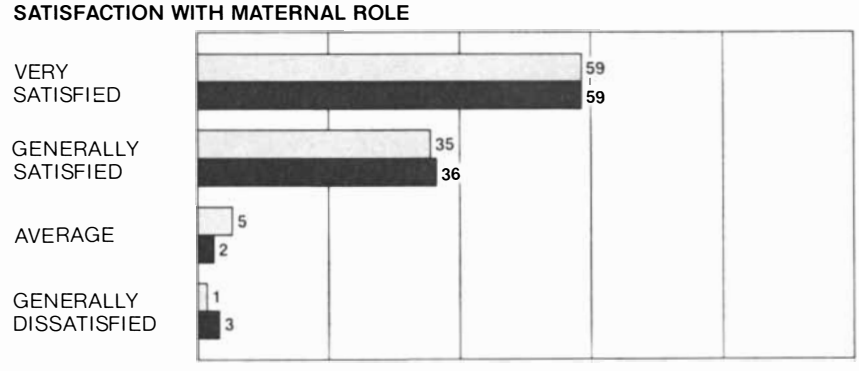
For analytical purposes I have been stressing the striking contrasts between women with a contemporary sex-role ideology and women with a traditional ideology. It is equally enlightening, however, to look at the values these usually divergent groups share. Both rated honesty and understanding people as crucial qualities, and both depreciated perseverance, the ability to work under pressure, ambition, competitiveness, physical stamina and realism. Somewhat surprisingly, both groups gave low ratings to the qualities of self-confidence, enthusiasm, courage, physical attractiveness and sexuality.

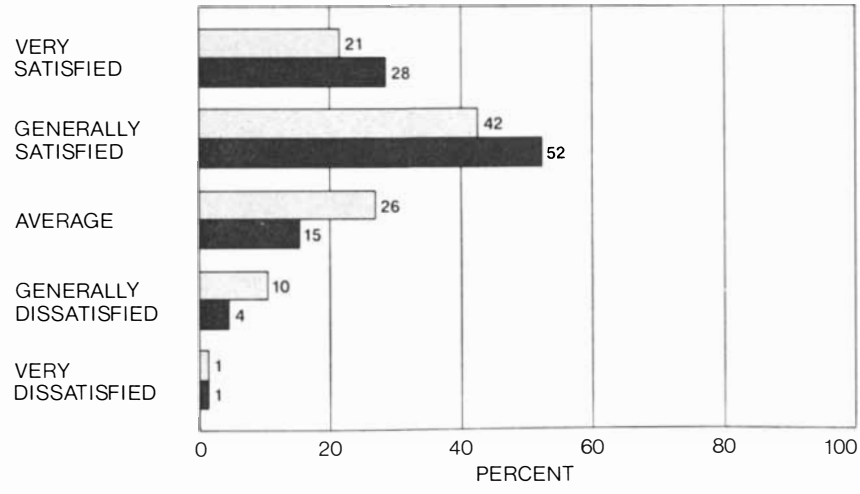
I have noted that childhood religion was unexpectedly found to be unrelated to present sex-role ideology. In view of this finding it was surprising to discover that there was a strong association between the present religious affiliation of a woman and her attitude toward sex roles. Women who espoused atheism, who had no formal religion or who professed Judaism or Eastern religions clearly tended to favor the contemporary sex-role ideology. Adult Protestants and Catholics were more likely to hold the traditional viewpoint. This is all the more puzzling because there was also



INFLUENCE OF SEX-ROLE IDEOLOGY on time of marriage was significant, with women holding a contemporary view more likely to complete their college studies before marrying.

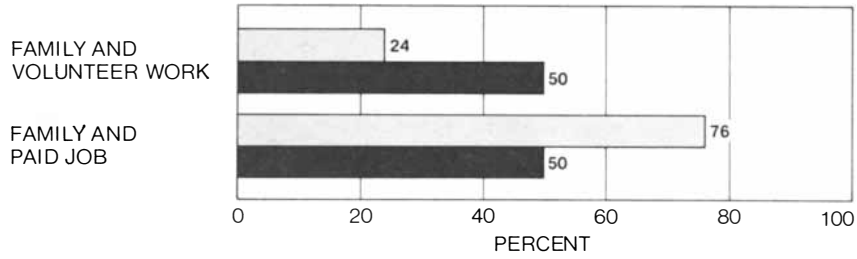






SIMILAR SATISFACTION profiles of women with contemporary and traditional views were found regarding wife and mother roles (*top and middle graphs*). Women with contemporary views, however, expressed less satisfaction about homemaking responsibilities.

© 1971 SCIENTIFIC AMERICAN, INC



IDEAL LIFE ROLES differed for women with contemporary and traditional sex-role attitudes. Women with contemporary view (*gray*) favored family and paid employment, whereas women with traditional view (*black*) split equally between family and employment.

a distinct association between childhood religion and present religion.

The key to the puzzle appears to be religious conversion. For women who had maintained the religion they had had as children there was no link between sex-role ideology and present religion. For those women who had shifted their religious affiliation, however, there was a clear and most interesting connection between attitudes toward the female role and present religion [*see bottom illustration on page 40*].

In our analysis of the adolescent relationships these women had had with their parents we noted that as adolescents women with the contemporary viewpoint were in the process of disengaging themselves from the family patterns and values with which they had grown up. Women in the traditional category seemed more willing to accept the frame of reference provided by their parents and to continue to live within it. If this interpretation is correct, there should be some indirect confirmation of it in terms of rejection or continuance of childhood religion. More explicitly, rejection of childhood religion might serve as an index of the degree to which a woman has disengaged herself from her parents and family traditions. Women with the contemporary ideology therefore should be more likely than those with the traditional viewpoint to experience religious conversion. Data from the survey confirm this hypothesis: more than a third of the women with the contemporary ideology had moved away from their childhood religion, compared with only a fourth of the women in the traditional group. Women with the contemporary position who had rejected their childhood religion were more likely than those with the traditional view to turn to atheism or to have no formal religion; they were also considerably more likely than women with the traditional position to take up Eastern religions (in which, we might speculate, they sought for new meaning in their life). Women in the traditional category who had converted were much more likely than women with the contemporary viewpoint to move toward the familiar religions of Protestantism and Catholicism.

It was interesting to examine whether or not women who attached great importance to intellectual qualities tended to act on these values in their daily life. One indicator is to see if women with the contemporary ideology were more likely than those with the traditional viewpoint to postpone marriage until the completion of their college studies, an action that would be consistent with placing a high value on intellectual activity. A look at the educational level of the respondents at the time of their marriage immediately brings out the fact that women with the contemporary ideology were indeed more likely than those with the traditional viewpoint to marry after the completion of their college studies. Women in the traditional group, who did not set such a high value on intellectual qualities, tended to marry while they were still in college.

It would appear that sex-role ideology is an important factor in predicting values. Does it also influence the degree of satisfaction women derive from informal activities with other women? Do women with the contemporary sex-role ideology derive less enjoyment from casual interaction with other women? This turns out to be the case. Women with the contemporary viewpoint said that they enjoyed activities such as chatting and card-playing much less than other women of their age; women in the traditional group tended to rate themselves as being average in this respect.

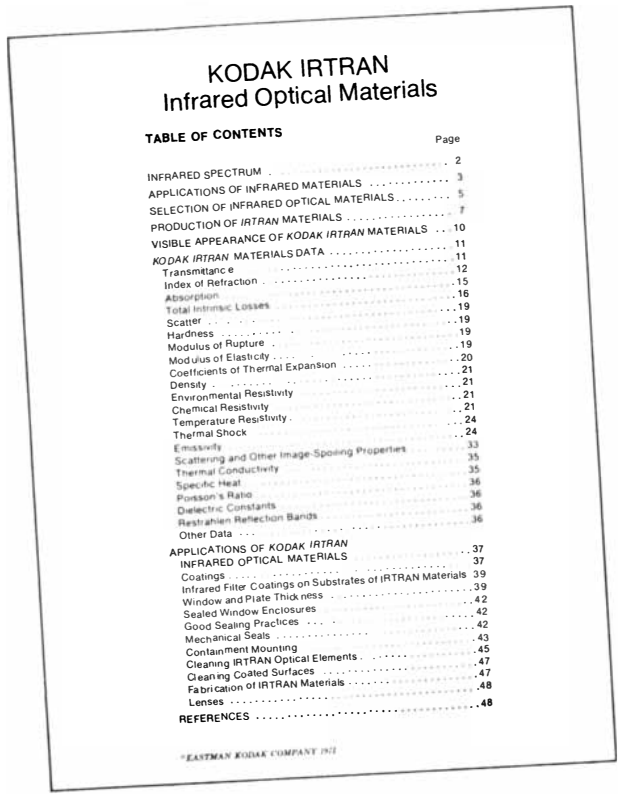
If these common female activities outside the home are less than exciting to women with the contemporary ideology, what about the three major roles of a married woman within the family: her conjugal role, her maternal role and her homemaker role? Women who saw the relationship between the sexes in contemporary terms were just as likely to express satisfaction with their husband as women who took the more traditional view of marriage. Moreover, among the women in the sample who had children, mothers with the contemporary ideology were just as satisfied with their maternal role as were mothers in the traditional group. When it came to satisfaction with the role of homemaker, a different picture emerged. Women with the contemporary viewpoint were noticeably less enthusiastic about homemaking. Contrary to television advertisements for detergents and other household products, not all women find delight in cleaning house or washing clothes. Women with the contemporary position were not reluctant to convey their dislike for these household tasks. On the other hand, activities that allow more self-expression or creativity, such as cooking, entertaining, interior decorating, sewing and shopping, were equally acceptable to women in both categories.

The two groups of women have distinctly different concepts of the ideal life. For the woman in the traditional category the ideal life consists in devoting her energies to family and volunteer social activities. Fifty percent of the women in this group reported that their choice for the most satisfying way of life involves family and volunteer projects; only 24 percent of the women in the contemporary group gave this response. More than three-quarters of the women in the contemporary category said that the most satisfactory life-style is a combination of family and full-time or part-time employment.

Although it is clear that ideology affects life goals and choices, does it also produce differences in the self-concepts of women? Self-esteem, the acceptance of oneself as an individual, is an important component of one's self-image. How do the two groups in our survey compare in self-esteem? Curiously, the women in both groups have remarkably similar self-esteem profiles. Women in both categories express equal confidence in their competence as wife and mother. Their sense of competence regarding their ability as student, employee and community participant is also similar. Wives with the traditional viewpoint rate their ability to solve complex problems just as favorably as contemporary wives. The traditional and contemporary sex-role viewpoints lead to two distinct life patterns, but within each ideological position women are able to find fulfillment and meaning in their life.

© 1971 SCIENTIFIC AMERICAN, INC

We want to be useful ...and even interesting



KODAK IRTRAN
Infrared Optical Materials

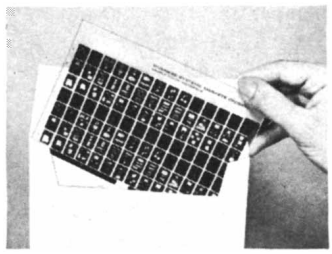
TABLE OF CONTENTS

©EASTMAN KODAK COMPANY 1971

A publication as specialized as this ought to be on microfiche

In this form updating would be a lot cheaper. So the content would be kept fresher.

The information would reach more people who could benefit from it but who are not going to send for the book. They aren't sure yet that they will need it. They *are* sure that they are running out of shelf and file space.



Microfiche is read and consulted like this ▼ Do you have such a device at home? In your office? No? Too bad.

Hunting one up and going there could be a bit much to ask. Which is why the above-cited book comes printed on 8½″ x 11″ paper pages and contributes nothing to overcoming either your paper glut or the high cost of publishing detailed technical information.

Let's break this circle. Users of scientific and technical information can come to consider a microfiche reader an essential item of home and office furniture.

Technical publishing outfits eager to speed the day should keep in touch with Dave Kithil, Kodak, Rochester, N.Y. 14650.

Kodak

New badge with an old name

Detailed, friendly explanations for readers new to the subject, or not so new

Many tables, many numbers, many graphs

Years in the writing. Not by a committee, but by one man who has spent much of his business life discussing infrared-transmitting optical materials with customers ranging up to the most technically sophisticated ones.

If this table of contents looks pertinent to your work, let us send you a free copy of the book. *Write Kodak Apparatus Division, Special Products, Rochester, N.Y. 14650 and ask for "U-72."* If you want to talk to the man himself, or even to give him a rush order, try 716-325-2000, ext. 62170.

but the printing industry will still be with us.

Microfiche has its place, but printer's ink will flow on and on.

The more people who depend on the printing industry for a living and the more they know, the better for Kodak. The intricate complex of businesses and crafts centered on the art of printing and packaging is more than a principal market for specialized Kodak products. One way or another, it provides a life role for a not inconsiderable segment of mankind. Finding a role in life does seem to be a common problem.

Perhaps we should try to attract more kids to printing. (The more people who look to our customer industries for prosperity, the better for us.)

Perhaps we shouldn't. Recruiting should run on something more dependable than enthusiasm. Distress now afflicts the "overeducated." Strongly held still—perhaps rightly so—is a concept of education that comes down from a time when college prepared men for the ministry, medicine, law, or teaching. We are painfully aware that a lot of the teaching in preparation for less classical professions is out of date.

Even if we wanted to, we couldn't mount a campaign powerful enough to lure large numbers of kids into printing and the graphic arts, but we have collaborated with new-style academics, the printing and allied industries, and their unions in a measurement just completed of 1) actual manpower needs in these fields, 2) how changes in technology promise to affect the needs.

Interest in the findings should be made known to W. F. Flack, Kodak, Rochester, N.Y. 14650.

© 1971 SCIENTIFIC AMERICAN, INC

Yes,
I'd like
to read
why
Science &
Mechanics
Magazine
chose the
Renault 12
"Small
Car of the
Year."

Renault, Inc.,
P.O. Box 1972, Long Island City, N Y 11101.
Gentlemen: Please send me the Science & Mechanics Magazine article.

Mr. Mrs. Miss

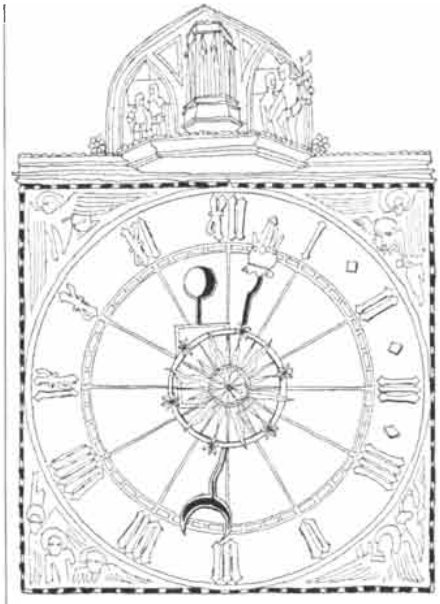
Address

City State Zip

RENAULT

World's largest producer of front-wheel drive cars.

SC-1



SCIENCE AND

Merchants of Death

The recently concluded war between Pakistan and India points up one of the most striking military facts of the postwar period. Every war of the past 25 years has been fought in the underdeveloped, or "third world," countries with weapons supplied almost entirely by the developed countries, principally by the two "great powers." Whereas the U.S. and the U.S.S.R. have avoided any direct confrontation that could lead to a general nuclear war, they have by their massive supplies of arms to the third world exacerbated many local conflicts that could escalate into such a nuclear war.

Over the past 20 years the flow of major weapons—planes, tanks, ships and missiles—to third-world countries has increased dramatically. Between 1950 and 1970 the total annual value of these exports (at constant prices) increased sevenfold, so that by 1970 it amounted to more than $1.5 billion. This growth is equivalent to an average annual rate of increase of 9 percent, which is nearly twice the average annual rate of increase of the gross national product of the third-world countries.

These and other facts and figures detailing the political, economic and military dimensions of the buildup of large arsenals of increasingly sophisticated weapons in the underdeveloped countries are presented in a 910-page volume titled *The Arms Trade with the Third World,* just published by the Stockholm International Peace Research Institute (SIPRI). The study, prepared over nearly four years by a team of SIPRI researchers headed by the British economist Frank Blackaby, presents the statistics for the movements of major weapons to each of the third-world countries, year by year for the past 20 years, and analyzes them in depth against the background of the political and economic history of the postwar period.

Because the trade in arms is usually not reflected in public trade statistics (it is either concealed or not reported) SIPRI had to construct its own arms-trade statistics. The method used was to build up country registers for some 91 third-world countries and enter on them all information, derived from a wide variety of sources, about arms deals in major weapons. In all, some 4,000 such arms deals were identified.

One reason given by SIPRI for taking up the study was that the international trade in arms, which received a great deal of attention in the years between World War I and World War II, had been largely neglected in the postwar period. In contrast to the interwar period, when the buildup of armaments was often attributed to the greed of private manufacturers, today the export of weapons is controlled by governments. Hence, according to SIPRI, "a study of the arms trade is crucial not merely because this is the route by which the stocks of weapons in the third world are built up; it is also one means by which the competition between rich countries is extended throughout the world and absorbs local disputes and wars."

One of the main features of the international arms trade demonstrated by the SIPRI study is the dominance of a few supplying countries. Of the supply of major weapons to third-world countries during the period 1950–1969 an average of 87 percent came from four countries: the U.S., the U.S.S.R., Britain and France. In the past few years this proportion has increased to more than 90 percent. Although the U.S.S.R. entered the market as an arms supplier to the developing countries relatively late, today the U.S.S.R. is the world's second-largest arms supplier, following closely behind the U.S. Together these two countries account for about two-thirds of the total arms deliveries.

On tabulating the main recipients of

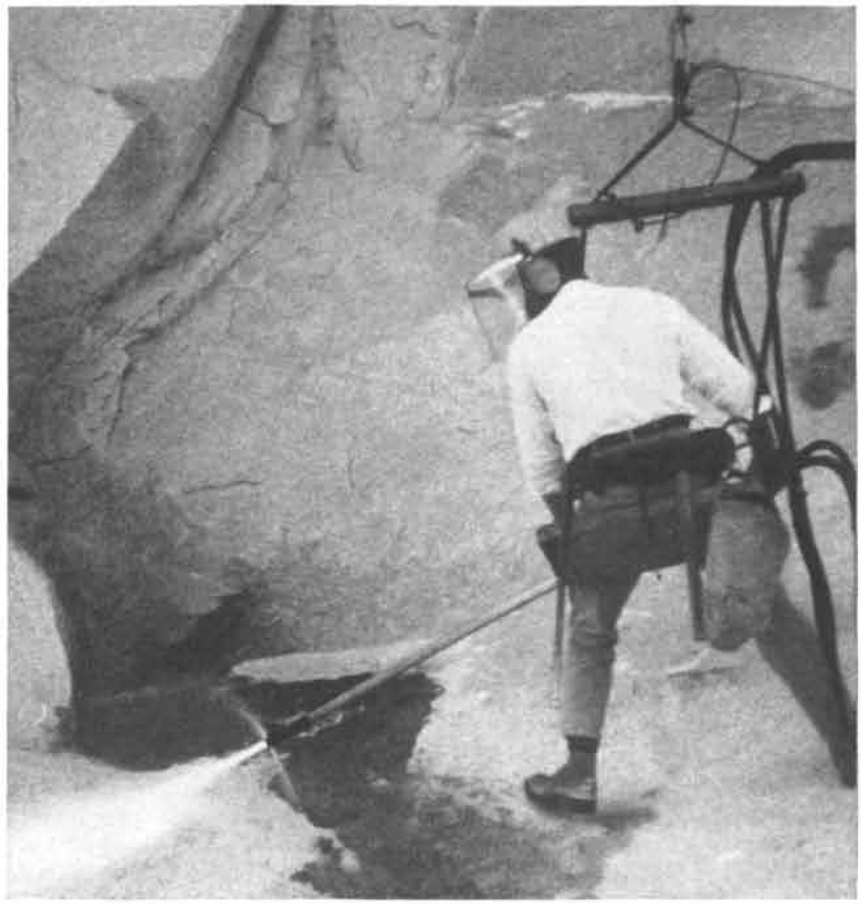
© 1971 SCIENTIFIC AMERICAN, INC

these weapons the SIPRI team found that 70 percent of all major-weapons deliveries have gone to the Middle East and Asia. The Middle East has accounted for nearly 25 percent of the total major weapons supplied since 1950, and the Far East, including Vietnam, accounted for a further 30 percent. The Indian subcontinent has absorbed about 15 percent of the total. Africa has accounted for 8 percent, although this share is increasing. In recent years the increase in major-weapons imports has been particularly rapid in the countries of the Persian Gulf.

In analyzing proposals to regulate the arms trade the SIPRI study concludes that it is doubtful whether such proposals are the most appropriate instrument for limiting the buildup of stocks of weapons in third-world countries—"mainly because such proposals inevitably discriminate against nonproducing countries. Any proposal for international registration would therefore certainly have to cover production as well as trade, if it were to have any chance of acceptance by third-world countries. The most fruitful approach to actual regulation or limitation of weapons would seem to be through regional recipient-country agreements which would cover production as well as trade. In the regions where the competing military commitments of the great powers are important, there is no prospect of any such agreement's being reached until there is more progress in settlement between the United States and the Soviet Union. It is in regions where the links with the major arms race are weakest that the possibility of agreement is greatest—for instance, Latin America and sub-Saharan Africa. One of the important characteristics of the arms trade is that, along with other factors, it links arms races in the third world to the major East-West confrontations; one object of any proposal should be to weaken this link and reduce the military dependence of the nonproducing countries."

Food and Population

The largest harvest in recent years appears to have been reaped from the world's farms in 1971, according to preliminary reports reaching the Food and



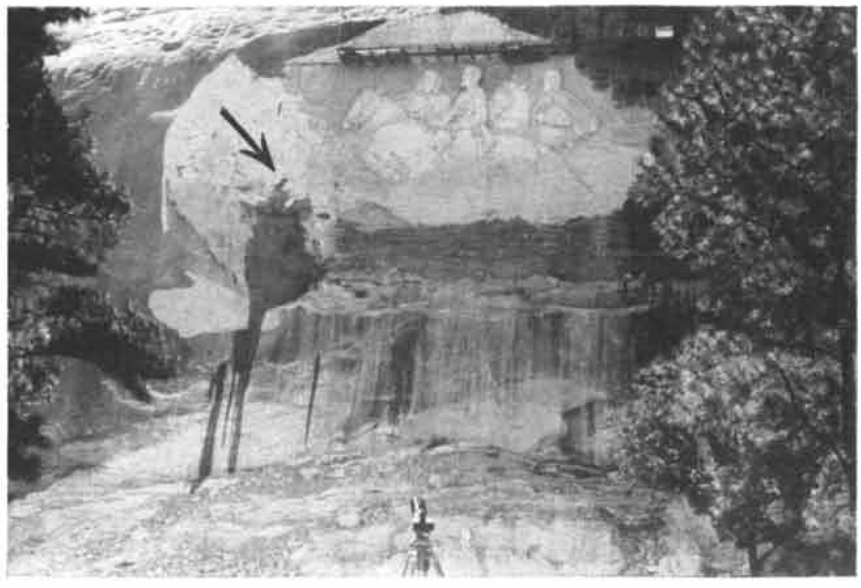
Another of those "impossible" views with a Questar

At historic Stone Mountain near Atlanta, a carver works on the Confederate Memorial. (See covering shot below taken with 50mm. camera lens, showing Questar Field Model mounted on Linhof tripod, Topcon camera attached, and worker indicated by black arrow.) The Questar photo, above, on Tri-X, 1/60 second, shows him in constant motion as he swings from a harness on the sheer granite wall, guiding himself with his feet and vibrating with the thrust of his powerful thermo jet torch. The photoengraving process permitting, you will notice such detail as his belt loops and a band around his left ankle. Photographs by Mr. and Mrs. Ralph Davis of Sarasota.

QUESTAR, THE WORLD'S FINEST, MOST VERSATILE TELESCOPE, IS PRICED FROM $865. OUR BOOKLET CONTAINS MORE THAN 100 PHOTOGRAPHS BY QUESTAR OWNERS. TO MAIL ANYWHERE IN N. AMERICA, $1.00. BY AIR TO REST OF WESTERN HEMISHERE, $2.50; EUROPE AND N. AFRICA, $3.00; ELSEWHERE $3.50

QUESTAR

BOX 20, NEW HOPE, PA. 18938



© 1971 SCIENTIFIC AMERICAN, INC

Agriculture Organization of the United Nations. Eric Ojala, the assistant director general, told a conference of the FAO in Rome that the increase over 1970 would be about 4 percent. In 1970 the harvest rose 2 percent over 1969, and in 1969 there was no gain over the preceding year.

In developing countries the rise was between 2 and 3 percent from 1970 to 1971, Ojala said. At this rate the production of food about kept pace with the rise of population in those countries. Most of the overall increase of 4 percent is attributable to a surge of 7 to 8 percent in the developed market-economy countries, according to Ojala. In eastern Europe and the U.S.S.R., he said, the gain in 1971 was 1 or 2 percent, compared with 6 percent in 1970. Discussing the "green revolution," Ojala cautioned that "the fabric of this progress in the developing world as a whole remains fragile and its base narrow, both geographically and in terms of commodities."

The Sirens of Titan

The exploration of planets beyond Mars will begin in late February or early March with the launching of *Pioneer 10* on a voyage to the vicinity of Jupiter, a trip that will take at least 19 months and cover more than 600 million miles. This will be the opening shot in a new chapter of space exploration that should provide a rich flow of information about the five outer planets: Jupiter, Saturn, Uranus, Neptune and Pluto. An outline of the opportunities in planetary exploration and a list of recommendations for the years immediately ahead have recently been published by the National Academy of Sciences in a report called *Outer Planets Exploration 1972–1985*.

Emphasizing that "the major part of the sun's planetary system lies beyond the orbit of Mars," a study group of the Space Science Board under the direction of Francis S. Johnson of the University of Texas describes the scope of effort that could be supported by three levels of expenditure: a high-budget program of $400 million per year, a more modest program of $250 million per year and a minimum program of about $40 million per year. The first two budgets, but not the third, would finance "grand tours" of the outer planets in the late 1970's, made possible by a favorable planetary alignment enabling a spacecraft to receive a gravity assist from Jupiter and thus reach the planets beyond with less expenditure of energy and in less time than would otherwise be possible. Even so at least eight years will be needed to reach Pluto. The favorable planetary alignment will not come again for about 180 years.

The spacecraft proposed for the grand-tour missions is designated TOPS, for thermoelectric outer-planet spacecraft. Beyond Mars sunlight is too weak to serve as a source of power, hence the craft will require a thermoelectric generator of about 400 watts. TOPS will weigh 1,400 pounds, carry 205 pounds of scientific instruments and be able to transmit data at an initial rate of 100,000 bits per second. *Mariner 9*, now in orbit around Mars, weighs 1,200 pounds, has an instrument payload of 152 pounds and can transmit data at a maximum rate of 16,200 bits per second. Perhaps the most unusual feature of the grand-tour spacecraft will be a computer called STAR ("self-test and repair"), provided with enough redundancy to operate for at least 10 years. The TOPS spacecraft would be launched by a seven-segment Titan vehicle.

According to the Space Science Board report, the minimum goal of the grand-tour missions should be to measure the density, composition and temperature of the atmospheres of the outer planets; the magnetic fields of the planets and the kinds of particle trapped in the planetary radiation belts; the variation in the properties of the solar wind with distance and how the wind interacts with each of the planets; the amount of interstellar hydrogen, and the intensity and composition of galactic cosmic rays in the region where the solar system merges into interstellar space. The grand-tour vehicles should also be equipped to take closeup pictures of the planets and their major satellites. Two of Jupiter's satellites (Ganymede and Callisto) are about the same size as the planet Mercury; two others (Io and Europa) are about the size of our moon. Titan (a satellite of Saturn) and Triton (a satellite of Neptune) are both larger than the moon but a little smaller than Mercury. The Space Science Board urges that as much attention be given to the satellites as to the planets themselves.

Pioneer 10 represents a smaller and less expensive spacecraft system well suited to making studies of Jupiter and Saturn. The Pioneers, well-tested, spin-stabilized craft, have been steadily upgraded in size and performance; the latest models weigh 560 pounds and carry 65 pounds of instruments. *Pioneer 6*, launched in 1965, is still operating. The Space Science Board recommends that a Pioneer be launched toward either Jupiter or Saturn at least once every two years during the next decade, a rate that could be supported even with the lowest budget. Three kinds of mission are proposed: simple planetary flybys, orbiting missions and atmospheric-penetration missions. The board observes that entry probes, equipped with heat shields, should be able to make measurements in the Jovian atmosphere until they have descended to the point where the pressure approaches, or possibly exceeds, 100 times the pressure of the earth's atmosphere at sea level. There is great uncertainty, however, about the ability of instruments to withstand the intense radiation assumed to exist in Jupiter's radiation belts. Information on this point will be critical in determining how close a grand-tour spacecraft can be allowed to approach Jupiter when it is being accelerated by the planet's gravitational attraction. The maximum budget would provide for four grand-tour launches, beginning in 1976 or 1977. If the intermediate budget is adopted, only two grand tours can be financed.

For the period in the 1980's when the alignment of planets will no longer be favorable for gravity-assisted trips to the most remote planets, low-thrust electric propulsion systems will be needed. One proposed system using solar cells as a power source should be able to propel a craft with a 1,300-pound payload to Saturn in 2.3 years, to Uranus in 5.2 years, to Neptune in 7.3 years and to Pluto in 10 years. An alternative nuclear-electric propulsion system should be able to propel a 150-pound payload to Pluto with a flight time of less than four years.

Decelerated Accelerator

A plague of wet magnets that has caused a delay of some six months in the start of experiments at the National Accelerator Laboratory in Illinois appears to have been stemmed, according to officials at the laboratory. Since the target date of last July 1 was a year ahead of the original schedule, it is likely that notwithstanding the difficulty the big accelerator will be in operation sooner than originally planned. At full power the machine will produce protons of up to 500 billion electron volts (GeV). The largest accelerator now in operation (at Serpukhov in the U.S.S.R.) develops slightly more than 70 GeV.

The function of the magnets, of which there are about 1,000 in the four-mile circular tunnel of the accelerator's main ring, is to guide and control the proton beam in the two-by-five-inch vacuum chamber that threads through the magnets. According to plan, the tunnel was

© 1971 SCIENTIFIC AMERICAN, INC

buried last winter under some 20 feet of earth. Then the earth froze and the interior of the tunnel became cold. Condensation occurring when warm summer air entered the tunnel was the source of the moisture that affected some 250 of the magnets.

The laboratory has dealt with the problem by sucking water out of the affected magnets, baking them, replacing the magnet coil if necessary and impregnating the magnets with epoxy resin. At the same time the tunnel has been dried and heated. Since the budget for the laboratory had a contingency fund, officials hope the final cost of the accelerator will still be within the $250 million originally authorized by Congress. Presumably the problem of wet magnets will not recur next summer because the operation of the tunnel this winter will keep the tunnel warm with the heat of the magnets. Moreover, since the magnets have a circulating-water cooling system that functions when they are in operation, it will be possible (as it was not last winter) to pump hot water through the system if necessary to keep the tunnel warm.

Moonglow

Thermocouples placed in two holes drilled on the surface of the moon by the astronauts of *Apollo 15* show that heat is flowing outward through the lunar surface at about three times the rate that had been expected. The source of the heat is presumed to be the decay of radioactive elements and confirms the high level of radioactivity previously measured in samples of lunar soil. The puzzling aspect of the related measurements is that if the entire moon were as radioactive as the surface soil, the entire body, in the words of one investigator, "would be just a molten puddle."

The heat-flow experiment was conducted under the direction of Marcus G. Langseth, Jr., of the Lamont-Doherty Geological Observatory of Columbia University and Sidney P. Clark, Jr., of Yale University. The actual heat flow recorded by the lunar instruments is about 50 microcalories per square centimeter per minute, or roughly half the average value for the earth. During the formation of the earth radioactive elements such as uranium, thorium and radioactive potassium became concentrated in lightweight rocks that formed the crust while much of the iron sank to form the core.

One early suggestion was that the moon experienced a rain of molten, highly radioactive rock during the final stages of its formation. A new hypothesis, however, has recently been put forward by a team of seismic investigators from the Lamont-Doherty Observatory, the General Dynamics Corporation, the University of Hawaii and the Massachusetts Institute of Technology. According to this hypothesis, based on seismic evidence, the moon was sufficiently plastic during the first billion years or so of its formation to support a fractionation process that depleted the lunar core of radioactive elements and carried them to the surface. Subsequently the core cooled, and it is now quiescent.

Ethics and Anthropology

A storm that first arose at the 1970 annual meeting of the American Anthropological Association broke again at the 1971 annual meeting in New York in November. At issue was the suspicion that association members doing Government-financed research in Thailand were gathering information about local tribal groups that would be useful to the planners of "counterinsurgency" measures. An earlier episode, involving a secret research effort in Latin America called Project Camelot, had caused the association to issue a declaration of ethical principles in 1967. The association later created a nine-member committee to inquire into post-Camelot ethical problems, with Eric R. Wolf as chairman.

The controversy over research in Thailand became public in April, 1970, when the Student Mobilization Committee to End the War in Vietnam released to the press excerpts from several surreptitiously copied documents identifying five association members in contexts that implied their involvement in counterinsurgency research. The student group had previously furnished copies of the documents to some members of the association's Ethics Committee; their press releases included a reply from Wolf and Joseph C. Jorgensen acknowledging receipt of the material and deploring its implications. Wolf and Jorgensen were reprimanded by the association's executive board, and both men resigned from the Ethics Committee. The membership then voted at its 1970 meeting to establish an *ad hoc* committee, with Margaret Mead as chairman, to report at the next annual meeting on the facts with respect to members' activities in Thailand and to assess the propriety of Wolf and Jorgensen's actions.

The Mead report was presented at the annual meeting; it consisted of four main sections. The first section, concerning the work in Thailand, presented the cautiously worded conclusion that "no civilian member" of the association had contravened the post-Camelot declaration of ethics. The committee nonetheless declared that "pervasive corruption" in the scholarly community with respect to Government financing "may also have prepared anthropologists... to close their eyes to misuse of their data... and, in the long run, misuse of their talents vis-à-vis science." "To our previously articulated ethical principles," the report stated, "we must now add, *anthropologists must not endanger identifiable villages or communities.*"

The second section, concerning the actions of the former members of the Ethics Committee, found them to have acted "hastily, unfairly and unwisely" in making public statements without "first having consulted" their accused colleagues. The third section, dealing with ways of preventing possible misuse of anthropological data, recommended that, when in doubt, all workers should alter the names of people and places and perhaps also consider delaying publication of field data for as long as five years.

Finally, the report recommended a new committee system. It proposed that the Ethics Committee be reorganized and be assigned a narrower function, confined to "matters of scholarly and scientific ethics" and specifically excluding "the field of applied anthropology." In addition the report proposed the establishment of a new standing committee to deal with Government policies that impinge on anthropologists' various fields of special competence.

In a tense and protracted evening meeting some 500 members of the council of the association were asked to signify "receipt" of the report, an action not formally implying either approval or disapproval. It was agreed that each section of the report, beginning with the Thailand findings, would be put to a vote. The majority of those present then voted against receipt of each section in turn and ultimately voted rejection of the report as a whole by referring it back to the executive board.

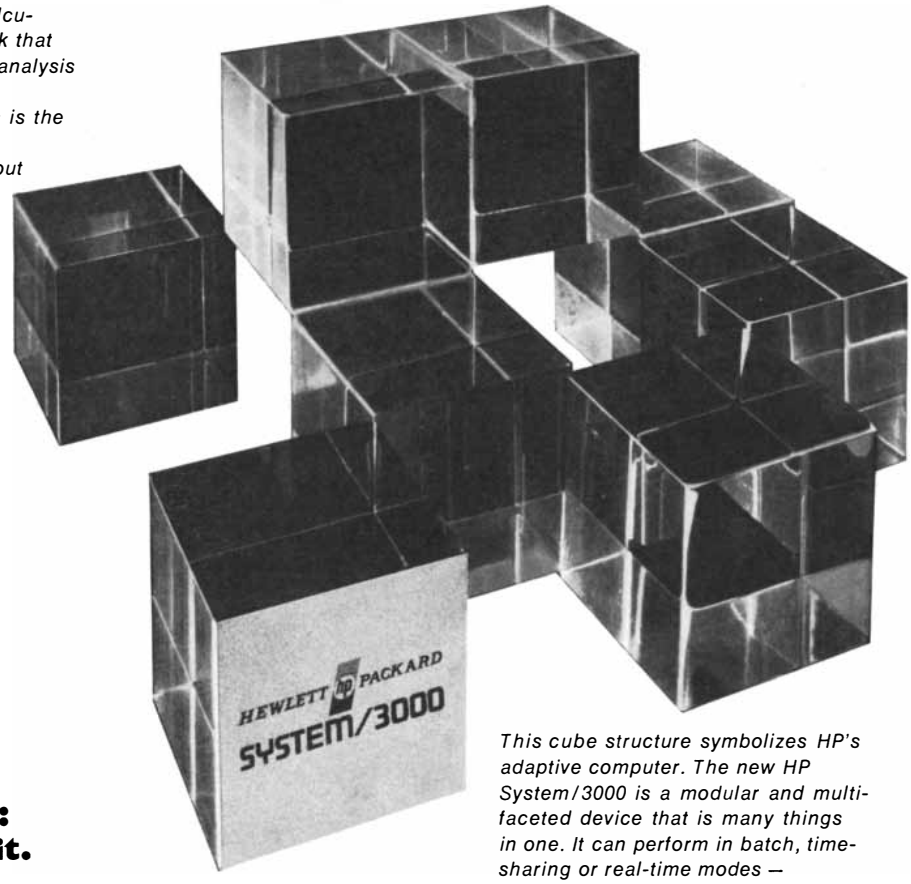
Each to His Own Blood

As is well known, the principal problems of blood transfusion are inadvertent hepatitis infection and making sure that the donee's blood matches the donor's. Why not solve both problems by transfusing the donor with his own blood? To be sure, this cannot be done in an emergency, but perhaps two-thirds of all transfusions are administered during and after surgery, when there is

© 1971 SCIENTIFIC AMERICAN, INC

Some things are changing for the better.

Many people know us as an instrument manufacturer: we make more than 2,000 products for measurement, test and analysis. Others know us as a computer company: more than 10,000 own our programmable calculators and computers. We prefer to think that our business is to serve measurement, analysis and computation needs . . . in science, industry, medicine and education. This is the rationale behind every new instrument, computer or system that we tell you about in these ads. This month:



This cube structure symbolizes HP's adaptive computer. The new HP System/3000 is a modular and multi-faceted device that is many things in one. It can perform in batch, time-sharing or real-time modes — simultaneously — and use three different computer languages at the same time.

Computer proliferation: a better way to control it.

Until now many computer users in high-technology areas faced a puzzling fork in the road. With several computers already in use, they had to decide either to continue the proliferation of dedicated systems or turn to a large-scale system costing at least a half-million dollars.

Now there is a third and happier choice: HP's new adaptive computer System/3000. It is the first small-scale computer that handles batch, real-time and time-sharing all at once, each in more than one computer language — at half the price of a good alternative.

For time-shared tasks, it handles up to 64 users. For real-time processing, it allows several programs to be active simultaneously, completely protected one from the other. In batch processing, it permits execution of a variety of programs at extremely high speed.

And System/3000 is multi-lingual in all modes, including FORTRAN, BASIC and the new SPL/3000, a high-level systems programming language that combines the power of a Compiler with the efficiency of an Assembler.

How System/3000 accomplishes all this is a long story. Part of it involves advanced computer architecture, virtual memory that assures available core for every program in any mode, a huge micro-programmed instruction set, and a multiprogramming executive that insures dynamic allocation of system resources. For the rest of the story on the adaptive computer write for our brochure on System/3000.

© 1971 SCIENTIFIC AMERICAN, INC

Computer proliferation: a better way to do it.

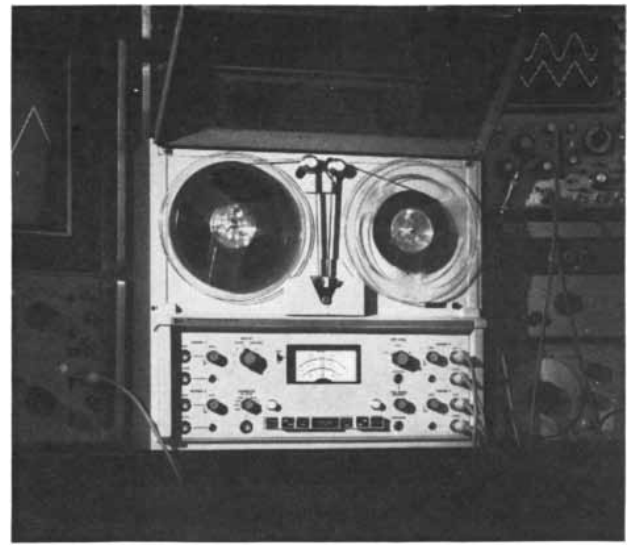
For users who need to access a large data bank and also require ease of I/O interface, computer proliferation isn't always a problem. It often can be the solution.

To perform these tasks at a new low price, HP now offers a family of small disc-based computer systems dedicated to time-sharing, real-time or batch processing. Each system uses HP's versatile 2100 digital computer, and each is available with an impressive amount of storage. A choice of mass storage devices provides 5 to 47 million bytes of disc storage.

An HP disc-based system can be tailored to your specific needs, with more than a dozen peripherals and general-purpose interfaces for a variety of instruments. The new systems also use all of HP's existing software of over 500 library programs and applications packages.

The dedicated batch system is priced from $33,000. Time-share and real-time systems start at about $50,000. For more information, write for our brochure on dedicated disc-based computer systems.

Q.E.D. There are computation problems in every business that Hewlett-Packard can solve.



Instrumentation quality tape recording at a bargain price.

Most scientists would use portable instrumentation tape recorders for analog recording if only they performed as well as the big expensive laboratory machines. Unfortunately, their small size usually meant reduced performance.

Then came the HP 3960. Truly portable in size (50 pounds) and low in price ($4,270 for a fully-equipped four-channel instrument), the 3960 actually outperforms most laboratory machines costing five times more.

If this sounds too good to be true, listen to some of the 3960's capabilities. At 15/16 ips, its FM signal-to-noise ratio of better than 200:1 lets you play back signals that would be buried in noise (ECG's for example) on many lab machines as well as on any other portable.

The 3960 lets you mix and interchange four FM direct record/reproduce channels at will. You have a choice of three electrically-switched speeds, for a time-base expansion of 16:1 or 10:1 . . . without signal degradation. Tape drive is bidirectional so that you don't have to rewind either to continue recording or to play back.

Built-in facilities let you calibrate the 3960's FM electronics without external equipment. And an integral peak-reading meter lets you optimize record level without using a scope. Options include a 5 to 30 foot loop adaptor, an interrupting voice channel, and an inverter for 12 or 28 VDC . . . all integrally mounted.

Write for Application Note 89, a tape recording handbook useful to scientists interested in tape recording techniques for vibration and test analysis, research and clinical medicine, acoustics, oceanography and other environmentally difficult research projects.

For more complete information, write Hewlett-Packard, 1508 Page Mill Road, Palo Alto, California 94304. In Europe: 1217 Meyrin-Geneva, Switzerland.

00121

HEWLETT hp PACKARD

© 1971 SCIENTIFIC AMERICAN, INC

usually considerable advance warning. Melvin M. Newman, Roger Hamstra and Matthew Block of the University of Colorado Medical Center have been looking into the matter; they report in *The Journal of the American Medical Association* that between 1962 and 1968 a total of 178 patients were given transfusions of their own blood. The blood was drawn at least 48 hours before surgery. Each withdrawal was followed by intravenous injection of iron dextran to stimulate the production of new red cells. Only one patient had a severe reaction, consisting of fainting and temporary respiratory arrest shortly after the injection of iron dextran was begun. Eight other patients had mild reactions such as chills, low fever, nausea or tingling of the hands.

Although the first successful self-transfusion during surgery was reported in 1921, the technique has been used infrequently, and mainly in cases where the blood type was rare and difficult to match. Surgeons have feared that withdrawing blood before surgery would impair the patient's condition by decreasing his blood volume or depleting his "vital resources." Favorable reports began appearing in 1962, but in all these cases the amount of blood that could be withdrawn was limited by the inability of the patient to rapidly replace red cells. The rate of red-cell production is limited by the body's ability to mobilize its stores of iron. In a normal man the maximum increase of red-cell production in response to acute blood loss is about two- or threefold. Supplying additional iron by injection of iron dextran boosts the rate of red-cell production four to six times. It has been shown that with the administration of iron dextran a normal adult male was able to maintain a stable level of red cells for 22 weeks in spite of the withdrawal of 500 milliliters of blood five times a week.

For patients who are to receive their own blood, the procedure is simple. When it was expected that only one unit (500 milliliters) of blood was needed, the blood was withdrawn 48 to 72 hours before surgery. When two units were required, the first was drawn five or six days before surgery and the second 48 hours before. When more than two units were required, at least 48 hours were allowed to elapse between withdrawals. Not everyone is able to supply his own blood for later transfusion. In particular, people who are unable to increase their red-cell production in response to the injection of iron dextran are not eligible. Newman, Hamstra and Block estimate, however, that at least half of the patients scheduled for elective surgery can be their own blood donor.

Among the advantages of self-donated blood is the reduction of the cost of a unit of blood from $40 to a handling charge of $15 or less. Newman, Hamstra and Block also report that in self-donation there has been a significant amount of blood left over; this blood has been added to the community blood bank.

The American Way of Death

When the death rate of Americans leveled off in the 1950's after nearly half a century of continuous decline, there was some question about whether the leveling off was a temporary halt in the downtrend or whether it marked the beginning of a new uptrend. An upward trend in the death rate of white males (including Mexicans and Puerto Ricans) over the age of 45 became evident in the decade 1954–1963. The upturn in mortality rate continued into the period of 1963–1968 for older white males and also spread to younger white males (ages 15 to 44), who had a dramatic rise in their death rate with increases of more than 10 percent in some age brackets. Males of other races (blacks, American Indians, mixtures of white and other races) showed a decline in mortality rate up to 1963 and then a sharp rise.

The increase in male mortality in the U.S. population is documented in a report by the National Center for Health Statistics: *Leading Components of Upturn in Mortality for Men.* A leading cause of death of males is still heart disease, but the long upturn in mortality from this cause appears to be leveling off, declining from a rate of about 376 per 100,000 population in 1960 to about 362 in 1967. In all age groups men have a higher mortality rate than women, but male mortality with respect to female reaches a peak in the age group 20 to 24. Automobile accidents are the leading cause of death for white males and a major cause of death for nonwhites in this age group.

There was a major increase in violent deaths between 1963 and 1968 for men in the 15-to-24 age bracket. In addition to automobile accidents there was a sharp rise in suicide and homicide for both whites and nonwhites. Lung cancer, cirrhosis of the liver and circulatory diseases are taking a progressively larger toll of younger men than they did a few decades ago.

Changes in the mortality rates of women differ from those of men. Older white females (over 60) continued to have a decreasing death rate up to 1968, whereas the death rate of younger white women (15 to 59) leveled off in the 1958–1963 period and then turned slightly upward in the next five-year period. Females of other races showed continuing declines in mortality for most age groups up to 1968.

The greater mortality rate of males has contributed to the reversal in the ratio of males to females in the U.S. In 1910 there were about 106 males for every 100 females. The ratio had dropped to 100 males to 100 females in 1940 and continued to decline, reaching 95 males to 100 females in 1970. Deaths occurring in foreign countries, including war deaths, although contributing to the change in the male-female ratio, were not included in the mortality rates in the Government report.

By Eggs Alone

An experiment in which weanling rats were fed diets consisting of various single foods has confirmed that eggs evidently provide all the nutrients a rat requires to grow and stay healthy but that a variety of cereals, "enriched" according to commercial standards, were uniformly inadequate. Most rats limited to exclusive diets of puffed rice, wheat flakes, shredded wheat and macaroni (all "enriched") were barely able to hold their weanling weight (about 60 grams) and a number failed to survive a 90-day experiment. Rats fed roasted peanuts did little better or no better. Rats fed all-meat frankfurters and hamburger gained weight for 40 to 50 days but then began to fail. Rats fed canned tuna fish grew somewhat more slowly but continued to grow for about 80 days. Only milk approached eggs in nutritional value, but after about 60 days its well-known deficiency in iron and copper made itself felt. In general rats fed eggs and milk were enabled to multiply their initial weight by a factor of six or seven in 75 days.

The experiment was conducted by Roger J. Williams, James D. Heffley and Charles W. Bode of the University of Texas in response to criticism made of earlier experiments in which they had demonstrated that rats fed nothing but bread inexpensively improved according to a special formula grew seven times faster than rats fed commercial "enriched" bread. The critics in effect said that man does not live by bread alone and that the experiments "would have given the same or similar results if you had begun with milk, eggs, meat, or any other food."

They also conducted new tests in

© 1971 SCIENTIFIC AMERICAN, INC

which they added their special enriching formula to commercial puffed rice, wheat flakes, shredded wheat and macaroni. They found that the supplement significantly improved the three wheat products but did not improve the puffed rice.

The term "enriched," as commercially used, means a product to which thiamin, riboflavin, niacin and iron have been added. Williams believes that bread, an important part of many diets, could be vastly improved at a cost of less than a cent a loaf by the addition of the following nutrients in the quantities shown for each pound of "enriched" flour: two milligrams of pyridoxine, 4.5 milligrams of pantothenate, 2.2 micrograms of cobalamine, 2,160 units of vitamin A, 20 milligrams of vitamin E, .5 milligram of folic acid, .5 gram of L-lysine, 300 milligrams of calcium, 713 milligrams of phosphate, 150 milligrams of magnesium (oxide), 20 milligrams of manganese (sulfate) and four milligrams of copper (sulfate).

Men out of Asia

That the New World was first populated by waves of emigrants from Asia has long been accepted. Each year archaeological investigations from Alaska and the Canadian Arctic down to Tierra del Fuego add detail to this main conclusion and even suggest in a general way what parts of Asia gave rise to the emigrants. A site in British Columbia has now yielded artifacts that strongly imply a specific homeland for the men who fashioned them: northern Japan.

Working near Telegraph Creek in northwestern British Columbia in 1969 and 1970, Jason W. Smith and his colleagues from the University of Calgary unearthed several hundred delicate obsidian "microblades," flakes seldom more than an inch long or a fifth of an inch wide, along with eight of the "cores" that yield such blades. Because artifacts made of obsidian, a volcanic glass, undergo surface hydration at a uniform rate, the age of the excavated blades and cores could be determined. The oldest blades had been produced around 9000 B.C. and the cores were shaped at various times between 4500 and 1500 B.C. One oddity is that some cores may not have been formed for the production of blades but may originally have been projectile points; evidently they were reutilized as raw material once they had been broken in use.

Similar microblade industries elsewhere in northwestern North America possess a sequential antiquity that suggests a steady push into the region from

This valuable new book gives you all the latest facts on FAX

There may come a day when no one will ever have to run out for his morning newspaper, or even wait for it to be delivered. It will be printed—pictures and all—right in the subscriber's own home!

And the postal system, as we know it, may cease to exist. Instead, people will transmit letters instantaneously, over any distance—the way they now make phone calls.

These are only two of the exciting possibilities of FAX—facsimile communication—which is already widely used by business, industry, government, and the news media. In this comprehensive book, a FAX expert affiliated with Bell Telephone Laboratories explores in detail the history of this versatile communications medium... its technology and hardware... its present applications... and its thrilling promise for the future. *Illustrated with 88 photographs and 72 line drawings.*

FAX

THE PRINCIPLES AND PRACTICE OF FACSIMILE COMMUNICATION

by Daniel M. Costigan

$10, now at your bookstore, or send remittance (plus applicable sales tax) to

CHILTON BOOK COMPANY

401 Walnut St., Philadelphia, Pa. 19106

© 1971 SCIENTIFIC AMERICAN, INC

Plymouth Cricket comes with power front disc brakes, standard. It's a great little stopper.

CHRYSLER
Plymouth

CHRYSLER MOTORS CORPORATION

Coming through with the kind of car America wants.

ANTIQUITY

A Quarterly Review of Archaeology
edited by Glyn Daniel

Founded in 1927 by O. G. S. Crawford, ANTIQUITY has appeared regularly ever since and won acclaim the world over as the most authoritative journal in its field. While written by specialists, the articles, notes and reviews are popular in character and indispensable to all interested in the development of man and his past.

Dr Glyn Daniel of the Faculty of Archaeology and Anthropology in the University of Cambridge, and Fellow of St John's College, has been Editor since 1956.

The annual subscription, postage included, is $8. Subscription forms and bankers' orders are available on request from

ANTIQUITY PUBLICATIONS LIMITED
104 Hills Road, Cambridge, England CB2 1LW

eastern Asia. The closest Asian parallel is an advanced Paleolithic microblade tradition that arose between 18,000 and 13,000 B.C. and is well represented on the northernmost of the Japanese islands: Hokkaido. Writing in *Arctic and Alpine Research,* Smith suggests that in both the New World and Asia (other microblade findings have been made in Mongolia, in northern China, in southern and eastern Siberia and in the Kamchatka Peninsula) the industries be collectively designated the East Asian–Northwest American Microblade Tradition.

Driving Blind

A motorist groping through a heavy fog may well wonder if radar would help. Although it is unlikely that radar for private automobiles will be available in the near future, many of the technical problems have been solved in an ingenious system developed by the Mullard Research Laboratories for airport vehicles. According to K. L. Fuller of the Mullard Laboratories, such a system is needed because aircraft can now land blind in a fog; ground vehicles should be able to operate under the same conditions in guiding planes to the terminal and also in the event of an accident.

The Mullard system is called AVOID (for airfield vehicle obstacle indication device). Fuller describes it in *Philips Technical Review.* The main problem confronting the designers was that, whereas other radar systems need only detect objects at ranges measured in hundreds of feet or miles, a ground-vehicle system must pick up objects at distances of a few feet. Thus it is difficult to design a system that works with the usual pulsed signal (the echoes come back too fast) and swinging antenna (it cannot scan the target area fast enough). The problem was basically solved with a frequency-modulated signal (the echoing waves come back slightly out of phase) and a stationary antenna that scans the target area electronically.

The image presented by the AVOID system is a direct representation of the scene ahead of the driver rather than the usual maplike plan view of the radarscope. It is projected on a television picture tube in the cab of the airport vehicle. The system was tested in an extensive blind-driving program at two locations: a parking lot and an unoccupied airfield. The driver needed a short period to familiarize himself with the system, but once he had done so he was able to drive around the two test sites without colliding with numerous obstacles in his path.

© 1971 SCIENTIFIC AMERICAN, INC

First price is publisher's list. **Boldface** shows member's price.

832. China and Russia: The "Great Game"
By O. Edmund Clubb. Three centuries of "massive contest," always short of war. $12.95/**$8.75**

105. Huey Long
By T. Harry Williams. Winner of the National Book Award and Pulitzer Prize. $12.50/**$8.50**

554. The Trial of the Germans: Nuremberg, 1945-46
By Eugene Davidson. $12.50/**$8.95**

657. The Creation of the American Republic, 1776-1787
By Gordon S. Wood. Bancroft Award winner. $15.00/**$9.95**

104. The Indian Heritage of America
By Alvin M. Josephy, Jr. $10.00/**$7.50**

677. Witchcraft at Salem
By Chadwick Hansen. $6.95/**$5.75**

687. Anti-Intellectualism In American Life
By Richard Hofstadter. Pulitzer Prize winner. $7.95/**$5.95**

708. Jefferson and the Rights of Man
By Dumas Malone. $10.00/**$7.95**

775. The Rising Sun: The Decline and Fall of the Japanese Empire, 1936-45
By John Toland. $12.95/**$8.95**

476. The Life of Lenin
By Louis Fischer. Winner of the National Book Award. $12.50/**$8.50**

726. A History of the African People
By Robert W. July. $15.00/**$8.95**

750. The German Dictatorship: The Origins, Structure, and Consequences of National Socialism
By Karl Dietrich Bracher. $13.95/**$9.50**

671. Medieval History: The Life and Death of a Civilization (2nd Edition)
By Norman F. Cantor. $10.95/**$8.50**

746. Empire of the Steppes: A History of Central Asia
By René Grousset. $17.50/**$9.95**

727. Their Tattered Flags: The Epic of the Confederacy
By Frank E. Vandiver. $10.00/**$6.95**

606. Henry VIII
By John J. Scarisbrick. $10.95/**$7.65**

842. The Black Image in the White Mind
By George M. Fredrickson. America's shifting attitudes toward blacks from 1817-1914 and the author's chilling conclusions. $10.00/**$7.75**

836. 20th Century China
By O. Edmund Clubb. $10.00/**$7.50**

791. Napoleon: From 18 Brumaire to Tilsit, 1799-1807
By Georges Lefebvre. $10.00/**$7.50**

103. From Slavery to Freedom: A History of Negro Americans
By John Hope Franklin. $12.00/**$8.40**

369. The Spanish Civil War
By Hugh Thomas. $12.50/**$8.50**

460. Hitler: A Study in Tyranny
By Alan Bullock. $10.00/**$6.95**

771. The World of the French Revolution
By R. R. Palmer. $8.50/**$6.75**

543. Ancient Europe
By Stuart Piggott. From 6000 B.C. to the Roman era. $10.00/**$7.50**

The History Book Club
Stamford, Connecticut 06904

Please enroll me as a trial member and send me, for only 99 cents, the three books whose numbers I have filled in:

[ ] [ ] [ ]

Also send me, at the special low member's price, my first selection:

[ ]

Within two weeks, I may return the books and owe nothing. Or keep them and agree to take only four more selections in the coming year from among more than 150 offered each month in the Club Review. I may choose any of these instead of the Editors' Choice, or no book at all, by returning the reply card accompanying the Review. SA 06S

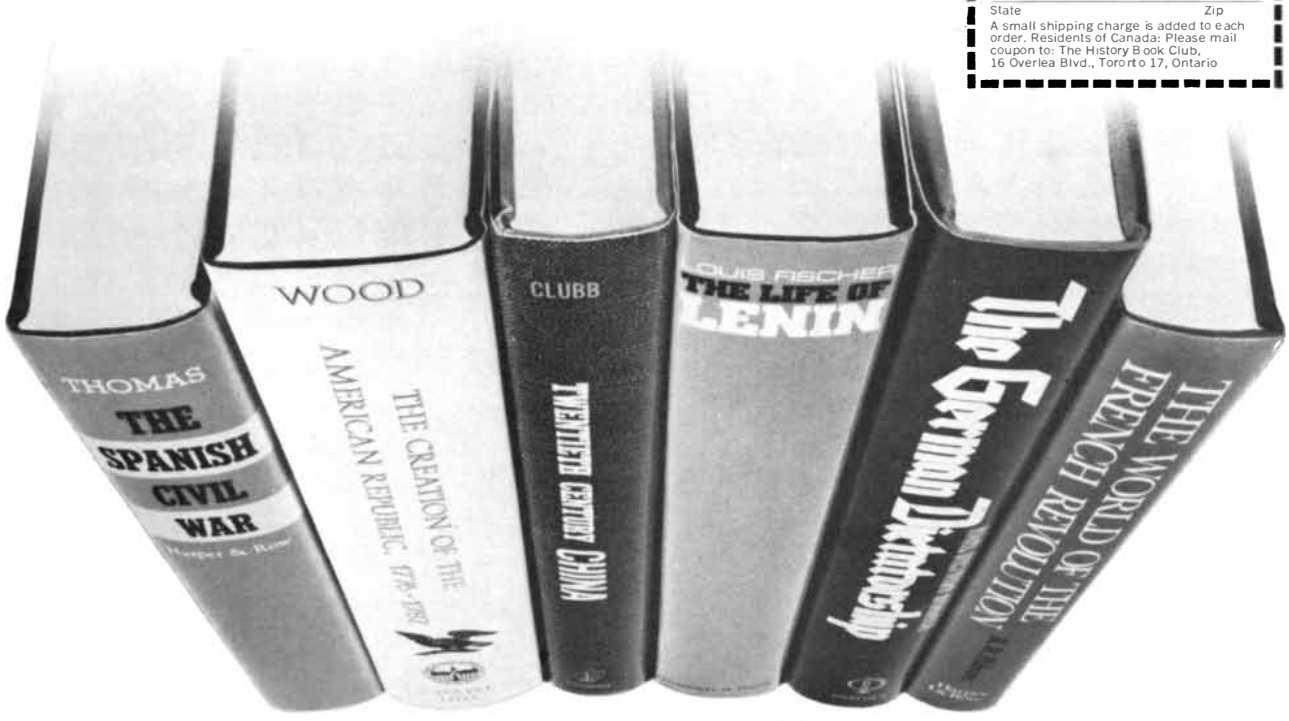
Print Name

Address

City

State Zip

A small shipping charge is added to each order. Residents of Canada: Please mail coupon to: The History Book Club, 16 Overlea Blvd., Toronto 17, Ontario



Special on revolutions. Any three for 99 cents.

In one violent thrust, revolution brings a nation's entire history into dramatic focus.

That's why we'd like to send you any three of the books on revolution shown here as your introduction to The History Book Club.

Of course, if you prefer, you may choose the company of tyrants, martyrs, politicians, kings. Become embroiled in wars. Or go on voyages of discovery.

In fact, you can have any three of the books we've listed. All for 99 cents, when you take a fourth at the low member's price.

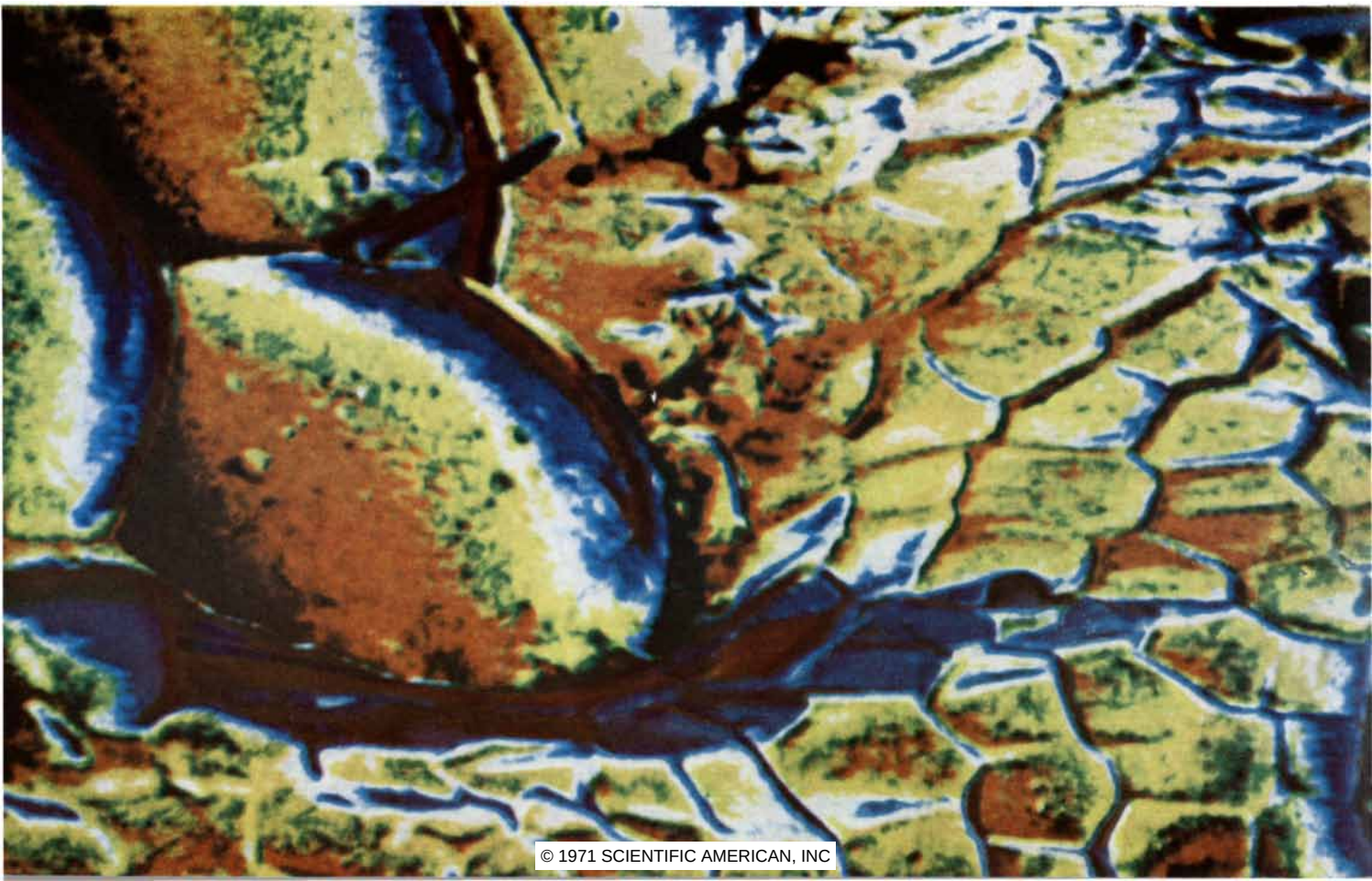
If you're not delighted, you may of course return them. But if you do keep them, you simply take four more new books during the coming year, from the 150 offered each month. You get these at savings of $2, $3 or $4 off bookstore prices. And you also earn other savings through bonus books you can choose from the Club's entire list.

The History Book Club

Stamford, Connecticut 06904

© 1971 SCIENTIFIC AMERICAN, INC

EYE OF FLOUR BEETLE, *Tribolium confusum*, is magnified about 4,500 diameters in these micrographs made with a scanning electron microscope. The rounded structures are individual lenses in the insect's eye. The microscope uses a television-like scanning system for producing images on a cathode ray tube. In the picture above variations in signal intensity are rendered as shades of gray. In the picture below color has been introduced by assigning a particular hue to a particular strength of signal. Color helps to bring out areas that have the same signal intensity. Thus the repetitive small changes in the scales surrounding the beetle's eye are more obvious in color than in black and white. These two scanning electron micrographs and those on subsequent pages were made in the laboratories of the two authors at the University of California at Berkeley. Their collaborator in making these two pictures was J. Pawley. The flour beetle is one of the few insects that can be observed in the scanning electron microscope while still alive.



© 1971 SCIENTIFIC AMERICAN, INC

THE SCANNING ELECTRON MICROSCOPE

The remarkably three-dimensional pictures from the instrument are made by scanning the microscope specimen with a fine electron beam synchronized with the electron beam of a cathode ray picture tube

by Thomas E. Everhart and Thomas L. Hayes

Within the past few years pictures made with the scanning electron microscope have been widely published, not only in scholarly journals but in large-circulation magazines as well. The outstanding characteristic of these images is their remarkable three-dimensional quality. Whether the image depicts the eye of a fruit fly, a red blood cell or a close-up of the human skin, it conveys a sense of reality normally lacking in micrographs made by other means.

The usual micrograph made either with a light microscope or with a conventional electron microscope (often referred to as the transmission electron microscope) is a two-dimensional image. What one sees is a pattern of light and dark areas produced by the passage of light or electrons through a thin slice of the specimen. The light microscope presents an image in two dimensions because it has a very limited depth of field, which means that it can be focused sharply only in one plane. If the sample is fairly transparent, one can adjust the focus at different depths, but since material above and below the plane of focus can interact with light in a manner that blurs the image, the light microscope works best with thin samples, viewed by transmitted light, or with flat samples, viewed by reflected light. Only at low magnifications (below roughly 200 diameters) is the light microscope useful for showing the shape of the specimen in depth.

The transmission electron microscope requires even thinner samples than the light microscope because only those electrons that emerge from the specimen with a single energy can be focused in a single image plane by the magnetic field of the objective lens. Therefore in most cases it is the electrons that have been deflected but not changed in energy that are utilized in the formation of the image. If the electrons emerged with a wide spectrum of energies, they would be bent by varying amounts in passing through the magnetic lens and the image would be blurred.

If a reasonable proportion of electrons are to emerge with their energy unaltered, the sample must be very thin even for beams of electrons with an energy of 50,000 to 100,000 electron volts. Thus the sample for transmission electron microscopy is usually not more than 500 angstroms thick, or about a thousandth of the diameter of a typical living cell.

Microscope Characteristics

The scanning electron microscope is able to provide images of three-dimensional objects because in its normal mode of operation it records not the electrons passing through the specimen but the secondary electrons that are released from the sample by the electron beam impinging on it. The sample can therefore be of any size and thickness that will fit in the instrument's evacuated sample chamber. The secondary electrons do not have to be focused but are simply collected. Moreover, because the envelope of the impinging beam resembles a sharp needle, the instrument achieves great depth of field. The transmission electron microscope also has a large depth of field, but normally the effective depth of field is greatly limited by the thinness of the specimen.

The scanning electron microscope is capable of a range of magnifications that overlaps the range of the light microscope or hand magnifying glass at the low end and the range of the transmission electron microscope at the high end. The light microscope has a maximum resolution of between 2,000 and 3,000 angstroms, that is, it cannot separate two points that are closer together than 2,000 or 3,000 angstroms (half the wavelength of blue or orange light). The maximum effective magnification of the light microscope is about 1,000 diameters; although greater magnifications are possible, the image becomes fuzzy and does not supply more information.

The best transmission electron microscopes have a resolution of between two and five angstroms, so that the maximum effective magnification exceeds a million diameters. The scanning electron microscope is useful over a magnification range from about 15 diameters to about 100,000 diameters, although the image begins to get fuzzy above 20,000 diameters. Most manufacturers guarantee a resolution of 200 angstroms, and a resolution of better than 100 angstroms can be attained with certain kinds of sample.

The broad magnification range of the scanning electron microscope, together with the ease of changing magnification, makes it easy to "zoom" from a gross image of the object to an image showing fine details. Unlike the zoom lens, which changes its focal length and hence its depth of field (as can be seen in televised football games), the lens of the scanning electron microscope does not change its focal length and depth of field in "zooming." Objects seen in the scanning electron microscope at a magnification of 15 diameters look much as they do when they are viewed with a strong magnifying glass or a low-power microscope. As one increases the magnification of a light microscope, however, the depth of field steadily decreases, and at a magnification of about 1,000 diameters the increasing resolution of detail ends. As one increases the magnification in a scanning electron microscope the relative depth of field does not decrease but remains constant, and the increasing resolution of detail ends at a magnification

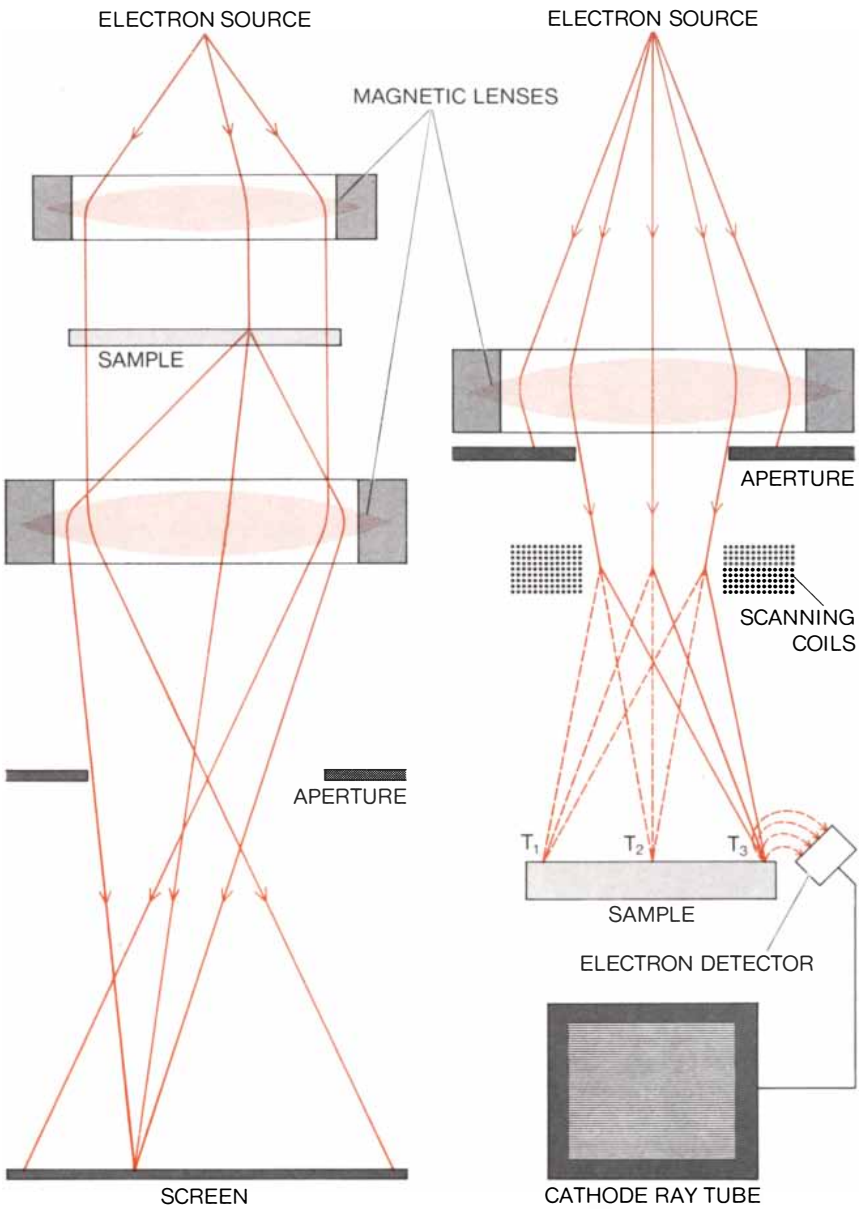
© 1971 SCIENTIFIC AMERICAN, INC

of about 20,000 diameters. At higher magnifications the signal from a given area of the sample is increased somewhat over the "noise" from that area by statistical averaging, and magnifications of 100,000 diameters are normally available and are useful.

Since thin slices are not required, the preparation of electrically conducting samples for the scanning electron microscope is generally much simpler than it is for the light microscope or the transmission electron microscope. The conducting sample can be lightly glued to the sample holder (often with a conducting paint that dries rapidly) and placed inside the microscope. A brief wait of about two minutes is needed to exhaust the gases from the sample chamber.

Samples that do not conduct electricity involve somewhat more effort. After the sample has been mounted a thin layer of metal (often gold) is evaporated onto it to hold the surface of the sample at a constant electric potential. The optimum thickness of the metal coating depends on the sample and the desired resolution. The coating can be likened to a layer of snow covering the ground. A thin layer obscures only small pebbles, a deeper layer obscures rocks and a very deep layer in the mountains can even cover cabins. It is possible for a layer to be too thin, for example if it fails to cover slightly recessed areas (comparable to the ground under trees) that will nevertheless be exposed to the incident electron beam. The desirable coating thickness ranges from perhaps 50 angstroms for work at the highest resolution to a few hundred angstroms for work at lower magnification.

Biological tissue requires specialized preparation. The tissue must first be fixed, to hold cells and other parts in position during subsequent steps. Next, the water is removed and the sample is dried. The sample is mounted on a stub, or holder, and is coated with metal. It is then ready for examination in the microscope.



TWO KINDS OF ELECTRON MICROSCOPE are compared in schematic diagrams that emphasize their similarities and clarify their differences. The transmission electron microscope (*left*) closely resembles a conventional light microscope. The sample is continuously illuminated over its entire area by a collimated beam of electrons. The electrons are scattered in passing through the sample and are focused by a magnetic lens, producing an image of the sample on a fluorescent screen. In the scanning electron microscope the source is imaged onto the sample by a magnetic lens; the small spot of electrons thus formed is deflected by scanning coils to illuminate the sample point by point at sequential times (T_1, T_2 *and* T_3). Secondary electrons, generated where the primary beam strikes the sample, are drawn to a detector and produce an electrical signal. This signal modulates the intensity of an electron beam inside a television tube scanned in step with the spot.

How the Image Is Formed

Early scanning electron microscopes were constructed in Germany in the mid-1930's by M. Knoll and slightly later by Manfred von Ardenne. Vladimir Zworykin and his co-workers at the RCA Laboratories constructed an improved version in the early 1940's. These instruments did not show the three-dimensional images that distinguish today's scanning electron microscopes because the signal-collection and amplification processes were not sufficiently well understood.

In the late 1940's Dennis McMullan, under the supervision of C. W. Oatley, constructed a much improved scanning electron microscope at the University of Cambridge. It incorporated a demountable secondary electron multiplier in the collection system and placed it in a more effective position than the one used by the RCA group. McMullan's design was subsequently improved on by later students. The first commercial scanning electron microscope evolved from the Cambridge designs and appeared in prototype form about 1963.

The transmission electron microscope is strictly analogous to the ordinary light microscope, but it achieves much higher resolution because electrons accelerated by a potential of, say, 100,000 volts have a much shorter wavelength (.4 angstrom) than visible light does (4,000 angstroms).

© 1971 SCIENTIFIC AMERICAN, INC

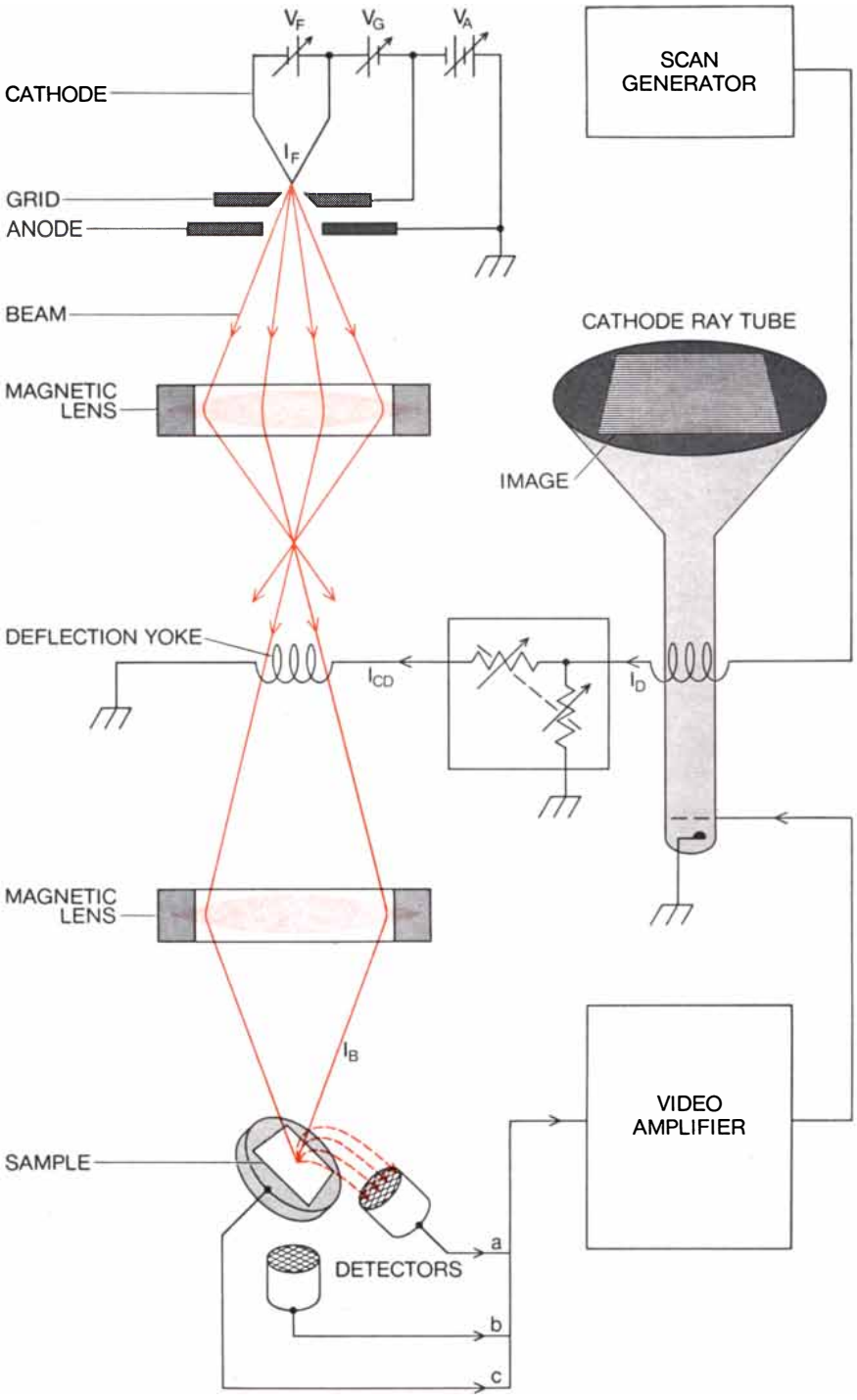
It is well known that the shorter the wavelength used in microscopy, the higher the resolution. Lens aberrations and other effects can also limit the resolution of a microscope; they limit the practical resolution of the best transmission electron microscopes to between two and five angstroms.

In transmission electron microscopy, as in light microscopy, the specimen is illuminated everywhere at once. The length of time a specimen can be viewed is limited by damage to the specimen and by drift in the performance of the instrument, but the image can be viewed as a whole at any one instant. The image appears on a fluorescent screen or is allowed to fall directly onto a photographic plate to expose a micrograph.

In scanning electron microscopy the specimen is scanned in a raster pattern by a focused electron beam and the image that results is formed by a time-sequencing technique similar to the one used in commercial television. There are two major differences. First, the standard television picture is made up of 525 horizontal lines, whereas the image produced by a scanning electron microscope can be varied from as few as 100 lines to more than 1,000. The finer scan is used to produce a micrograph. Second, the rate of scan that produces a micrograph is often much lower than the scanning rate in television. The reason is that it is difficult to focus many electrons into a scanning spot only 100 angstroms in diameter. Therefore the spot must dwell for a longer time at each of a million or so points on the sample to build up a secondary-electron signal strong enough to create an image substantially free of noise. Thus what one photographs on the microscope's picture screen to produce a micrograph is an image painted by a slow-moving electron beam that makes possible a statistically significant count of the signal generated at each point. Although one can speed up the scanning pattern to television rates for visual inspection, a time exposure of up to several minutes is needed to obtain high-resolution micrographs of the highest quality.

The Anatomy of the Microscope

The scanning electron microscope has four principal parts: an electron-optical system that produces the scanning electron probe; a sample where interaction between the probing electron beam and the material of the specimen produces the information signal; a detection system that collects the information-carriers emitted by the sample and amplifies the

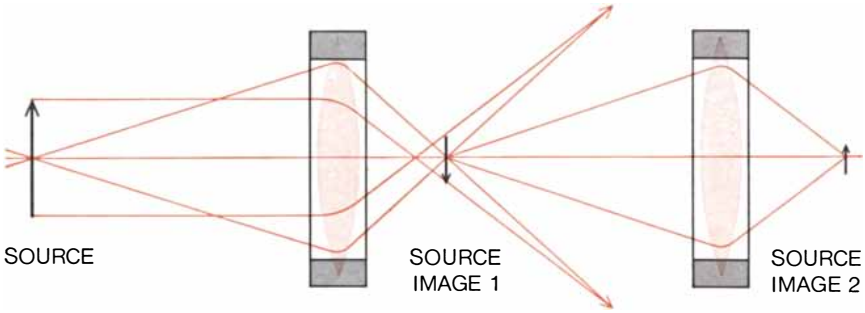


SCANNING ELECTRON MICROSCOPE is essentially a closed-circuit television system of some complexity in which the subject of the picture is illuminated by a flying spot of electrons. The electron beam originates at a highly negative tungsten cathode heated by current I_F. The emitted electrons are accelerated from the cathode through a grid and then through an anode that is at ground potential. This system of electrodes is called the electron gun. A system of magnetic lenses demagnifies the electron beam (*see top illustration on next page*). An adjustable current, I_{CD}, is applied to a deflection yoke to move the beam across the sample in a raster pattern. The ratio of this current to the current I_D in the deflection yoke of the cathode ray tube determines the magnification of the microscope. In commercial instruments the magnification can usually be varied continuously from about 15 diameters to 100,000. When the primary electron beam strikes the sample, it gives rise to secondary electrons whose number varies with the geometry and other properties of the specimen. The secondary electrons, collected by a detector (*a*), are normally used to produce a scanning electron micrograph, but backscattered electrons and photons emitted by the sample under bombardment can also be used to produce an image. Other kinds of images can be produced by transmitted electrons (*b*) and by currents induced in the sample (*c*).

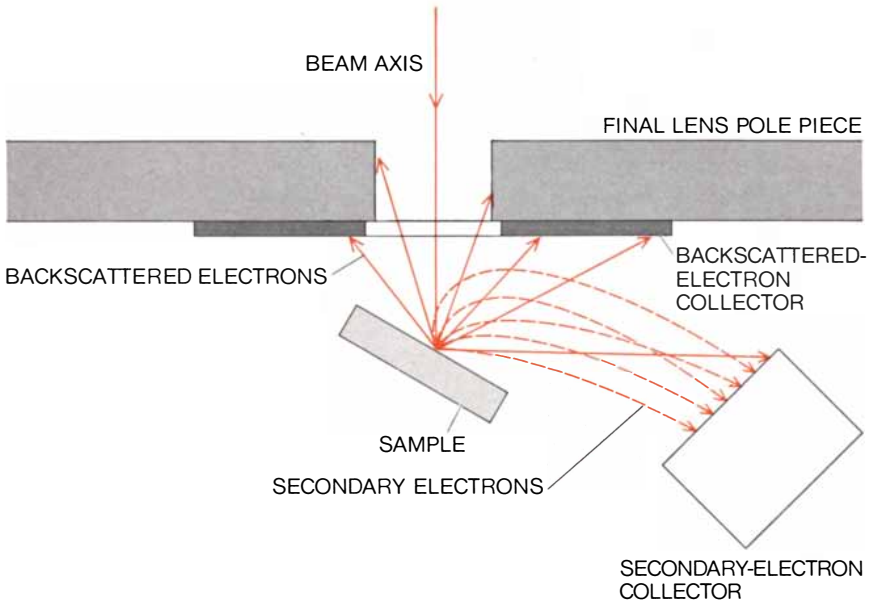
© 1971 SCIENTIFIC AMERICAN, INC

resulting electrical signal, and finally the display system that sequentially displays the information at the proper "address" in the image [*see illustration on preceding page*].

The electron beam of a scanning electron microscope usually originates at a heated tungsten cathode, typically held at 20,000 volts below ground potential. The beam is accelerated and focused by carefully shaped electrical and magnetic fields and must in addition be demagnified. Without demagnification the beam, "looking" back into the electron gun, would have a diameter of between 250,000 and 500,000 angstroms. The demagnification to approximately 100 angstroms is accomplished by electron lenses (magnetic fields) arranged to duplicate the effect observed in looking through the wrong end of a telescope [*see upper illustration below*]. An inevitable consequence of demagnification is that much of the radiation leaving the electron gun does not reach the sample. The electrons are intercepted by apertures required to minimize lens aberration. As a result the current of the electron beam is reduced from about 10^{-4} ampere at the electron gun to perhaps 10^{-10} or 10^{-12} ampere at the sample. (A beam of 10^{-4} ampere provides about 10^{15} electrons per second, while a beam of 10^{-12} ampere provides only about 6×10^{6} electrons per second.)



DEMAGNIFYING ELECTRON OPTICS must be used in a scanning electron microscope to reduce the size of the electron source from its actual diameter of 20 to 50 microns (250,000 to 500,000 angstroms) to a desired beam diameter of about 100 angstroms. The demagnification is accomplished by using two or more magnetic lenses in series, each capable of demagnifying the beam by a factor of from 20 to 100. The result is analogous to sending light backward through a telescope. Much of the radiation focused by a given lens misses the next lens, thus reducing the number of electrons reaching the sample.



TWO KINDS OF DETECTOR are commonly used in scanning electron microscopes, one to collect backscattered electrons, the other to collect secondary electrons. Backscattered electrons are high-energy electrons from the primary beam that have been turned around within the sample. Secondary electrons, produced near the surface of the sample by the impact of the primary beam and by the backscattered electrons, have less energy and are readily attracted to a collector held about 200 volts positive with respect to ground.

Other fundamental limitations of electron optics and electron emission help to determine the ultimate performance of the probe-forming system. For example, electron lenses that are rotationally symmetrical have certain inherent aberrations, notably spherical aberration and chromatic aberration, that prevent the designer from making the convergence angle of the beam too large [*see top illustration on opposite page*]. As the convergence angle is increased the probe current increases. This is desirable for improving the signal-to-noise ratio but undesirable in that the spot size also increases and the depth of field decreases. On the other hand, if the convergence angle is decreased, the beam current likewise decreases, and for convergence angles that are small enough the resolution is limited by diffraction effects.

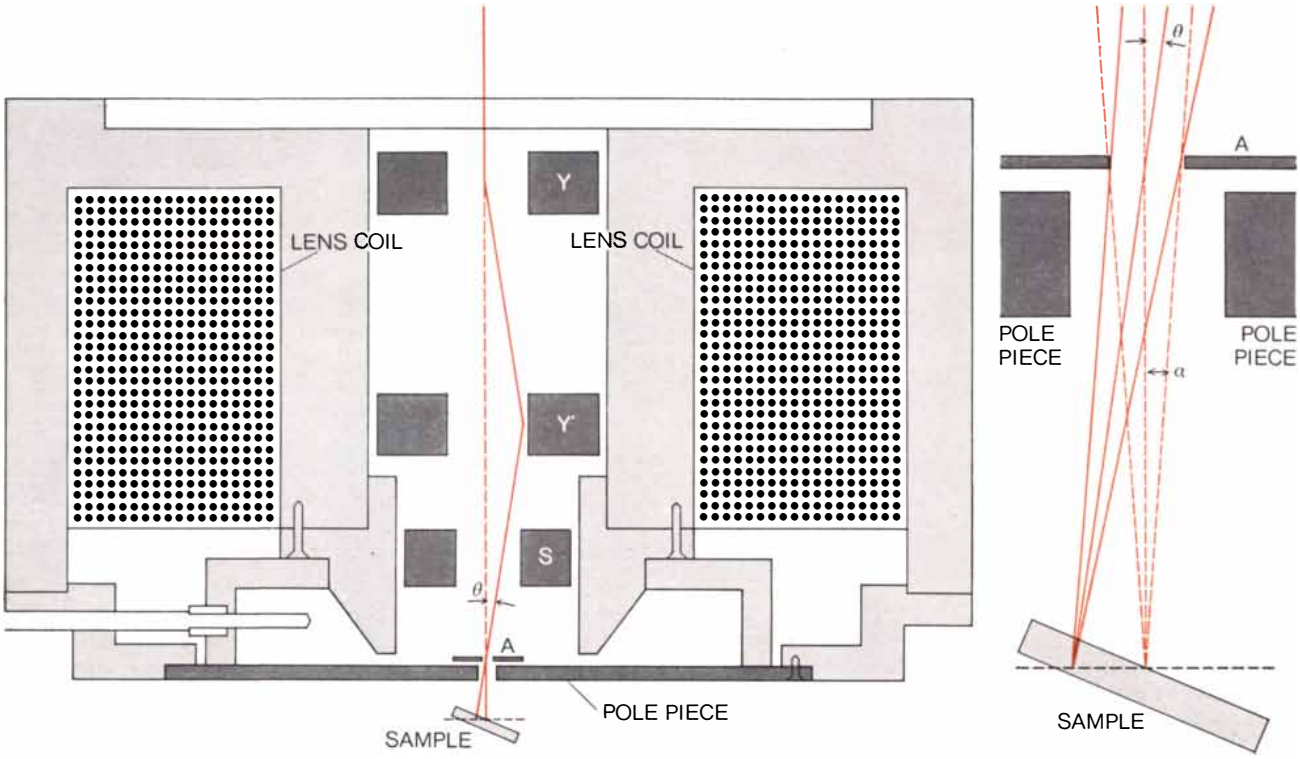
An optimum value for the beam convergence angle is between .005 and .01 radian, which means that the ratio of the focal length to the aperture diameter of the electron-optical system lies between 100 and 50. In ordinary cameras this ratio, or f number, commonly falls between 1.4 and 22. As most photographers know, the higher the f number, the greater the depth of field.

The amount of current that can be focused into the final spot is directly proportional to the current density available at the cathode, typically between one ampere and 10 amperes per square centimeter. To achieve high signal-to-noise ratios at the highest possible resolution, special high-density sources have been developed to improve the current density available at the electron gun. At the Thomas J. Watson Research Center of the International Business Machines Corporation, Alec N. Broers has developed a lanthanum hexaboride cathode that is about 100 times better than the normal tungsten cathode. His cathode has been installed on several commercial scanning electron microscopes.

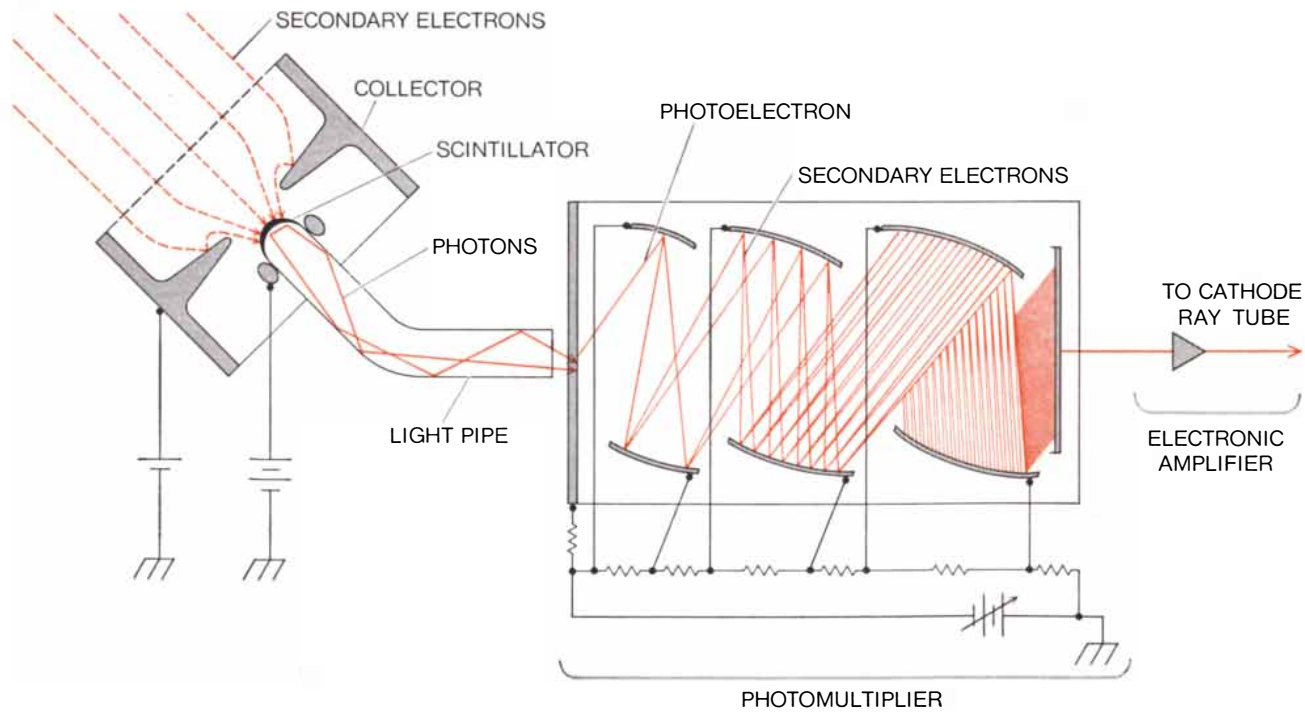
At the University of Chicago, Albert V. Crewe and his co-workers have incorporated a field-emission source with an even higher emission-current density in a scanning electron microscope. Although the cathode requires an ultra-high-vacuum environment for stable operation, a resolution of about five angstroms is possible in Crewe's microscope when operating in the scanning-transmission mode on ultrathin specimens. This instrument has detected individual uranium atoms in special preparations [see "A High-Resolution Scanning Electron Microscope," by Albert V. Crewe; SCIENTIFIC AMERICAN, April, 1971].

Although the probe that scans the

© 1971 SCIENTIFIC AMERICAN, INC

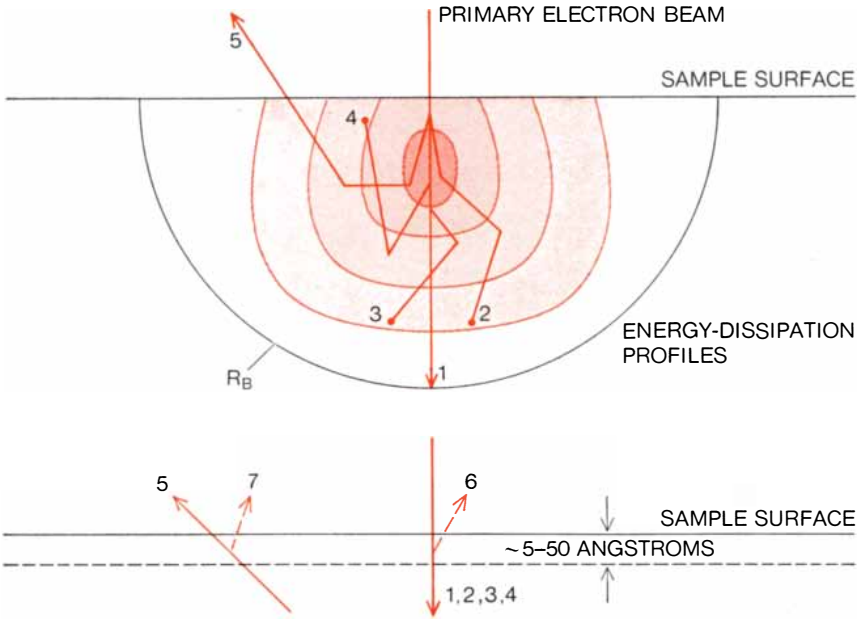


ROTATIONALLY SYMMETRIC FINAL LENS with built-in deflection system is the heart of the electron optics in a scanning electron microscope. As the electron beam passes through the final lens (*left*) it is first deflected away from the axis by one deflection yoke (*Y*), then back toward the axis by a second yoke (*Y'*) so that it crosses the axis at the final aperture (*A*) at an angle theta (θ). The stigmator (*S*) corrects for residual asymmetries. The diagram at the right shows how the beam is focused on the object (*broken lines*) and then both deflected and focused (*solid lines*). Because the angle alpha (α) at which the beam converges is usually less than half a degree, the beam diameter does not increase greatly either slightly in front of or slightly in back of the plane of best focus. This gives the imaging system great depth of field so that all parts of an object appear to be about equally in focus.

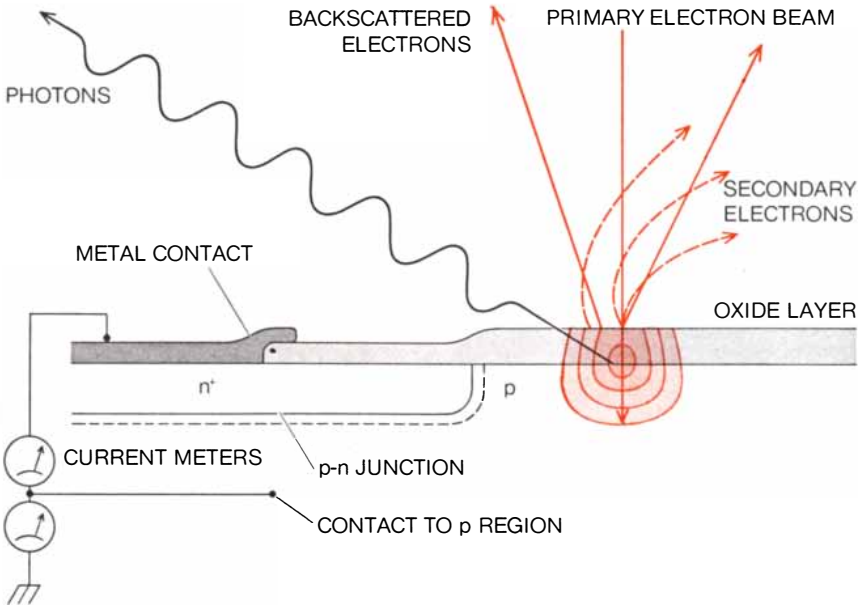


DETECTION AND AMPLIFICATION SYSTEM in a scanning electron microscope introduces a minimum of noise. The flow of secondary electrons entering the collector can be as small as 10^{-12} ampere. The secondary electrons are accelerated to about 10,000 volts before they strike the scintillator, where each electron produces many photons. The photons are guided to a photomultiplier where photoelectrons are excited. Each photoelectron triggers a cascade yielding from 100,000 to 50 million additional electrons.

© 1971 SCIENTIFIC AMERICAN, INC



TRAJECTORIES OF ELECTRONS inside a sample are statistical and thus can be predicted only on the average. Typical trajectories (*upper diagram*) are labeled *1* through *5*. The various contour lines show depths at which scattered electrons lose equal amounts of energy. The outermost contour, known as the Bohr-Bethe range (R_B), denotes the greatest distance a primary electron can travel in the sample. Trajectory *5* represents a backscattered electron. The probability of backscattering increases with the atomic number of the target material. Secondary electrons can be generated at any point along the path of a scattered primary electron. Only the secondary electrons generated near the surface of the sample are likely to have enough energy to escape (*6 and 7 in lower diagram*) and thus to contribute to the usual image produced by a scanning electron microscope. Since secondary electrons produced by backscattered electrons (7) can emerge as far as one or two micrometers from the point of entry of the primary beam, they tend to degrade the resolution of the image.



ADDITIONAL INFORMATION SIGNALS can be generated when the sample is an electronic device incorporating a semiconductor or an insulator. The diagram shows the kinds of signal one can measure if the sample is a transistor containing a *p-n* junction. Like other samples, it will yield secondary electrons, backscattered electrons and photons of various frequencies (X rays, ultraviolet radiation, visible light, infrared radiation). In addition semiconductors and insulators undergo a sharp change in conductivity when struck by electrons. In silicon each incident electron can produce 5,000 or more charge carriers in the form of electron-hole pairs. The upper meter would show the current flow induced across a *p-n* junction; the lower meter would measure the flow from the device to ground.

sample in a scanning electron microscope must be sharply focused, the secondary electrons that produce the actual image need not be. This seeming paradox disappears if one recalls that we are building up the image as a series of sequential picture elements. Points on the object and on the image correspond to one another because they are struck by the microscope electron beam and the cathode ray tube electron beam at the same instant.

Collecting Secondary Electrons

When the electrons in the probe strike a thick specimen, several things can happen [*see top illustration at left*]. As the electrons penetrate into the specimen, they lose energy and slow down. They may also be scattered through large angles. Finally, some of the electrons may reemerge from the top surface, traveling in a direction more or less opposite to that of the incident beam. By collecting these "backscattered" electrons one can obtain information about the kinds of atom in the specimen. In principle, however, any signal generated when the beam strikes a given point in the object, regardless of its form, can be used to produce an image.

The image one normally uses is produced by the secondary electrons: electrons excited within the specimen by the high-energy electrons in the probe. The probability that the secondary electrons will escape from the sample decreases exponentially as their point of generation moves away from the surface. As a result those that do emerge come from a surface layer that is only five to 50 angstroms thick for metals and somewhat thicker for insulators. These secondary electrons can be thought of as spraying out of the surface in various directions with energies of a few electron volts.

The secondary electrons are drawn to a collector held at a potential of about 200 volts positive with respect to ground [*see bottom illustration on preceding page*]. There is no need to focus the secondary electrons because those collected at any given instant all come from the immediate vicinity of the point where the probe has entered the specimen. At the highest resolution only those secondary electrons excited by the primary beam entering the sample provide a useful signal; the secondary electrons excited by backscattered electrons leaving the sample produce a background that lowers the quality of the image. The strength of the signal produced at the collector is proportional to the number

© 1971 SCIENTIFIC AMERICAN, INC

of secondary electrons impinging on it. The signal is fed to a cathode ray tube and modulates the tube's electron beam as it moves across the tube face synchronously with the path of the electron probe across the surface of the sample. This produces the scanning electron microscope image; a photograph of the image on the face of the tube is a scanning electron micrograph.

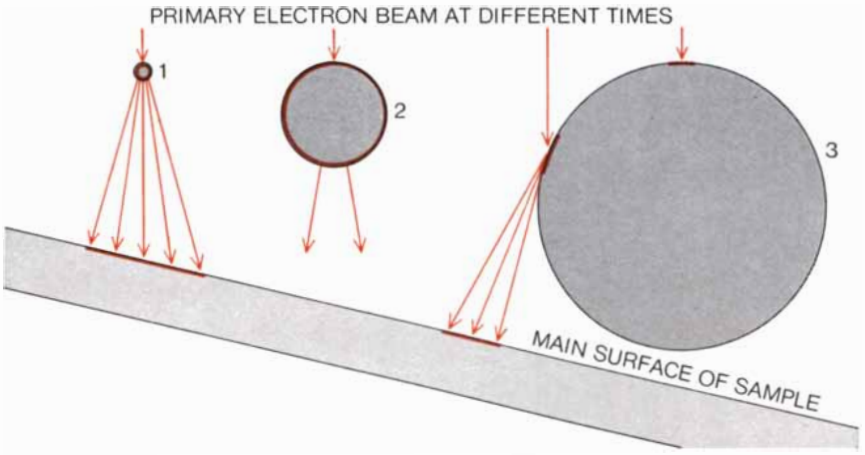
Information and Imagery

Questions that must be asked and answered in scanning electron microscopy include the following: How large a region does the information come from when the beam strikes the sample? How fast can the information be generated and with what efficiency? What geometrical, physical, chemical or other properties of the sample determine the information?

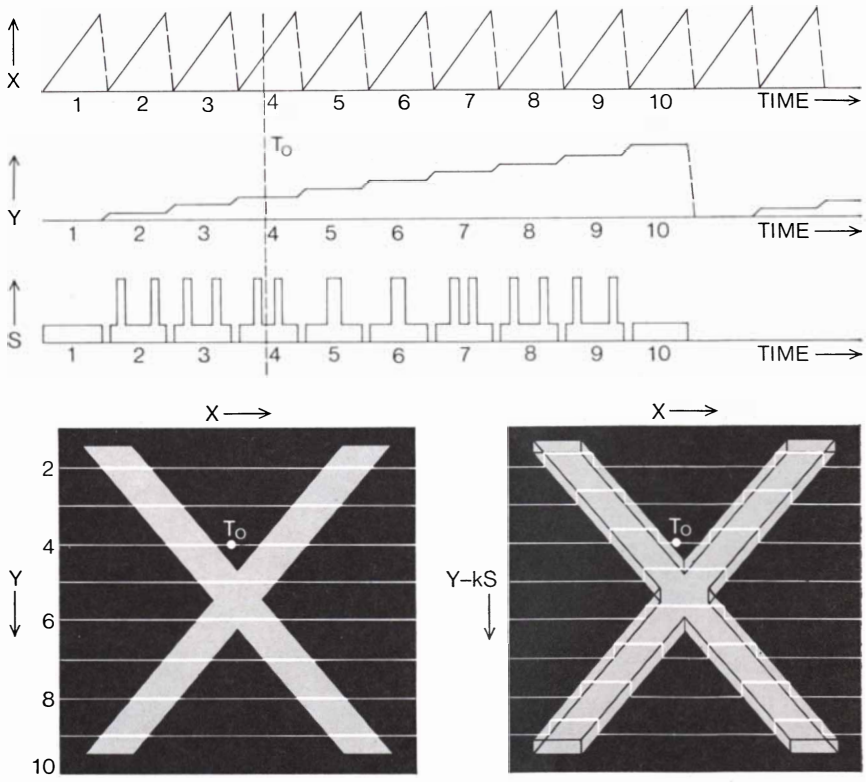
If the image is supplied by secondary electrons, which is normally the case, one knows such electrons have a short mean free path in solids: about 10 angstroms in metals and about 50 angstroms in insulators. Many of the secondary electrons emerge from a circle whose diameter is at most 20 angstroms larger than the diameter of the electron probe itself.

Other secondary electrons, produced by backscattered electrons leaving the sample, emerge from different points within a larger circle, about one to five microns in diameter for normal beam energies. Secondary electrons are generated in less than 10^{-10} second, so that there are essentially no late-emerging electrons to blur the image when the probe has moved on to the next point. For each 10 primary electrons striking a sample between one secondary electron and 20 are generated, depending on the sample material and the angle between the beam and the surface of the sample. X rays, which can also be collected to produce an image, are generated much less efficiently. Between 1,000 and 100,000 primary electrons are needed to give rise to a single X-ray photon, the exact number again depending on the sample material. X rays provide a useful signal, however, because they can reveal what elements are in the sample. For this reason sensitive X-ray detectors are finding wide application in scanning electron microscopy.

If the sample is a semiconductor or an insulator, its electrical conductivity will usually be sharply increased by electron-beam bombardment. For this reason the scanning electron microscope is extremely useful in the examination of integrated circuits, which often incorpo-



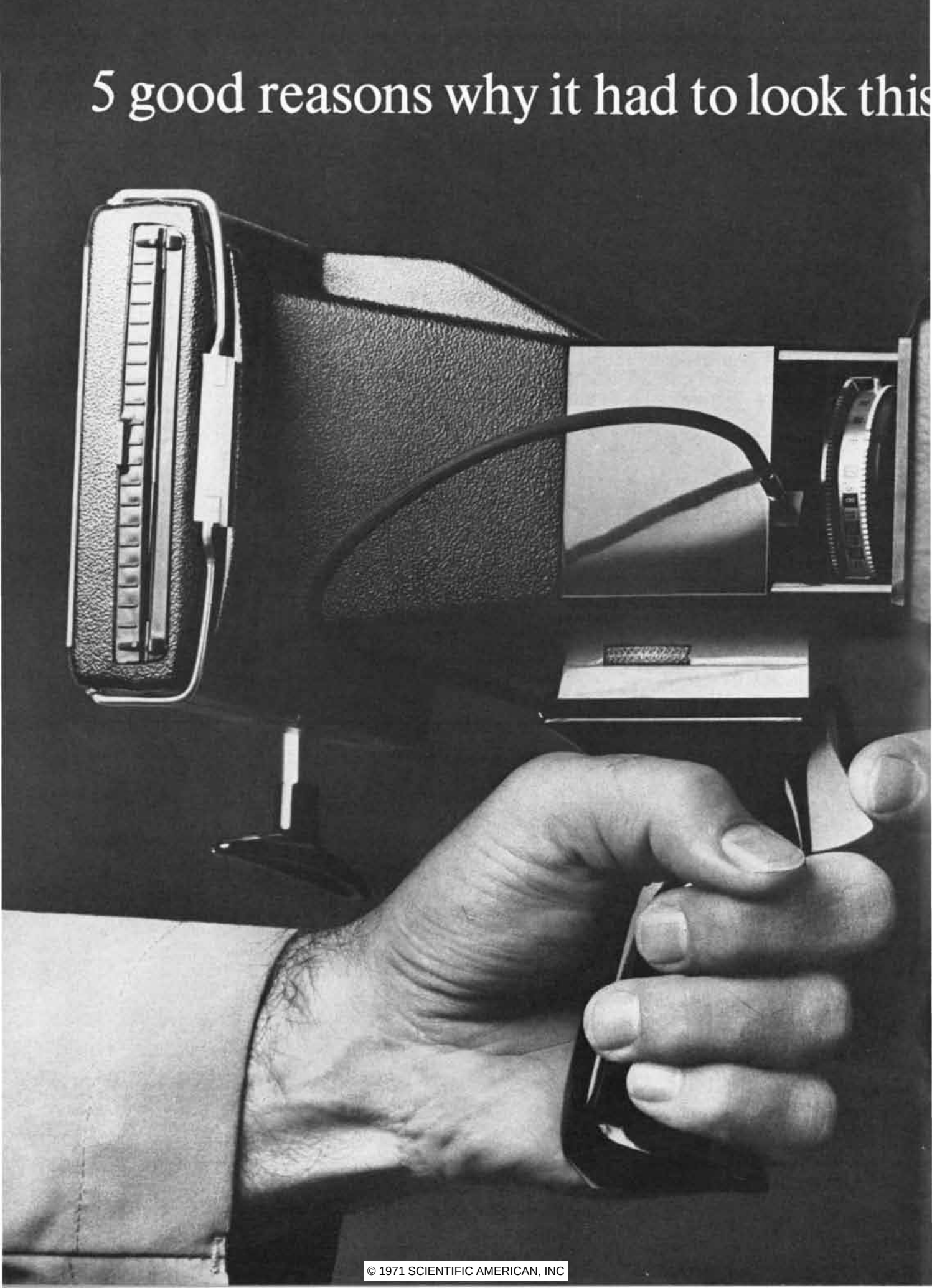
WHEN INTERPRETING IMAGES produced by a scanning electron microscope one must be aware that secondary electrons can arise from regions at some distance from the immediate area struck by the primary electron beam. The diagram shows what happens when the beam strikes fibers of different sizes: one much smaller than the average penetration range (*1*), one whose diameter is about .6 times that range (*2*) and one much larger (*3*). In the first case primary electrons easily penetrate the fiber and are scattered, thus generating secondary electrons (*color*) at three surfaces: the top of the fiber, the bottom of the fiber and the surface of the sample. The fiber will therefore seem brighter than it would if secondary electrons were generated only at its top surface. The second fiber absorbs most of the incident electrons but secondary electrons will escape from its entire surface, also making the fiber look brighter than might be expected. The third fiber is so large that it stops all electrons that strike the center. Electrons that strike the edge, however, tend to glance off and spray a sizable area below. This often makes the edge of a large fiber look particularly bright.



SCANNING SCHEME used in scanning electron microscopy is exactly like that used in conventional television except that a typical micrograph contains 1,000 horizontal lines compared with 525 lines in commercial television. Here the two coordinates of the display are labeled *X* and *Y*; the signal is denoted by *S*. The configuration of signal *S* in each time slot represents the spacing of on and off values needed to produce one of the 10 horizontal lines in portraying a large block *X* (*lower left*). The *X* would look continuous, of course, if 1,000 lines were used. The picture at the left is an intensity-modulated display. One can also deflect the electron beam from its normal raster position, *Y*, by an amount proportional to the signal strength ($Y - kS$), producing a deflection-modulated display (*lower right*).

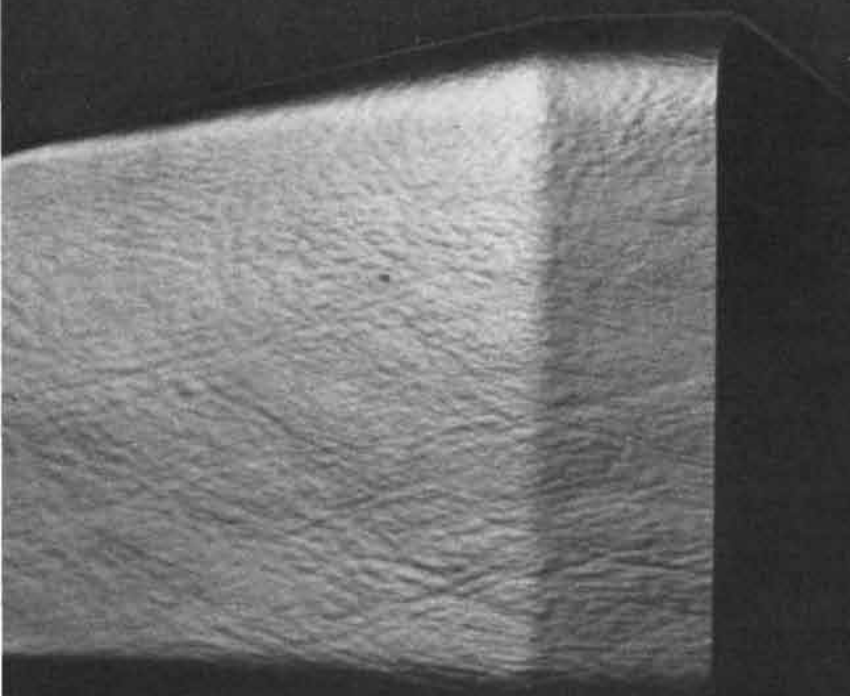
© 1971 SCIENTIFIC AMERICAN, INC

5 good reasons why it had to look this



© 1971 SCIENTIFIC AMERICAN, INC

way (the low price wasn't one of them).



Polaroid's new CR-9 Land Oscilloscope Camera. Under $180.

The CR-9 looks the way it looks because it's totally functional.

It's the most portable of all. It weighs under 24 ounces. There's no need to permanently mount it. And the pistol grip has a trigger-like shutter release. So one hand can easily operate it.

It's versatile. It fits almost any oscilloscope. There are 8 different hood sizes available. Just specify the size you need.

There's no focusing. You just place the hood against the display and shoot. It focuses for you. And frames the shot for you. And the controls for shutter speed and lens opening are easy to get at.

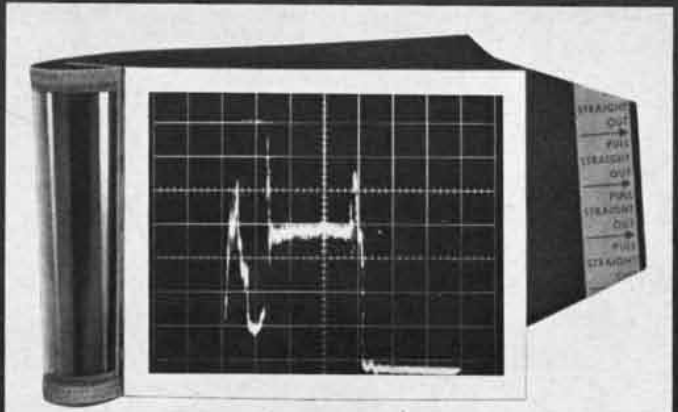
You get high-quality prints. The hood is light tight. And keeps film and display in parallel planes. And there's a 70 mm f 5.6 oscilloscope recording lens. So you get a virtually distortion-free record.

You get the print in 15 seconds. Because it's designed to use standard Polaroid Type 107 3000-speed Land film. And you can photograph high-speed, transient phenomena. Easily.

Best of all, it's the lowest priced oscilloscope camera of all. And not by accident. It was by design.

For more information, write to: Polaroid Corporation, Dept. 57-219, Cambridge, Mass. In Canada: 350 Carlingview Drive, Rexdale, Ont. In other parts of the world, contact the nearest Polaroid subsidiary or distributor.

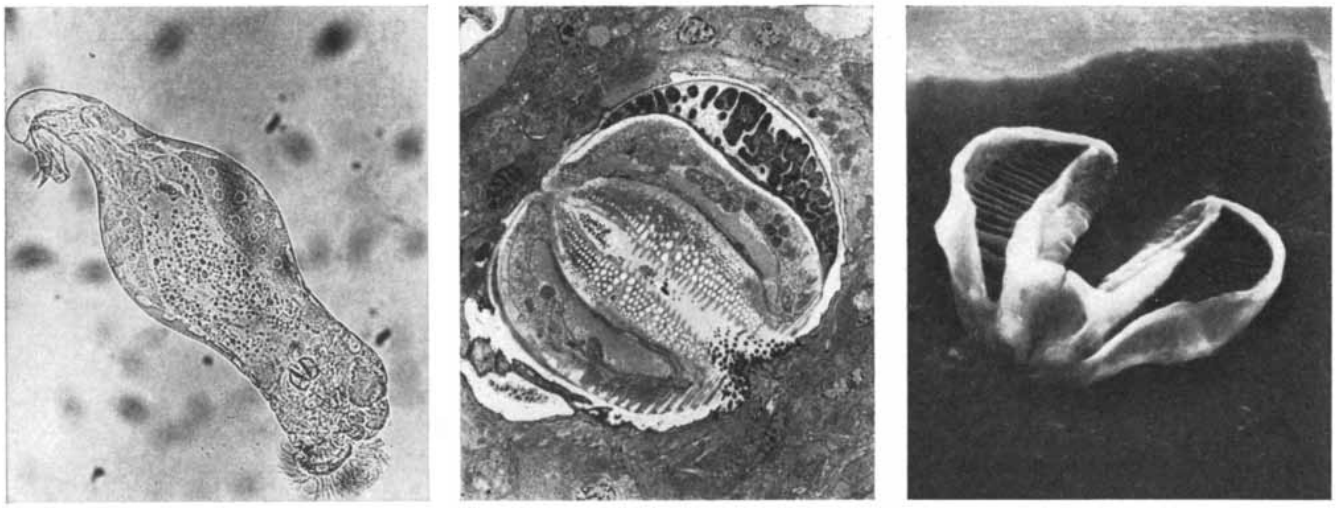
For the record, we want you to get an accurate picture.



This CR-9 recording is as distortion-free as you'd get with a $700 camera.

Polaroid® Price subject to change without notice.

© 1971 SCIENTIFIC AMERICAN, INC

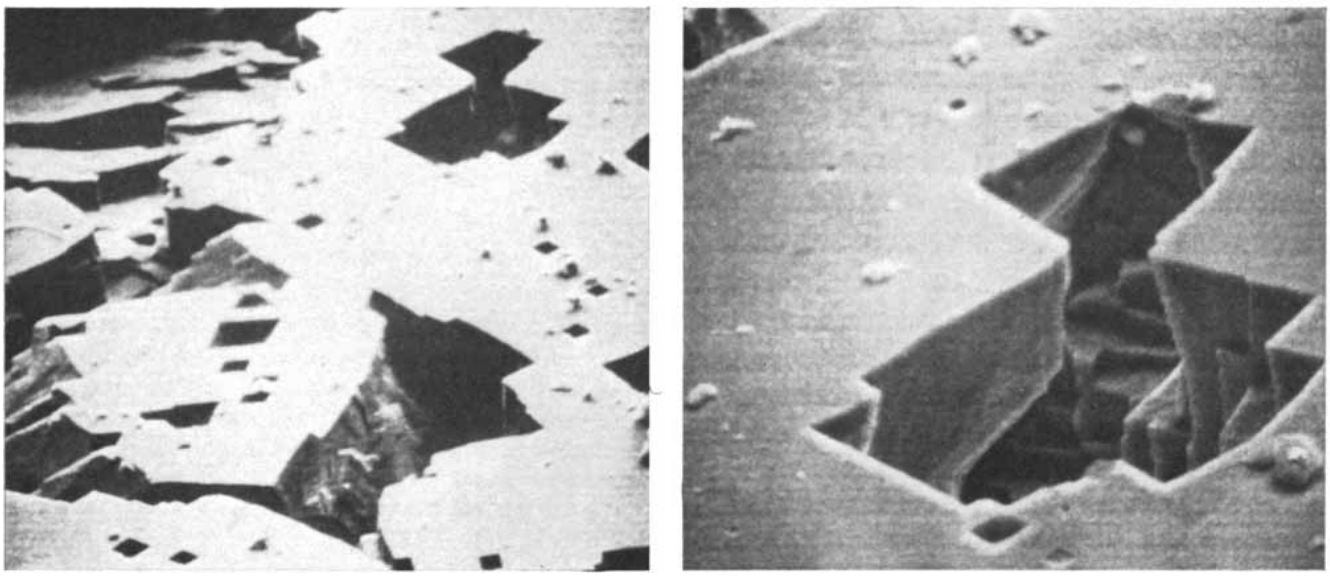


THREE KINDS OF MICROSCOPE yield widely differing images. A standard light microscope produced the picture of a rotifer, magnified 240 diameters, shown at the left. A transmission electron microscope produced the image of a thin section through the animal (*middle*); the magnification is 2,400 diameters. The close-up of the rotifer's jaws (*right*), also magnified 2,400 diameters, was produced by a scanning electron microscope. The first two micrographs were made by James K. Koehler of the University of Washington.

rate thousands of transistors or other circuit elements in a chip only a tenth of an inch square. In silicon, for example, an incident beam of 18,000-volt electrons will produce about 5,000 electrons and 5,000 "holes." These current-carriers are produced in pairs, and if they are detected by a *p-n* junction in the integrated circuit, the signal current can be much stronger than the current of the incident beam. The designer of the circuit can tell from such signals if his device is functioning as it has been designed, if the relative positioning of succeeding process steps was within tolerance limits and so on. Moreover, he can inspect the surface geometry of the circuit chip to see if its metal conductors have the desired integrity and quality [*see illustration on page 69*].

The resolution provided by the scanning electron microscope is of course much higher than that provided by the light microscope, which formerly served as the standard inspection tool. For the next generation of electron devices the beam of the scanning electron microscope is being used to make the patterns of the devices themselves. The dimensions of these patterns are much smaller than those of their predecessors, which were created by light-optical methods.

What accounts for the distribution of light and dark areas that gives the impression of a three-dimensional surface in a typical scanning electron micrograph? First of all, the sample is usually not held perpendicularly to the scanning beam but is tilted at an angle of from 30 to 60 degrees. For heavily contoured biological specimens this is often not necessary. The probability that second-



TWO DIFFERENT KINDS OF SIGNAL were used to make these scanning electron micrographs of a heavily etched surface of aluminum. The image at the left was made with highly energetic backscattered electrons that travel virtually in straight lines. The holes in the surface of the aluminum appear black because the electrons escaping from the holes in a straight trajectory cannot be "seen" by the detector of backscattered electrons. The micrograph at the right shows a close-up of one of the jagged holes, easily recognizable from its shape, when the image is produced by secondary electrons. The interior is clearly depicted because secondary electrons emerge with low energy and their trajectories bend readily to strike a detector with a positive potential of a few hundred volts.

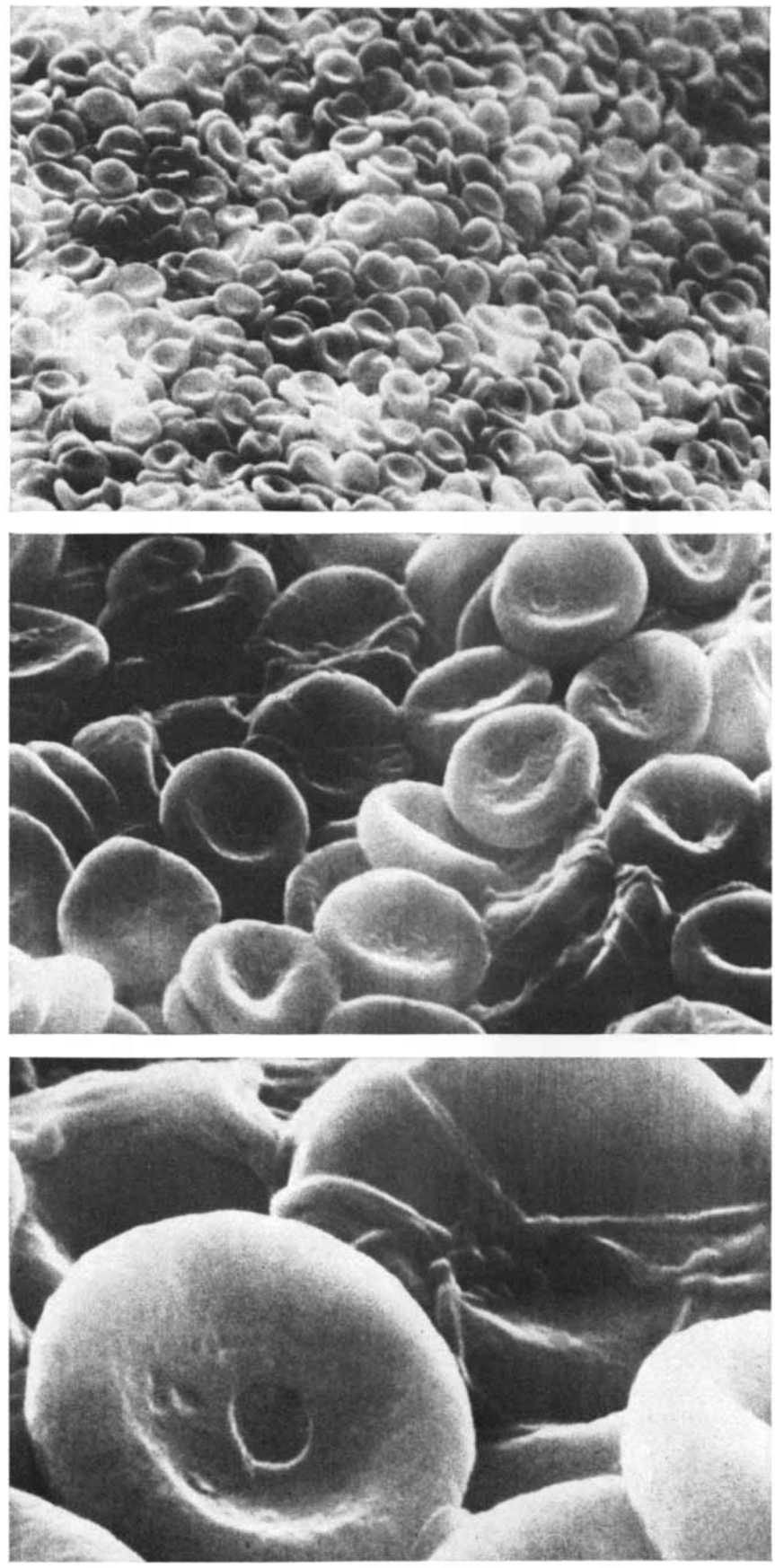
© 1971 SCIENTIFIC AMERICAN, INC

ary electrons will be generated increases as the angle between the incident beam and a line perpendicular to the surface of the sample at the point of incidence increases. In other words, the beam generates more secondary electrons when it strikes a sloping surface or a sharply curving edge than when it strikes a flat surface. Other effects are subtler and involve the way structures lying above a primary surface either absorb or deflect the electrons in the primary beam. For example, a small fiber can look brighter than one would expect on the basis of size alone [*see top illustration on page 61*].

These effects let the viewer infer a three-dimensional structure from a two-dimensional micrograph. To determine whether or not the inference is correct, stereoscopic pairs of images can be created easily by making two micrographs in which the angle of the sample to the incident beam is changed by several degrees. When viewed through a stereoscopic viewer, the resulting image is truly three-dimensional.

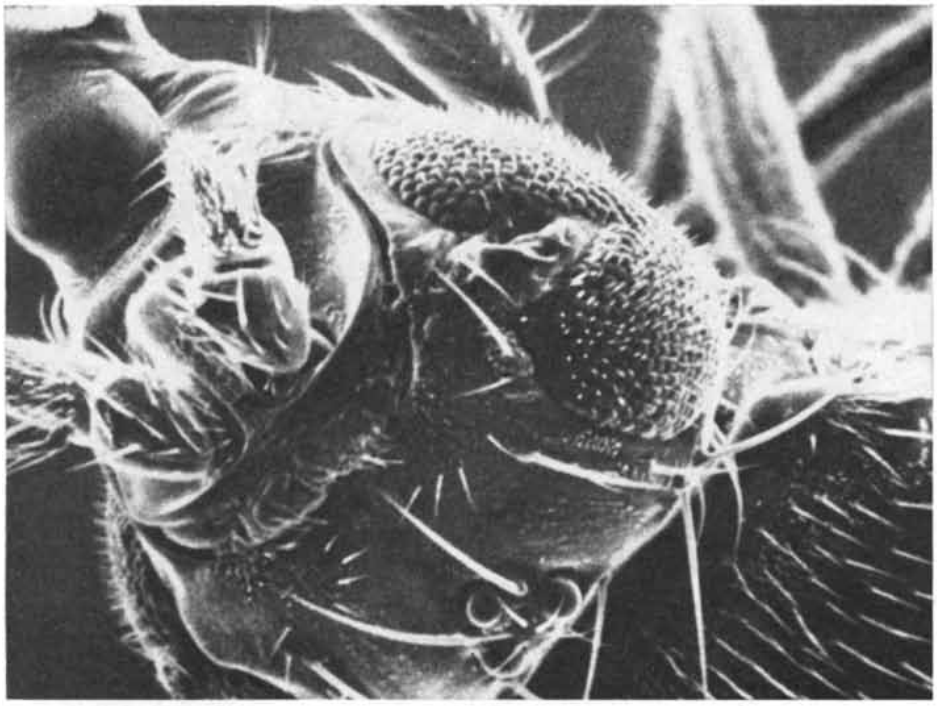
Biological specimens, which account for a large variety of samples inspected with the scanning electron microscope, can be prepared in several forms. One type of specimen consists of the biological tissue or organism complete and nonsectioned. Since the specimen must be placed in the instrument's vacuum chamber, water must be carefully removed. Freeze-drying and critical-point drying are the most often employed techniques for removing water without introducing serious artifacts. Occasionally simple air drying can be employed if the sample is resistant to deformation or if collapse of the surface is desirable to indicate underlying structures in the specimen. Fixation, staining, washing, drying and coating are all used at times in the preparation of samples. Since the information desired is often the topological relations among component parts of a microsystem, rather than the precise dimensions of a particular part, artifacts such as shrinkage can be tolerated, particularly if alternative microscopic methods are available to establish actual dimensions.

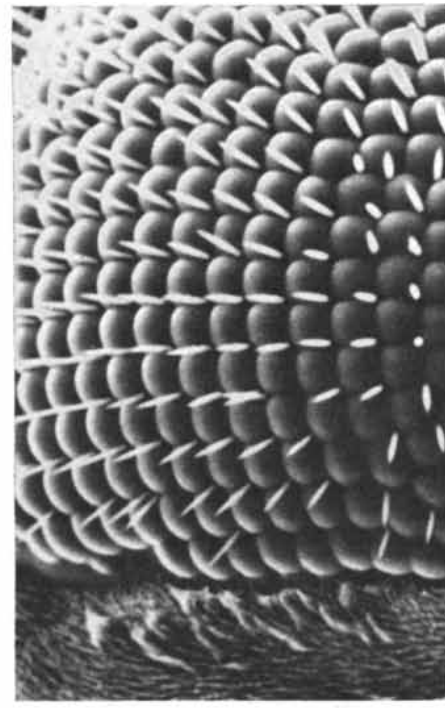
If one wants to compare images provided by a light microscope with those provided by a scanning electron microscope, one can readily mount the specimen on a glass slide that can be inserted into both instruments. With careful dissection techniques one can expose deeply lying sections of bulk specimens so that their topology can be studied by scanning electron microscopy. It is also possible to examine specimens that have been sectioned as they would be for or-



HUMAN BLOOD was allowed to clot in moist air, fixed in formaldehyde and viewed in the scanning electron microscope. Reading from top to bottom the sample is magnified 1,500, 5,000 and 15,000 diameters. The disks are red blood cells, held in a meshwork of fibrin strands. The cells are somewhat shrunken. The pictures were made by L. McDonald.

© 1971 SCIENTIFIC AMERICAN, INC



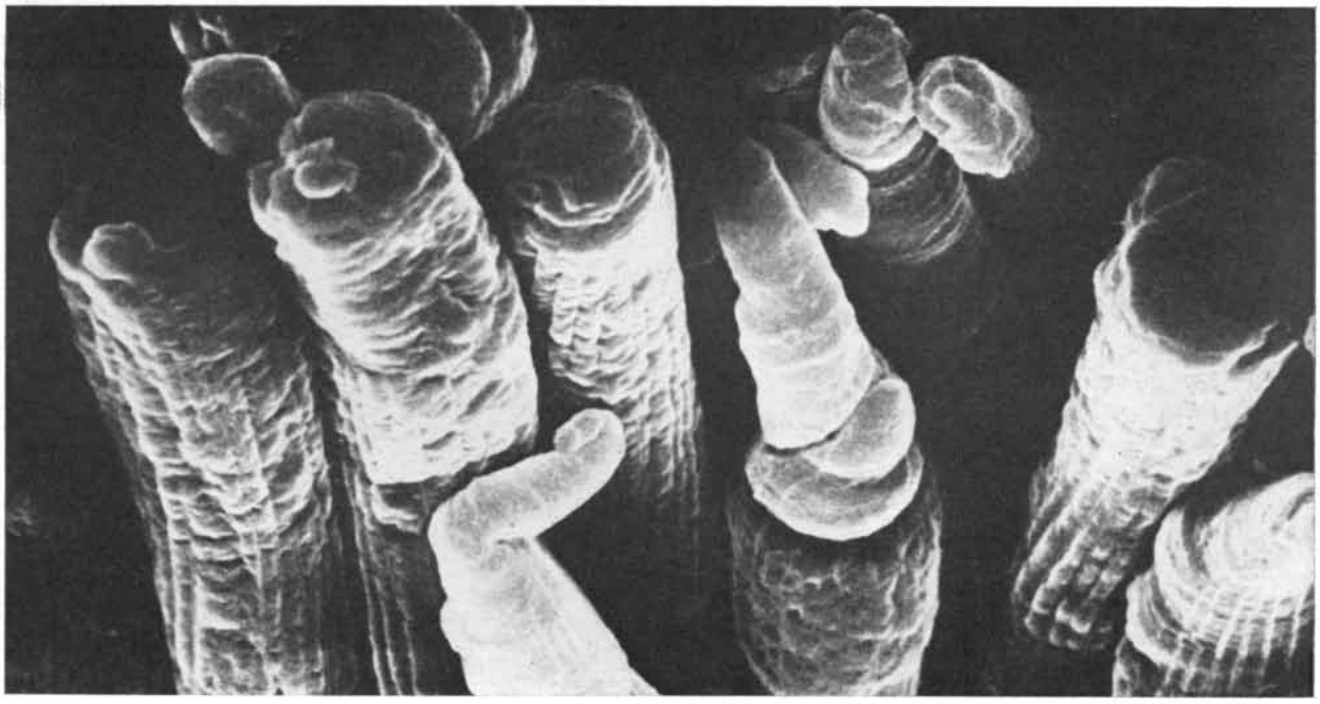


"ZOOM" VIEW OF FRUIT FLY, *Drosophila melanogaster*, was made with a scanning electron microscope. The fruit fly can be viewed in the microscope while still alive so that no metal coating or other preparative technique is needed. The three micrographs

dinary light or electron microscopy.

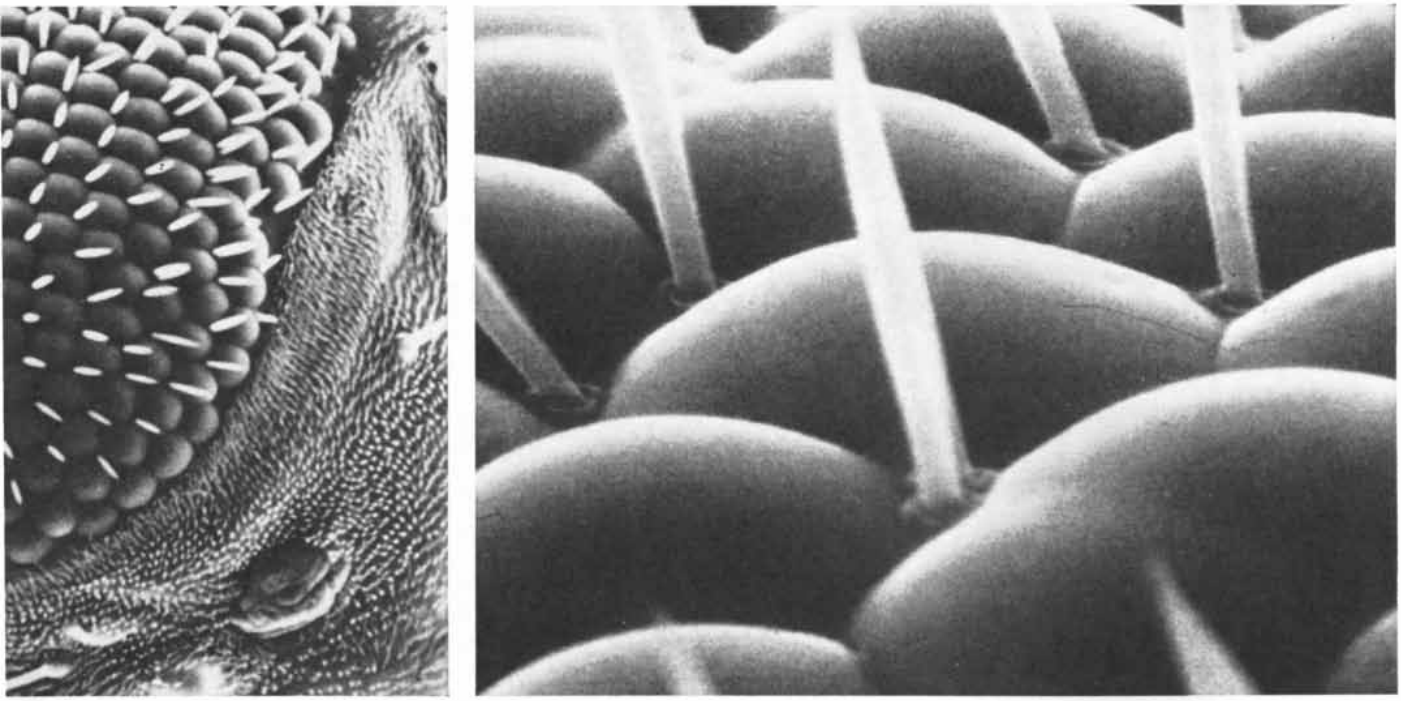
It may seem surprising, but there are even a few organisms, chiefly insects, that can be observed in the scanning electron microscope while they are still alive. The flour beetle (*Tribolium confusum*) is one of these hardy organisms [*see illustration on page 54*]. Individuals in all stages in the life cycle of the beetle (egg, larva, pupa and adult) are able to hold enough water to survive the high vacuum of the microscope's specimen chamber for as long as an hour. Since the living specimens act as conductors under low-current electron bombardment, they do not have to be coated with metal or otherwise altered.

The most striking attribute of the images presented by the scanning electron microscope is their ease of interpretation. An insect looks as an insect



ARCHITECTURE OF PHOTORECEPTORS in the retina of a salamander are revealed by the scanning electron microscope at a magnification of 2,500 diameters. Rods appear as cylinders; cones are tapered. The overall structure of these cells is difficult to make out in conventional microscopes since they cannot be viewed intact. The micrograph was made by E. R. Lewis and Y. Y. Zeevi.

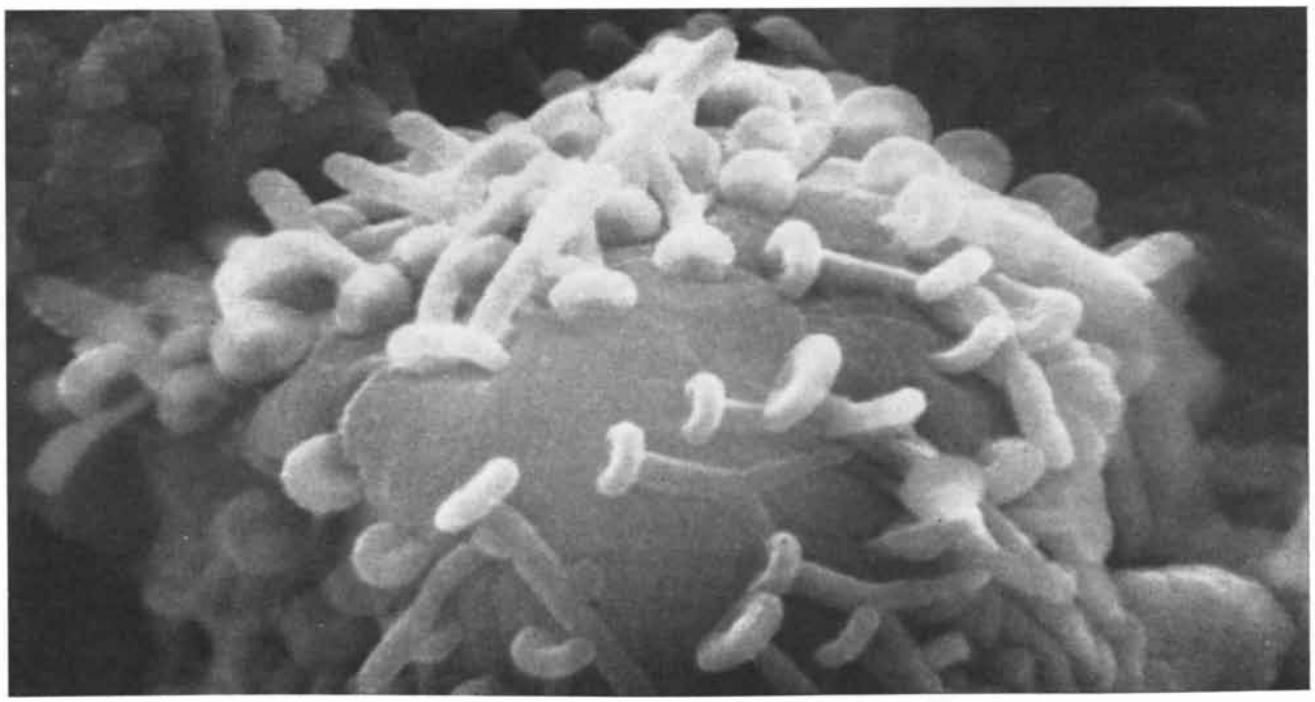
© 1971 SCIENTIFIC AMERICAN, INC



from left to right are at successive magnifications of 180, 450 and 4,500 diameters. The first micrograph shows the fly's entire head; subsequent views depict the eye alone in greater and greater detail. Micrographs were made by authors' colleague Hyman Hartman.

should; hence we trust our eyes when we see the insect as it is revealed by scanning electron microscopy. This faith is usually justified for a wide variety of objects when secondary electrons provide the signal. If two parts of the sample are at different voltages, however, the secondary electron collection from the two parts is different, changing the contrast of the image in unfamiliar ways. One must become familiar with these changes and consciously force one's visual system to interpret the image in a fresh way. If the signal used to produce the image is created not by secondary electrons but by X rays or currents generated by the primary electron beam, the resulting picture may bear no relation to topography; again one must learn to interpret unfamiliar information.

In addition the signal can be coded

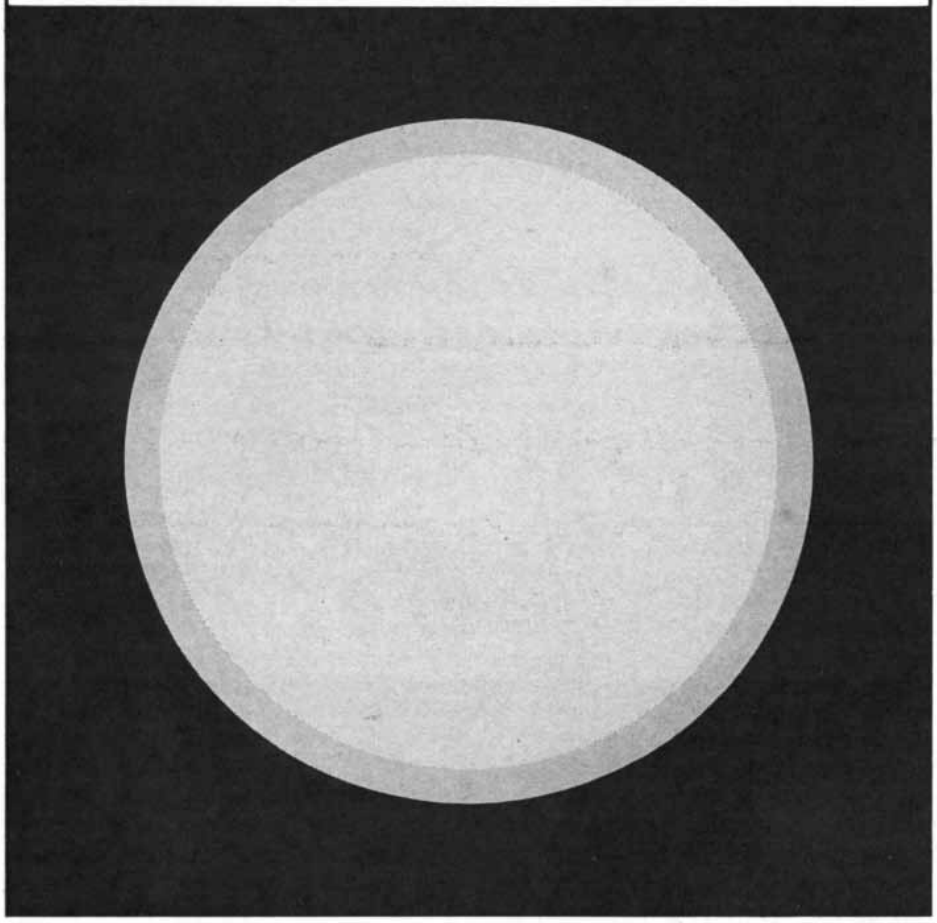


KNOBS ON NERVE CELL of a sea hare (*Aplysia*) are believed to be associated with synaptic junctions, where nerve impulses pass from one cell to another. This single view at a magnification of 11,000 diameters tells more about certain aspects of nerve tissue than one could glean from hundreds of serial sections of the same specimen. The scanning electron micrograph is by Lewis and Zeevi.

© 1971 SCIENTIFIC AMERICAN, INC

A SCIENTIFIC AMERICAN Book

Energy and Power



"This is just possibly the most important issue of any magazine published in recent years for the concerned layman, because the whole of it is devoted to the civilized world's most important subject—energy and power."
—*The Environmental Monthly,* September 1971

Now available in book form is the complete textual and illustrative content of the September 1971 issue of SCIENTIFIC AMERICAN, which placed in broad perspective the critical problem of our growing power consumption.

1971, 142 pages, illustrated, clothbound $6.50, paperbound $3.25

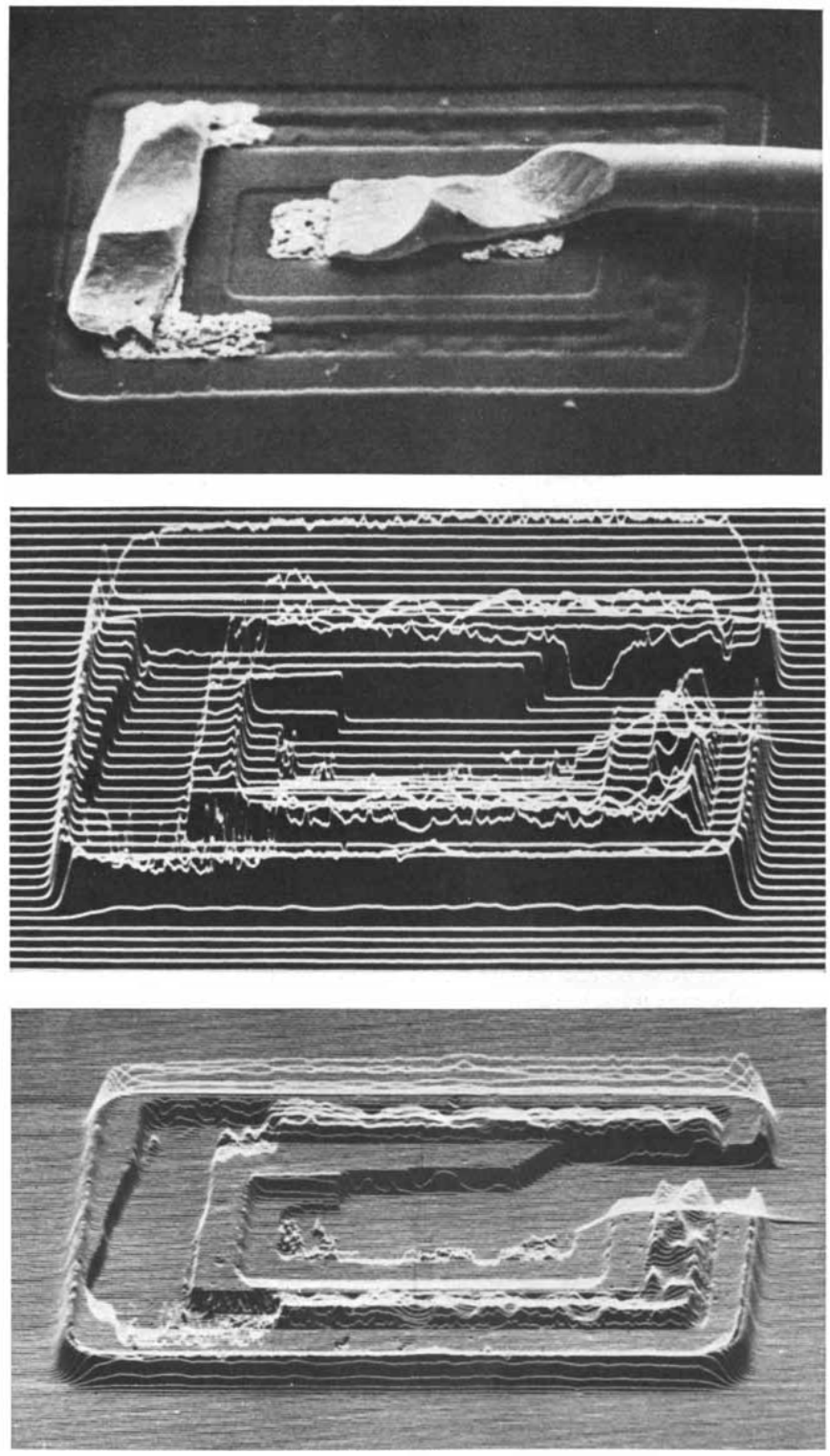
From your bookseller or Dept. B
W. H. FREEMAN AND COMPANY
660 Market Street, San Francisco, California 94104
58 Kings Road, Reading, England RG1 3AA

© 1971 SCIENTIFIC AMERICAN, INC

in artificial ways, for example by assigning different colors to represent different levels of intensity. Since the eye is more sensitive to differences in color than to slight variations in shades of gray, a color-coded image can reveal significant information that would otherwise be overlooked. At the same time one must be careful not to misinterpret the image by reading into it characteristics based on earlier and irrelevant visual experiences.

With these thoughts in mind, let us compare the information that can be obtained from the light microscope, the transmission electron microscope and the scanning electron microscope [*see top illustration on page 64*]. The first two instruments form images out of radiation passing through the object (or reflected from it) by means of a series of lenses. The contrast in the image comes from the scattering or absorption of photons or electrons or, at high resolutions, by shifts in the phase of the wave functions that represent photons or electrons. Because the energy of a photon of light is a few electron volts, the photon can interact with the chemical bonds of the sample, which have similar energies. This interaction provides color information, invaluable to the biologist or the pathologist, who can code the constituents of the cell by staining them in various ways. In transmission electron microscopy one can code selected parts of the sample with stains consisting of heavy metals whose massive nuclei scatter the incident electrons much more strongly than the light nuclei present in most biological samples do.

The scanning electron microscope differs from both of the older instruments by not being limited to the incident radiation for the formation of an image. In such an instrument no focusing is necessary after the electron beam strikes the sample. Instead the corresponding points on the specimen and in the image are correlated in time, thereby allowing any signal generated by the primary beam to serve as a potential source of the image. These signals can easily be stored on magnetic tapes or disks for redisplay at a later time or for processing by computer. Although computer processing of images made by the scanning electron microscope is still in an exploratory stage, it appears promising. Since the various types of microscope supply different kinds of information, they complement one another rather than competing. Each has unique characteristics that provide revelations about the microstructure of the real world.



MICROGRAPH OF TRANSISTOR made with secondary electrons in a scanning electron microscope shows (*top*) that a gold lead has broken and that a gold-aluminum alloy has formed between the lead and the thin aluminum layer deposited on the surface of the transistor to form a contact base. The alloy is visible as a bright region adjacent to the gold lead. Although the lead is broken, current can still be injected into the base by means of the microscope's electron beam. The injected current causes a current to flow between the transistor's emitter and collector and the flow can be displayed as a deflected wave form. Many such deflection traces can be used to build up a deflection-modulation image. An image containing only a few deflection traces is preferred for quantitative analysis of transistor performance (*middle*). A picture with many traces (*bottom*) simulates a complex surface whose height is proportional to the current flow between emitter and collector. Such images help to reveal how the induced current and the physical structure are related.

© 1971 SCIENTIFIC AMERICAN, INC

GEOTHERMAL POWER

The pressure on energy resources has generated new interest in the earth's heat. The emphasis is on exploring for new geothermal areas and developing new ways to extract work from steam and hot water

by Joseph Barnea

An old source of power for man's work has begun to attract new interest. Natural underground reservoirs of steam and hot water are now being tapped on a significant scale, and it will come as a surprise to many people to learn that the harnessing of this geothermal energy has already reached an aggregate capacity of a million kilowatts in plants around the world. At the present rate of development it is likely that by the end of this decade the production of electric power from steam fields will be quadrupled.

The heat of many geothermal reservoirs comes from a large body of molten rock that has been pushed up into the earth's crust from great depths by geologic forces. This dome of magma heats the rocks in the crust near the surface, which in turn heats the water in fissured or porous rocks to a temperature of perhaps 500 degrees Fahrenheit. Being at depths of as much as six miles, the water is under high pressure and is therefore liquid. Where the hot water can escape through a fissure it begins to boil, and part of it flashes off as steam. The geothermal energy can be tapped by a well driven into the fissure or down to the porous layer.

Interest in this source of energy has quickened in the past few years. Recent explorations have revealed that the resource is larger and more extensive than had been supposed. A generation ago the hot springs and steam fields that had long been known in a few localities around the world were believed to be merely local freaks of nature. There is evidence now that reservoirs of steam and hot water are actually widespread in the earth's crust. Signs of their presence have been detected on most of the continents and on a number of islands. It seems possible that such fields will also be found under the seas. Some of the explored fields are known to hold large quantities of energy. A single steam field in northern California, the Geysers field, is estimated to have a potential capacity of three million kilowatts, and surveys that have been made in the Imperial Valley of southern California have indicated a potential of 20 million kilowatts in that area.

The incentive for undertaking a major effort to tap geothermal fields has been heightened by projects showing that in addition to electric power they can yield other useful products. The geothermal steam or hot water can be applied to desalting seawater, to heating houses, greenhouses and swimming pools and to providing nonelectrical energy for refrigeration and air conditioning. Moreover, the hot water itself is a source of extractable minerals and can serve to provide potable water. These additional dividends increase the economic attractiveness of investment for the exploitation of this great earth resource, which up to now has served man mainly as a resort attraction (in the form of health spas) and as a somewhat esoteric and certainly minor source of power supporting small generating plants at a few sites around the globe.

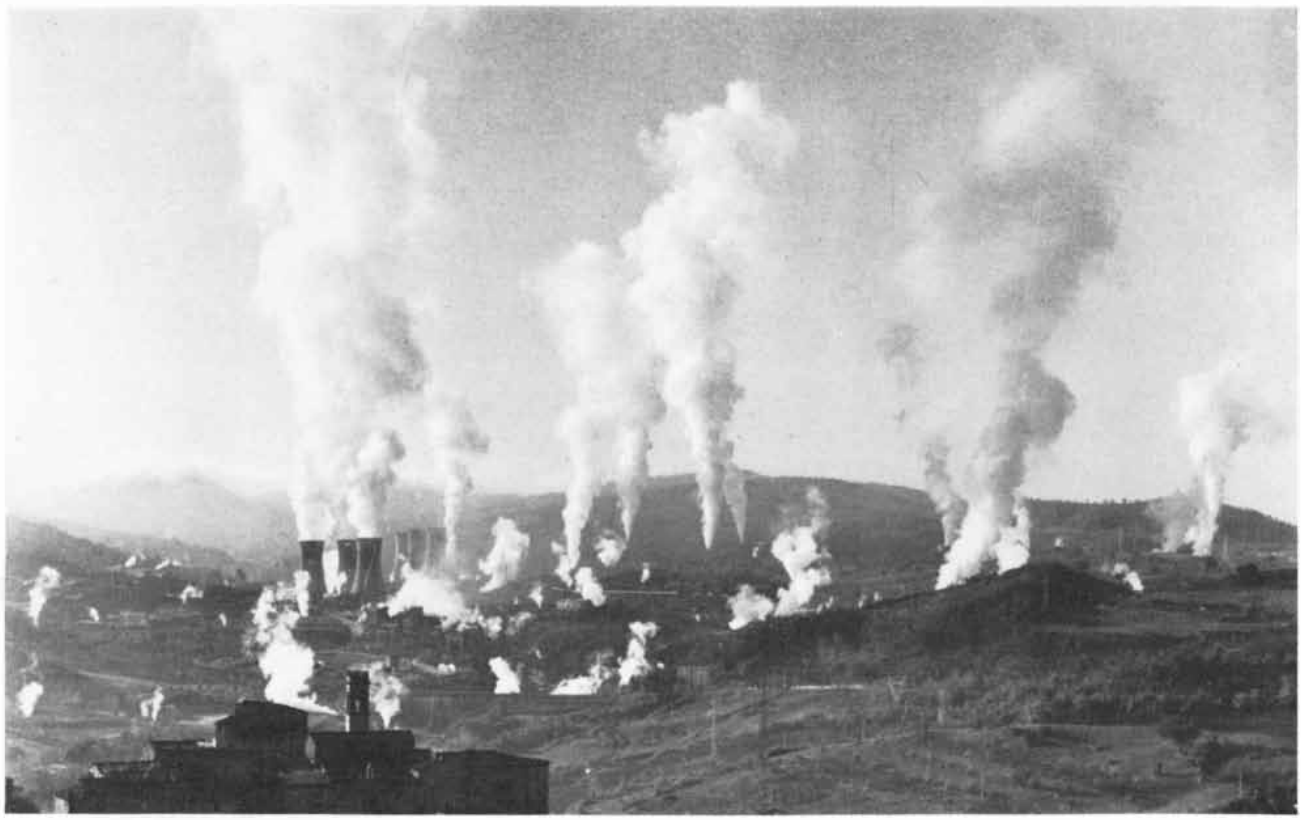
Hot springs, where water from heated strata flows naturally to the surface, have of course been known and used since ancient times. The Romans developed these watering places for medical and recreational purposes all around the Mediterranean and to the outskirts of their empire as far as Bath in the British Isles; there were also medical spas in ancient Japan and elsewhere in the Far East. Hot springs still flourish as health resorts today in Japan, in France and other centers in continental Europe, in Africa and in many other places outside the Anglo-Saxon world.

The realization that the steam in the crust might be tapped for power came at the beginning of this century. In 1904 the first electricity plant so powered was built and plugged into a steam field in northern Italy now known as the Larderello field. Over the following decades there was a slow and tentative growth of interest in geothermal energy. More plants were built on the Larderello field, and other small-scale projects for the use of natural steam or hot water for power, industrial purposes and heating were developed in Japan, Hungary, the U.S.S.R., Iceland, New Zealand and elsewhere. In the U.S. the first geothermal power plant, of 12,500-kilowatt capacity, was commissioned in 1960 on the Geysers field, which is by far the largest field yet discovered in the world.

To those investigators who early recognized the potentialities of geothermal energy the development of this resource has seemed agonizingly slow. There are several reasons why things have not gone more rapidly. Judging from the surface indications (the comparative rarity of hot springs or steam holes) the geothermal energy that might be available appeared to be highly localized and minor in amount. The explorations and the discovery of fields have been limited to those that show surface signs, because little information has been available on geological indications that might signal the presence of hidden fields. In the past such fields could be found only by speculative drilling, and it seemed that the expense of drilling would be justified only for fields located at a shallow depth. The paucity of research and information on the geothermal resources in the crust, the lack of guides for exploration and the shortage of trained specialists and

© 1971 SCIENTIFIC AMERICAN, INC

STEAM WELLS tap geothermal energy for the production of power at a plant operated by the Pacific Gas and Electric Company at The Geysers, about 90 miles north of San Francisco. Since 1960 the company has brought plant capacity at the site to 192,000 kilowatts.



LARDERELLO GEOTHERMAL FIELD in Italy has been used for generating electric power since 1904. It now has a capacity of 380,000 kilowatts. The chimney-like structures at left are hyperbolic cooling towers that are associated with the power plant at the site.

© 1971 SCIENTIFIC AMERICAN, INC

GEOTHERMAL SOURCE associated with a volcanic crater in the Rift Valley in Ethiopia was explored by means of infrared photography in a project carried out by the United Nations. The explorations were made by airplane using black-and-white infrared film, which shows hotter areas as white and cooler areas as progressively darker shades of gray. Promising photographs, such as the one shown here, were processed through a densitometer, which converts the density of the photograph in terms of ground temperature and applies a predetermined color scheme to indicate the differences. In this photograph the hottest areas are orange and the coolest ones are blue. The technique provided the first measurement of the extent of this geothermal source and range of temperatures in area.

© 1971 SCIENTIFIC AMERICAN, INC

technicians in this field have in the past combined to retard progress.

Nevertheless, the enterprise is now moving forward at an accelerating pace. It has been given impetus in the U.S. by the Geothermal Steam Act, adopted by Congress in December, 1970, which establishes the development of U.S. geothermal resources as a national goal. What is now needed is a worldwide expansion of efforts in research, exploration and the training of experts for this work.

The sources of usable geothermal energy in the earth fall into three classes: dry steam fields, wet steam fields and fields of lesser heat content consisting of water at temperatures below the boiling point (at atmospheric pressure). Each type of geothermal energy has its special uses and also capabilities for a variety of applications.

The dry steam fields are filled mainly with steam itself, under pressure and at relatively high temperatures. This steam is usable directly for the production of electric power. It can be piped right to the turbine and therefore simplifies the requirements for plant equipment; the investment in plant may be as low as $100 per kilowatt. In order to minimize piping costs the plant must be located close to the steam wells; moreover, since the steam emerges from the field at low pressure and large amounts of steam must be handled, the effective size of the turbines is limited. This means that the plant cannot be very large. The upper limit at present is about 55 megawatts. Power generators of this magnitude, each fed by 10 to 15 steam wells, are now being installed at the Geysers field in California.

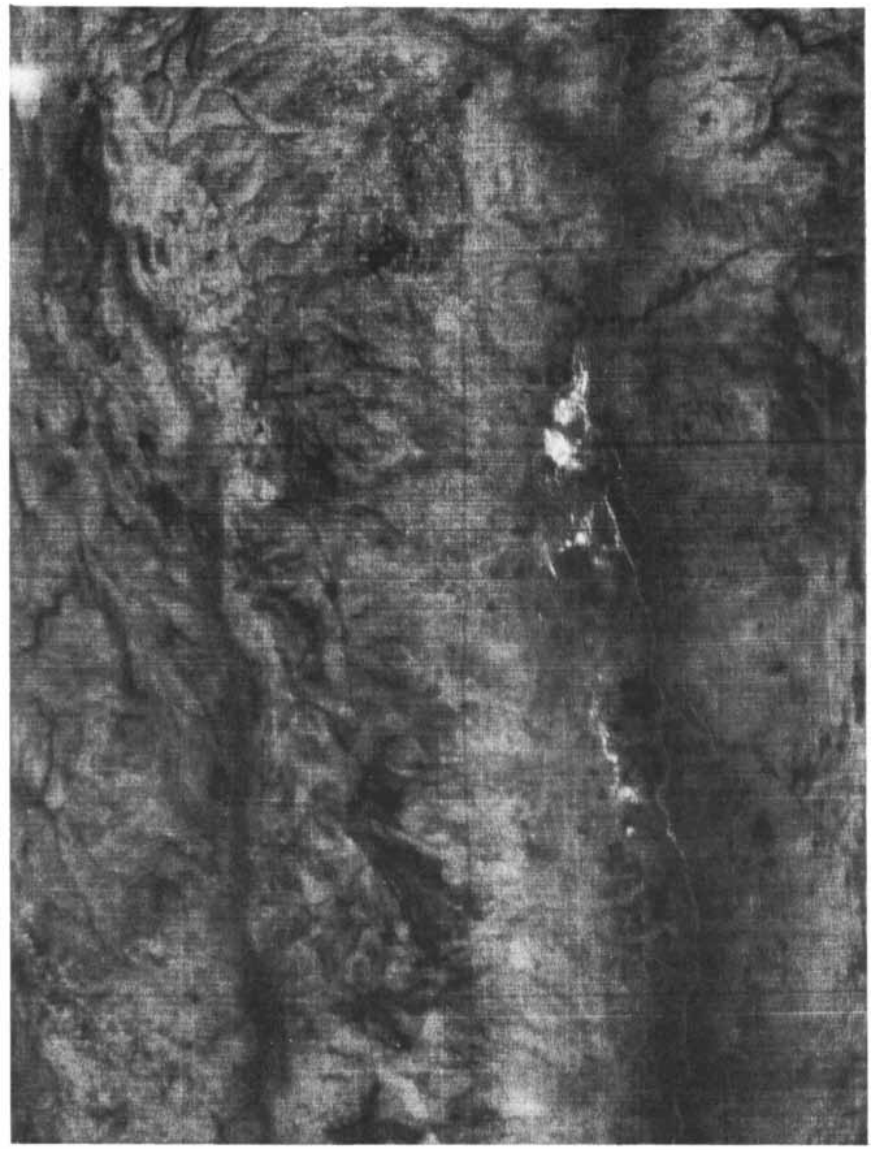
The steam from a dry field can be put to uses other than power production. The water condensed from the steam after it has given up its energy can provide a supply of fresh water. In locations near the ocean or a saltwater lake the steam could be employed as a heating medium in distillation plants for producing potable water by subatmospheric boiling, the steam in this case being provided without any cost for fuel.

So far the existence of five important dry steam fields has been established: the Larderello field in Italy, the Geysers field in California, the Valle Caldera field in New Mexico and two fields in Japan. In the absence of systematic exploration it is not yet possible to estimate how many other such fields may lie hidden in the earth's crust.

On the basis of discoveries made to date, it seems that wet steam fields may be 20 times more abundant than dry steam fields. The wet field is filled with hot water (above its boiling point at atmospheric pressure) that does not become steam until the pressure is released by drilling into the field. The superheated water in the wet field, typically at temperatures ranging from 180 to 370 degrees Celsius (about 350 to 700 degrees Fahrenheit), flashes into a mixture of steam and water as it comes to the surface. About 10 to 20 percent of the discharge, by weight, is steam; the rest is hot water. The steam can be used for power production; the hot water has a multitude of potential uses.

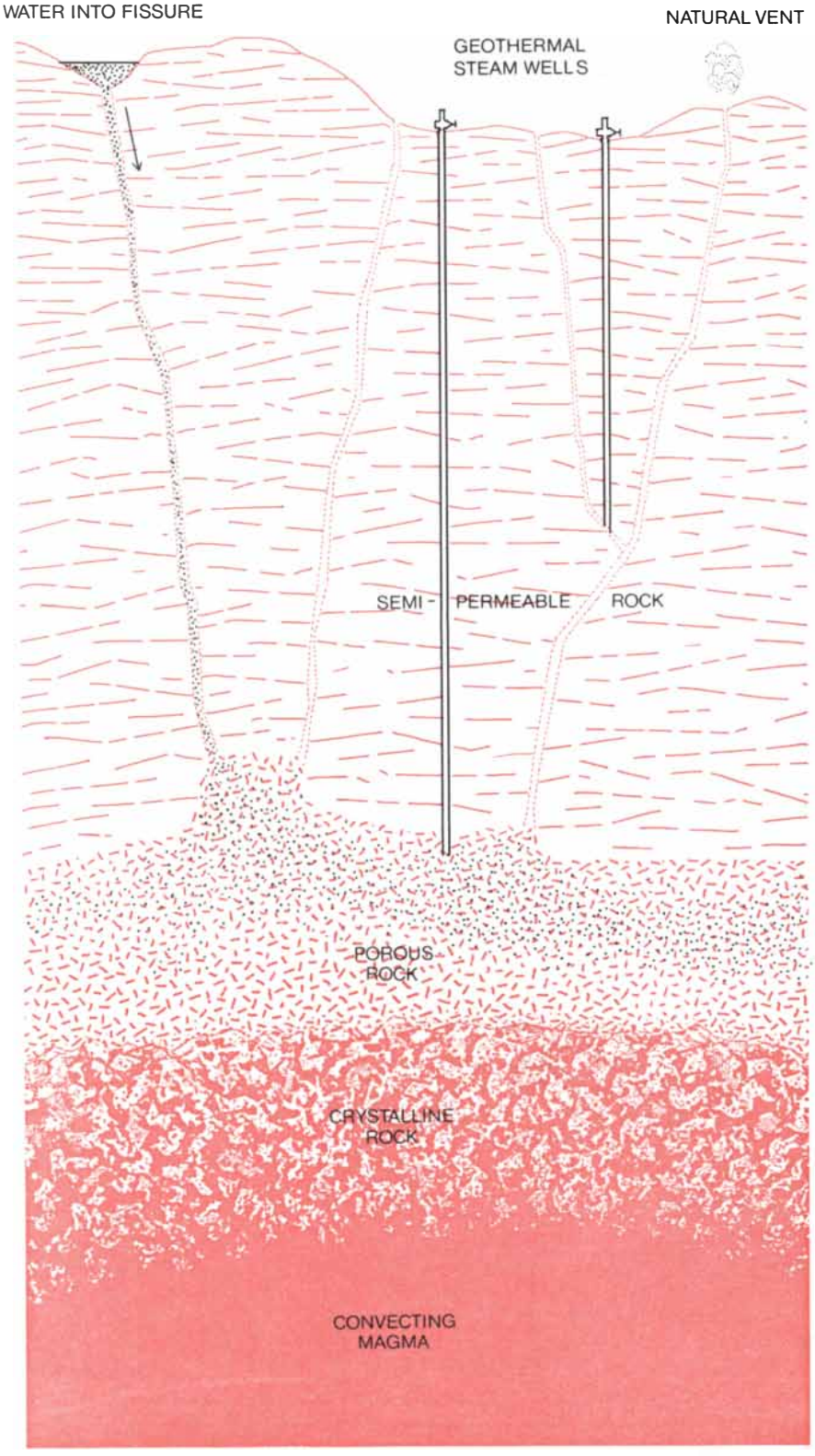
The pioneering stages of the harnessing of geothermal energy have been marked by concentration on a single application of the yield from the wells. At the Wairakei wet steam field in New Zealand, for example, the steam fraction is fed to a power-generating plant and the hot water is discarded into a river. In this respect geothermal energy has paralleled the history of the discovery of petroleum, which at first was used only for kerosene lamps. Now geothermal development is entering a more sophisticated stage through the analysis of its components and their combination into multipurpose projects.

There are already installations in which the steam of a wet steam field is devoted to the production of power and some of the hot geothermal water is distilled, without the addition of any more heat, to make fresh water. (Distillation is possible because the pressure in the flash-distillation plant is kept below at-



INFRARED VIEW of the steam field at The Geysers was obtained by the U.S. Geological Survey. The aerial photograph was made before dawn in order to minimize the effect of the sun on the temperature of the ground. Light areas at right center are geothermal areas.

© 1971 SCIENTIFIC AMERICAN, INC



GEOLOGICAL SETTING of a geothermal energy source is portrayed. The heat comes from magma, or molten rock, that has been pushed up into the earth's crust. By convection of the magma the heat moves through crystalline rock to a layer of porous rock containing water that has percolated down from the ground, sometimes to great depths. Over the porous rock is relatively impermeable rock that serves as a cap to contain the heat. Being deep in the ground, the water is under high pressure and is therefore liquid, although its temperature may be some 500 degrees Fahrenheit. It expands and rises in a natural vent; as the pressure drops, water begins to boil and produce steam. A well can tap the vent or the porous layer.

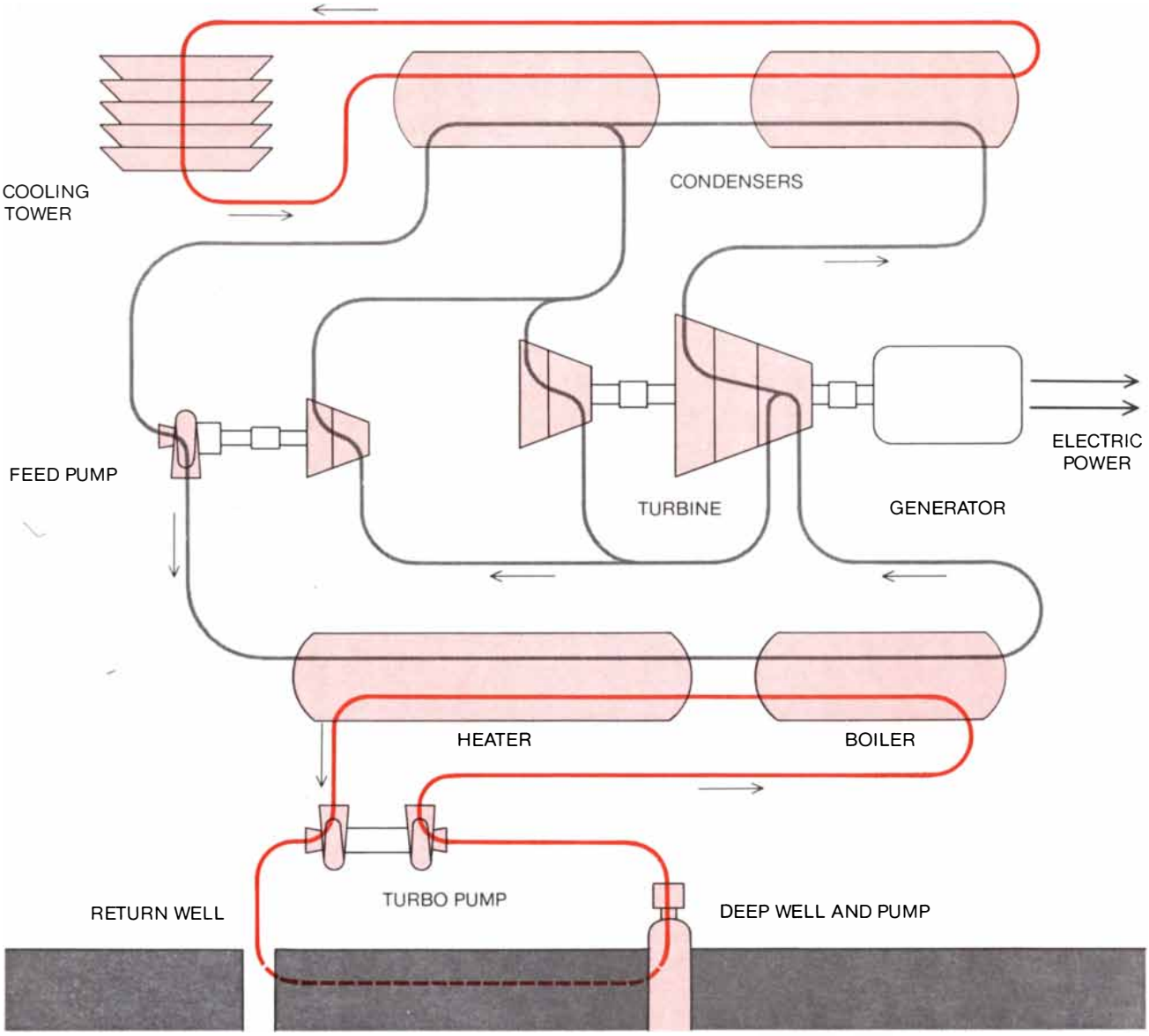
mospheric pressure by exhaust pumps.) A further step is planned at a wet steam field recently discovered at El Tatio in Chile. There the Chilean government in cooperation with the United Nations is investigating the development of a facility that will generate three products [*see illustration on page 77*]. The steam will first be used to produce electricity. Hot water produced from the steam will go through a desalination plant, producing fresh water, and the effluent from the hot-water feed will be concentrated in a mineral-rich brine from which valuable minerals will be extracted in evaporation ponds.

Following the accidental discovery of mineral-rich geothermal brines in southern California and in the Red Sea the UN began a systematic search for mineral brines as part of its program of geothermal investigations. Two discoveries of potential economic importance have been made so far, one in Ethiopia and one in Chile.

At Kawerau in New Zealand a paper and pulp company is using the hot water from a wet steam field for heating in industrial processes. In Iceland the hot water from such fields has long been applied to industrial uses and household and district heating. In Japan the applications include uses in experimental fish-farming projects, cleaning, cooking, soil-heating and bathing. Househeating with hot-well water is being developed on a large scale in several countries, notably Japan, the U.S.S.R. and Hungary (where the cost of such heating is reported to be only a fourth of that with fuel-burning systems). In the U.S. househeating from hot wells is being applied on a small scale in Boise, Idaho, and Klamath Falls, Ore.

The use of geothermal water in air conditioning is based on a process that employs water as the refrigerant and a solution of lithium bromide as a low-temperature absorbent fluid. As in other refrigerating systems the refrigerant is vaporized, thereby extracting heat from the surroundings. Then, however, the refrigerant is taken up by the absorbent. External heat (in this case supplied by geothermal water) drives the refrigerant off the absorbent as a gas; the gas is condensed to liquid, which returns to the evaporator to begin the cycle again. Two Russian investigators of applications of geothermal energy, A. N. Tikhonov and I. M. Dvorov, recently reported that a machine of this kind, used in a system providing refrigeration in summer and heat in winter, is being mass-produced

© 1971 SCIENTIFIC AMERICAN, INC



POWER PLANT using geothermal hot water instead of steam has been designed with a low-boiling-point heat absorbent such as Freon or isobutane as the driving fluid. Geothermal water is pumped through a heat-exchange system (*bottom*), where the absorbent takes up the heat. The absorbent evaporates and drives the system of turbines connected to the generator. Absorbent next goes to the condensers, where it is condensed into liquid again by water from the cooling tower and returned to heat exchanger for a new cycle.

in the U.S.S.R. Such a system has also been installed in a hotel in New Zealand, which reports that the energy cost, using geothermal water, is only a tenth of that for a system using electrically operated compressors.

The third type of geothermal field, called a low-temperature field, has only recently begun to receive attention. Fields in this class generally consist of large bodies of water in the range of 50 to 82 degrees C. (about 120 to 180 degrees F.). They are found in sedimentary deposits, notably in Hungary, where the field was discovered accidentally while drilling for petroleum was in progress. The hot water from this type of field is most efficiently used for heating: in houses, greenhouses, mines in cold climates and industrial plants. The use of such water from low-temperature fields in the U.S.S.R. is reported to have represented a saving of about 15 million tons of fuel in 1970.

The new Geothermal Steam Act of the U.S. stresses the multipurpose approach in the development of geothermal energy resources. To this end it will be necessary to plan on a comprehensive scale, treating the problem with an approach like that for the development of an entire river basin. This means that we shall need planners who are acquainted with all the technologies and economic considerations involved, from exploration to the numerous possible applications.

Much study has already been given to the costs of exploitation of geothermal energy for various purposes. Since a number of special factors are involved in this new technology, standards for estimating costs have not yet been developed; however, the UN, in response to a proposal made at the Symposium on the Development and Utilization of Geothermal Resources, which was held in Pisa in 1970, is expected to appoint a committee of experts to formulate uniform costing procedures, so

© 1971 SCIENTIFIC AMERICAN, INC

that costs in various situations and various countries can be compared.

Some of the costs are already well known from experience. Drilling a steam well costs from $50 to $150 per meter, depending on conditions, so that the drilling cost of a field 1,000 meters deep will be between $50,000 and $150,000 for one well. There are also ready answers on the costs of piping, valves and the various items of equipment for a power plant. The cost of operation for delivery of the heat from a steam field to the plant is likewise well established; with proper management this cost is only about one to three cents per million British thermal units.

What, then, are the special costs? The most important ones are related to the question of the life expectancy of the available heat supply in a field or a given well. This obviously is difficult to estimate. There are reasons to believe, however, that with proper management a geothermal field will last for many years, particularly if it is recharged by ground water or by artificial injection of gas or geothermal effluent water. At the present stage of development I believe the lifetime of a typical field can prudently be assumed to be about 30 years for purposes of estimating the amortization of the investment used in developing it. To the initial investment we must add a special cost having to do with maintenance: the wells have to be cleaned regularly and sometimes even redrilled because of the precipitation of chemicals from the steam or hot water.

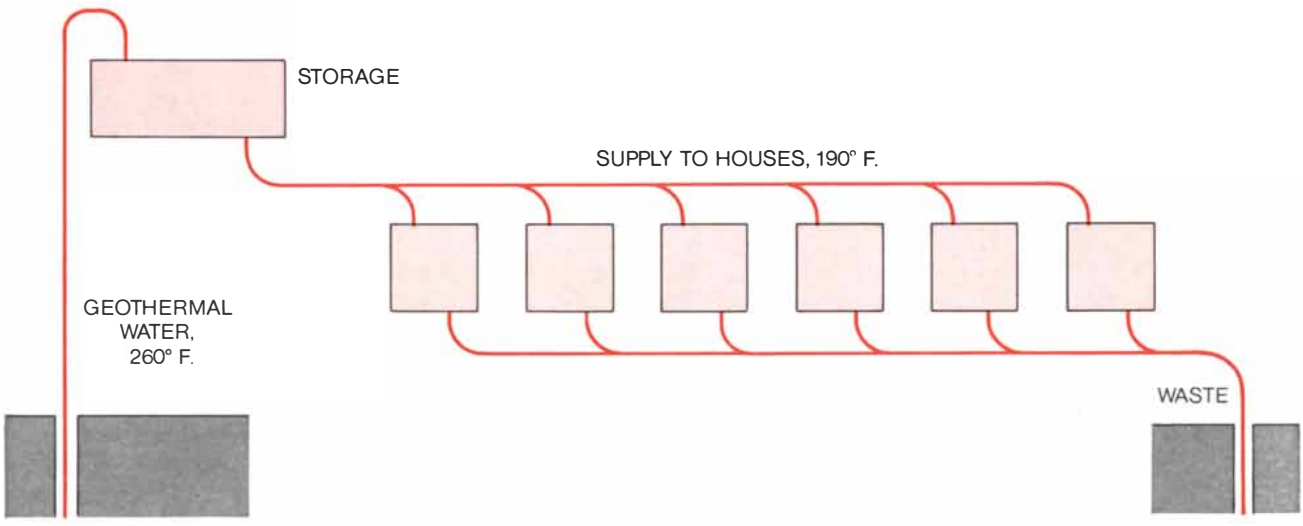
The experience thus far gained furnishes us with approximate cost figures for the various applications of geothermal energy. In a single-purpose installation producing only electric power at base load the cost of the power produced is between three and six mills per kilowatt-hour, including full amortization of all the investments over a reasonable period. In desalination plants the cost would probably be in the range of 20 to 50 cents per 1,000 gallons of freshwater yield—far below the costs of other desalination systems. For househeating, air conditioning and similar purposes the use of geothermal energy makes possible savings of up to 90 percent or more, as we have already noted. A hotel in the city of Rotorua in New Zealand reports that the operating cost of a heating and air-conditioning system based on the use of geothermal energy in a lithium bromide absorption installation is only 12 cents per million kilocalories, as against $2.40 per million kilocalories for an oil-burning system involving approximately the same investment in equipment.

These estimates are calculated from the experience of single-purpose facilities. With the development of multipurpose plants the dividends made possible by extraction of all the benefits in the crude outflow from the geothermal field (like the extraction of the various products from crude petroleum) should reduce the cost of the individual applications.

Not the least of the attractions of geothermal energy is that it can be used at little or no cost to the environment. Unlike fossil or fissionable fuels, it does not pollute the biosphere with combustion products or radiation; unlike hydropower systems, it does not flood fertile lands or generate stresses that may lead to earthquakes. It does present two hazards. The steam and hot water from many fields contain small amounts of boron and other chemicals, which can be harmful when discharged into streams. Trials at the Geysers field and at a geothermal field in El Salvador have shown, however, that the contaminated effluent can be injected back into the field without reducing production from the wells. There is reason to believe this problem will not be difficult to control. The other hazard is that the land may subside where large amounts of water are withdrawn from geothermal reservoirs. Some subsidence has occurred at the Wairakei field, which has been depleted of 70 million tons of water per year and as a result has changed in part from a wet to a dry steam field. This problem too can be controlled, by limiting withdrawals from the field to a safe rate and by recharging it with water, as is now done to prevent subsidence in petroleum fields.

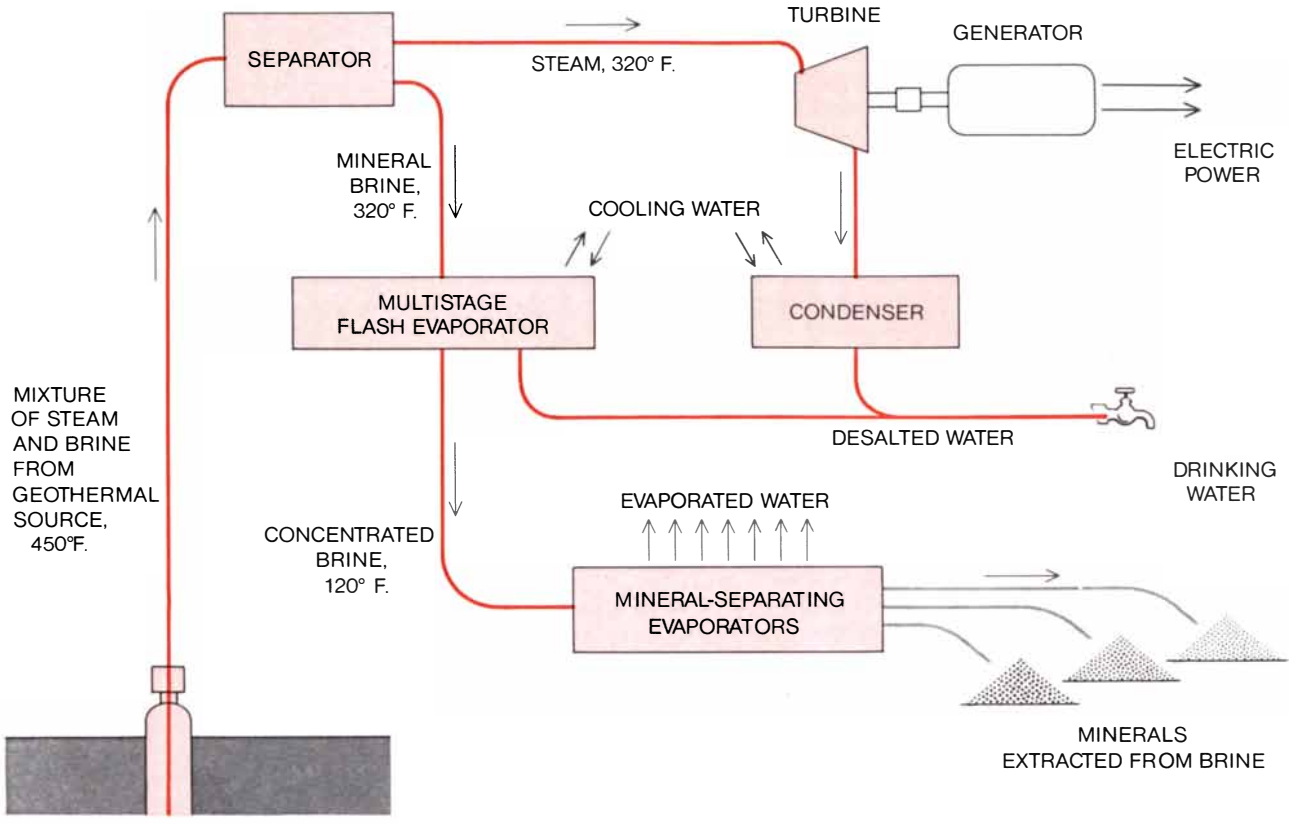
The UN is taking an active interest in geothermal energy. In cooperation with the government of Italy it conducted the symposium on geothermal energy in Pisa, where it was demonstrated that geothermal applications are marked by considerable international interest and collaboration. Although the scale of this research is severely limited in funding, a variety of imaginative ideas are being explored.

One project is concerned with the possibility of producing electricity in low-temperature fields. The heat from the geothermal water is used in a heat exchanger to boil a secondary fluid with a



HEATING OF HOUSES and other buildings is done in a few places by a scheme such as the one shown here. Geothermal water is pumped to a storage tank, from which it flows to the buildings. Such systems are in use or being developed in several countries.

© 1971 SCIENTIFIC AMERICAN, INC



MULTIPURPOSE DEVELOPMENT based on geothermal energy is being designed by the UN and the government of Chile for a geothermal field recently discovered in Chile. In this case the geothermal source produces a mixture of steam and mineral-rich brine. The steam and brine are separated, and the steam drives a turbine to produce electric power while the brine is put through an evaporator that concentrates it, thereby producing desalted water. The concentrated brine goes to a separator that extracts the minerals.

low boiling point, which then drives the power turbine. Such a plant, installed on a field providing water at 81 degrees C., is in operation at Kamchatka in the U.S.S.R.; it uses Freon as the secondary fluid. Similar small plants have recently been built in Japan.

Among all the research needs the paramount one is the development of techniques of exploration to search the earth for geothermal reservoirs, hidden as well as visible. This will call for extensive geological, geochemical and geophysical studies and testing. (The UN recently resorted to infrared surveying in a large-scale search from the air for possible geothermal sites in Ethiopia and Kenya.) From the standpoint of geology, interest naturally focuses on areas underlain by rocks of high porosity, since these are likely to hold large quantities of water. From the standpoint of utility and benefit, one hopes to find geothermal reservoirs in arid areas where underground water itself, as well as energy and minerals, would be a boon to the region.

From surface indications alone it appears there are belts of geothermal reservoirs along the western side of the Americas from Alaska all the way down to Chile, in the Middle East (Turkey) and East Africa throughout the African Rift Valley and in the Far East along the "Circle of Fire" of volcanic activity that surrounds the Pacific Ocean. In Turkey two-thirds of the country is believed to have geothermal potential, and there are good prospects for this resource in almost all the countries around the Mediterranean. The many spas of hot waters throughout Europe suggest that geothermal reservoirs should be widespread on that continent. Recent discoveries by drilling in Europe and elsewhere also indicate that a potential exists in many regions that had not previously been considered for exploration. The U.S. may have similar possibilities: drillers came on geothermal reservoirs in Louisiana and Texas recently during deep drilling for petroleum.

Inexpensive power and heating would be very helpful to many developing countries. Some of them, notably in Central America, are rich in geothermal resources—indeed, Central America has much more of this potential energy than it could use itself. Large-scale exploration and development of its abundant geothermal fields would be very worthwhile, however, particularly for the region's economy, if the power potential were fully developed and marketed in the U.S. by way of long-distance transmission lines.

As new information becomes available the magnitude of geothermal energy resources is beginning to be appreciated. On the basis of a reconnaissance, which included airborne infrared scanning over a large area, carried out by the government of Ethiopia and the UN it has been estimated that a part of the Afar region in Ethiopia may have an exploitable geothermal potential sufficient to meet the present need for electric power for the whole of Africa. There are in addition other areas in Ethiopia that are believed to have a geothermal potential of similar magnitude.

At this stage it is impossible to estimate the magnitude of the exploitable resources of geothermal energy that lie hidden under our feet in the earth's crust. The world's energy needs and exciting recent discoveries, however, certainly warrant a great effort of exploration for this ready-made store of energy.

© 1971 SCIENTIFIC AMERICAN, INC

The Spectrum of the Airglow

The airglow is the faint light emitted when atoms and molecules are dissociated high in the atmosphere. Its spectra yield clues to the physical conditions and chemical processes in the region

by M. F. Ingham

The dark sky that is the background to the stars is never absolutely black, even if one looks at it on a clear, moonless night from a place that is far away from the lights of towns. Instead it is faintly luminous and gray. This diffuse light comes mainly from four sources: integrated starlight, which is the combined effect of the many stars that are too faint to be seen individually; the zodiacal light, which is sunlight scattered by fine particles of dust in interplanetary space; the flickering aurora, which is often visible at high latitudes, and the airglow, which is the subject of this article.

The airglow results from a process of luminescence in the upper layers of the atmosphere, where atoms and molecules emit light both by day and at night. The process begins with the energy of the sun's ultraviolet light, which ionizes the atoms (removes one or more electrons from each affected atom) and dissociates the molecules. Later the atoms and molecules recombine in various ways, often emitting energy in the light that is the airglow. The aurora, on the other hand, is excited by energetic particles entering the atmosphere from outside along the lines of force of the earth's magnetic field. (This distinction between the airglow and the aurora is not altogether precise, but it will serve.)

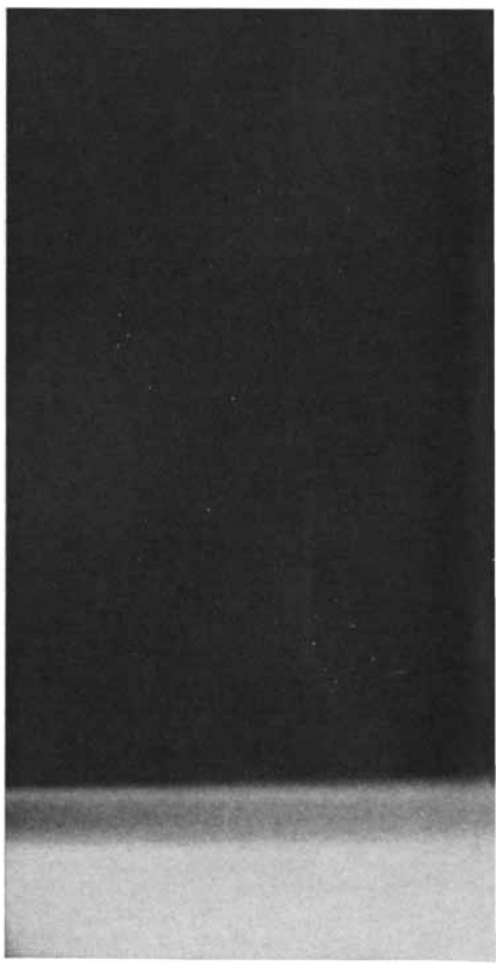
As a phenomenon for investigation the airglow lies on the border between astronomy and geophysics. To astronomers, however, it is largely a nuisance, being one of the causes of the sky brightness that sets a limit to the faintness of the stars that can be observed. To the geophysicist interested in the upper atmosphere the airglow affords essential clues to the physical conditions and chemical processes in the neutral (unionized) atmosphere between 90 and 300 kilometers (56 and 186 miles) above the surface of the earth and also in the ionosphere.

Depending on the direction in which one is observing, the four main components of the light of the night sky differ in relative brightness. The integrated starlight is brightest around the galactic equator, that is, the Milky Way. The zodiacal light is brightest near the ecliptic (the apparent path followed by the sun among the stars during the year and the projection on the sky of the plane of the earth's orbit around the sun). It is also bright close to the sun because the interplanetary dust is concentrated by gravitational forces toward the planes of the orbits of the planets and is densest near the sun. The brightness of the aurora depends on the observer's latitude; below 40 degrees latitude it can be ignored. The airglow increases in brightness toward the horizon for a reason I shall explain below.

Depending on the time of observation, the airglow is more precisely described as the dayglow, the twilightglow or the nightglow according to whether the emitting layer of the atmosphere is illuminated from above or from below or not at all [*see illustration on page 80*]. Since the photochemical processes that cause the airglow depend critically on the conditions of illumination, there are marked differences between the three forms of airglow. I shall mainly be discussing the nightglow, but I shall also describe the features that are peculiar to the twilightglow and the dayglow.

The nightglow can be observed with a photometer or a spectrometer. A photometer is essentially a small telescope with a colored filter that isolates a particular region of the spectrum; the human eye is replaced by a photomultiplier or some other light-sensitive electronic device. A spectrometer has a prism or a diffraction grating to spread the light into a spectrum, which is then recorded photographically or scanned photoelectrically. Either instrument can be used on the ground or sent aloft by balloon, rocket or satellite.

A photometer looks at a fairly large area of sky, typically 10 to 20 square



EDGE-ON VIEW OF AIRGLOW appears in a photograph made by U.S. astronauts from a Gemini spacecraft. The bright band

© 1971 SCIENTIFIC AMERICAN, INC

degrees. The wavelength of the light admitted through the filter can range from a few tens of angstroms to several hundred. As a result enough light is received for a reading to be taken in a matter of seconds or for a continuous record of the surface brightness of the sky to be kept throughout the night.

In using a photometer from the ground one must subtract from the observations the contribution of the zodiacal light and starlight to obtain a measure of the brightness of the airglow. It is also more meaningful to calculate the brightness as it would be seen from above the lower layers of the atmosphere. Here a further correction is necessary because some of the light is scattered out of the line of sight and never reaches the photometer.

None of the steps in correcting the observations is straightforward. The problem can be diminished but not eliminated by avoiding the Milky Way and the ecliptic, thus reducing the effects of starlight and the zodiacal light. The real problem, however, is that the three components are so mixed that knowledge of one depends on knowledge of the others, and only a process of successive approximation suffices to disentangle them. Finally, the correction for scattering is even less certain than the separation of the background, although the problem can be avoided altogether by observing from above the scattering atmosphere by means of an airborne photometer.

Working with a spectrometer to get a photographic spectrum of the nightglow takes a long time, entailing exposures of up to 100 hours spread over several nights, but a spectrum thus obtained contains a great deal of information. A close look at the uppermost spectrum on page 82, showing the nightglow between 3,700 and 6,600 angstroms, will indicate what can be learned in this way. (The range of wavelengths covered takes in most of the wavelengths that are visible to the human eye.)

The most prominent features in this portion of the spectrum are on the right-hand side, that is, in the red, yellow and green region. One sees prominently the red and green lines of atomic oxygen, which appear also in the aurora and give it its characteristic colors, and the yellow sodium doublet (two closely spaced lines). The sodium doublet is the same one that is emitted by sodium-vapor streetlights, but here it originates naturally in the upper atmosphere.

The remaining lines in this region belong, with one exception, to the Meinel band system (named for Aden B. Meinel of the University of Arizona, who has studied it) of the hydroxyl radical (OH). The system extends far into the infrared region; if more of the nightglow spectrum were shown, the Meinel system would be seen to become quite intense. The exception is the first Balmer line of hydrogen (H alpha) at 6,563 angstroms. (The Balmer lines are named for Johann Jakob Balmer, a 19th-century German mathematician and physicist who studied hydrogen emissions intensively.)

The blue and violet region of the spectrum is much more complicated. To un-

at bottom is the earth; the narrow, darker band above it is the lower atmosphere, and the thin, bright strip above that is the airglow. The pale cone of light between the airglow and the large, bright spot, which is an overexposed image of Venus, is the zodiacal light. The airglow results from atomic and molecular luminescence; the zodiacal light is sunlight scattered by fine particles of dust.

© 1971 SCIENTIFIC AMERICAN, INC

ravel it one must keep two things in mind. First, we are dealing here with both emission and absorption lines; the bright lines, representing light emitted by atoms and molecules and reaching the spectrometer, are emission lines, and the dark lines, representing light that has been absorbed on the way to the spectrometer and is therefore missing, are absorption lines. Second, there are actually three spectra on top of one another: the emission spectrum of the nightglow, with narrow, bright lines; the absorption spectrum of the zodiacal light, which is almost identical with the spectrum of sunlight, and the absorption spectrum of the integrated starlight. The spectrum of the aurora does not appear, since the spectrogram was obtained in Bolivia at 16 degrees south latitude.

Some of the dark bands in the spectrum are gaps between emissions. Others are absorption lines, such as two lines (designated *H* and *K*) of the calcium ion (Ca^+), which also form the most prominent features of the solar spectrum; the sharp line of neutral (un-ionized) calcium at 4,227 angstroms, and the broader line of neutral iron at 3,720 angstroms. Most of the emission lines belong to the Herzberg system of diatomic oxygen (O_2). (The system is named for Gerhard Herzberg of the Canadian National Research Council, who has just been awarded the 1971 Nobel prize in chemistry for his work on spectra.) The lines in the Herzberg system are grouped in bands, each beginning sharply on the violet, or short-wavelength, side and becoming weaker and more spread out toward the red. In this part of the spectrum most of the Herzberg bands overlap; two isolated ones can be seen beginning at 3,941 and 3,975 angstroms. Emission features not belonging to the Herzberg system are the diffuse bands at about 3,615, 3,700 and 3,910 angstroms, which have not been identified, and the sharp line at 4,359 angstroms, which is a mercury line from streetlights in nearby La Paz.

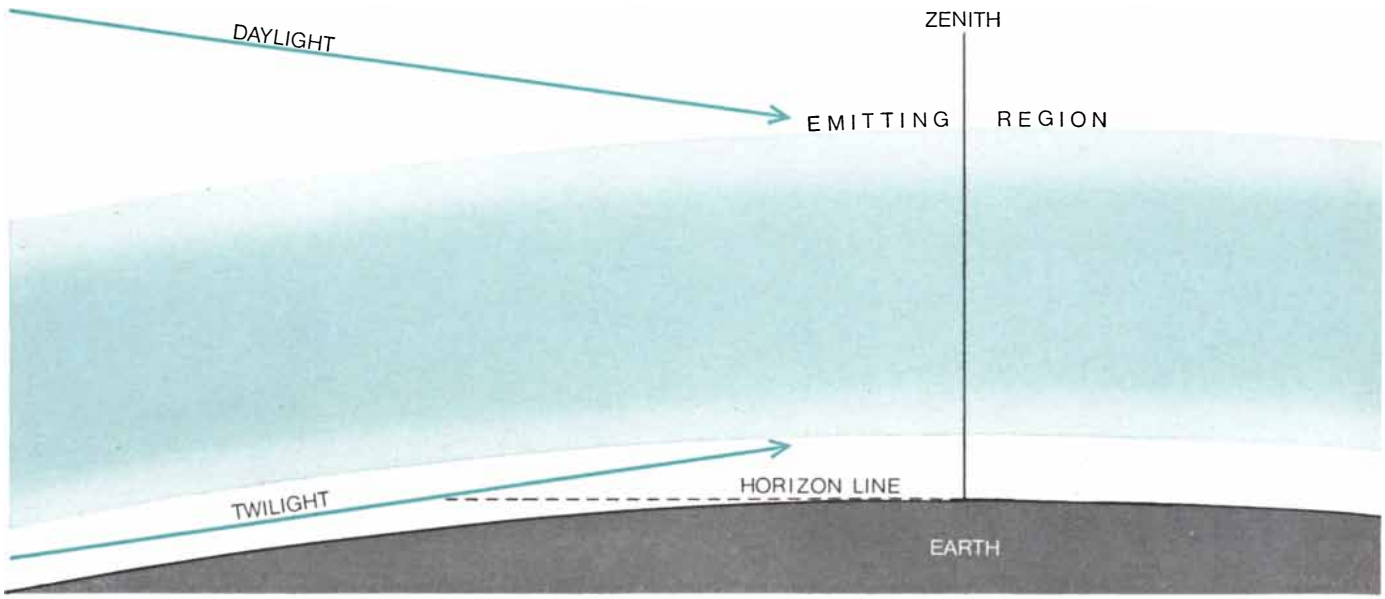
The features I have been pointing out encompass only a relatively short portion of the nightglow spectrum, which has in fact been recorded photographically from 3,100 angstroms in the near ultraviolet to 12,400 angstroms in the infrared. In the ultraviolet the Herzberg bands of O_2 predominate, and in the infrared the Meinel OH system is notably strong. This system extends almost to 50,000 angstroms. Because of absorption by water vapor, however, it can be observed only to 12,400. The Meinel band at 12,000 angstroms is 100 times as intense as the band at 6,500 angstroms.

Also in the infrared the oxygen molecule contributes two band systems to the spectrum: the atmospheric system, with bands at 7,619 and 8,645 angstroms, and the infrared atmospheric system, whose first two bands lie at 1.27 and 1.58 microns. The bands at 7,619 angstroms and 1.27 microns are absorbed by the O_2 in the lower atmosphere, below the emitting layer, and so cannot be observed from the ground.

Absorption by the atmospheric layer of ozone prevents observation from the ground at wavelengths shorter than 3,000 angstroms. Rocket flights have revealed, however, that the night sky is quite bright at the wavelength of the Lyman-alpha line of hydrogen at 1,216 angstroms (the first of a series of hydrogen lines named for Theodore Lyman of Harvard University).

The spectrum of the twilightglow differs from the nightglow spectrum in that certain features disappear shortly after the end of twilight and others are markedly stronger in twilight than they are during the night. In the first category the chief features are bands of the ionized nitrogen molecule (N_2^+); two lines of neutral lithium at 6,708 angstroms, which are highly variable in intensity; a line due to potassium at 7,699 angstroms, and a helium line at 10,830 angstroms. The nitrogen bands are degraded toward the violet, that is, they spread out and become weaker toward the left side of the spectrum [*see middle illustration on page 82*]. The two strongest nitrogen bands have heads, or sharp beginnings, at 3,914 and 4,278 angstroms.

The yellow lines of sodium, the red lines of oxygen and the infrared atmospheric system of O_2 are strongly enhanced in the twilightglow, and the green line of oxygen is somewhat en-



DIFFERENCE OF ILLUMINATION is involved in different names applied to the airglow. Depending on whether the emitting layer is illuminated by the sun from above, as indicated by the arrow labeled "daylight," from below, as indicated by the arrow labeled "twilight," or not at all, the airglow is more precisely called the dayglow, the twilightglow or the nightglow. Characteristics of

© 1971 SCIENTIFIC AMERICAN, INC

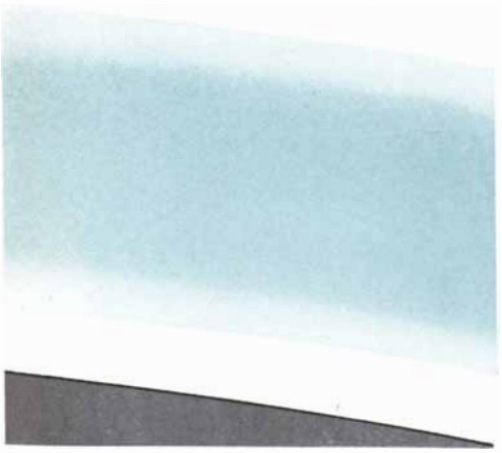
hanced. The evolution of the twilight-glow can be seen in the five twilight spectra at the bottom of the next page, which were obtained at intervals of three minutes in a darkening sky. Here the sodium doublet proceeds from absorption in the first spectrum, when the scattered sunlight in the sky was still strong enough to overcome the sodium emission, to emission in the last spectrum. The red oxygen line at 6,300 angstroms can be seen emerging, but the green line of oxygen does not appear.

Less can be said about the dayglow. Its existence has long been recognized, but observing it has only recently become possible. The problem is of course the brightness of the day sky, which completely swamps the emission lines. At sea level a pure blue sky is some 10 million times brighter at the zenith than the night sky, whereas the emission lines of the dayglow are seldom more than 1,000 times brighter than the emission lines of the nightglow. One means of overcoming this difficulty is to observe from balloons or rockets at great height; at 90 kilometers the sky is already a million times darker than it is at ground level. Another method, which makes it possible to observe the dayglow from the ground, is to isolate the emission lines from the continuous emissions by means of a spectrometer or some other device that provides high resolution.

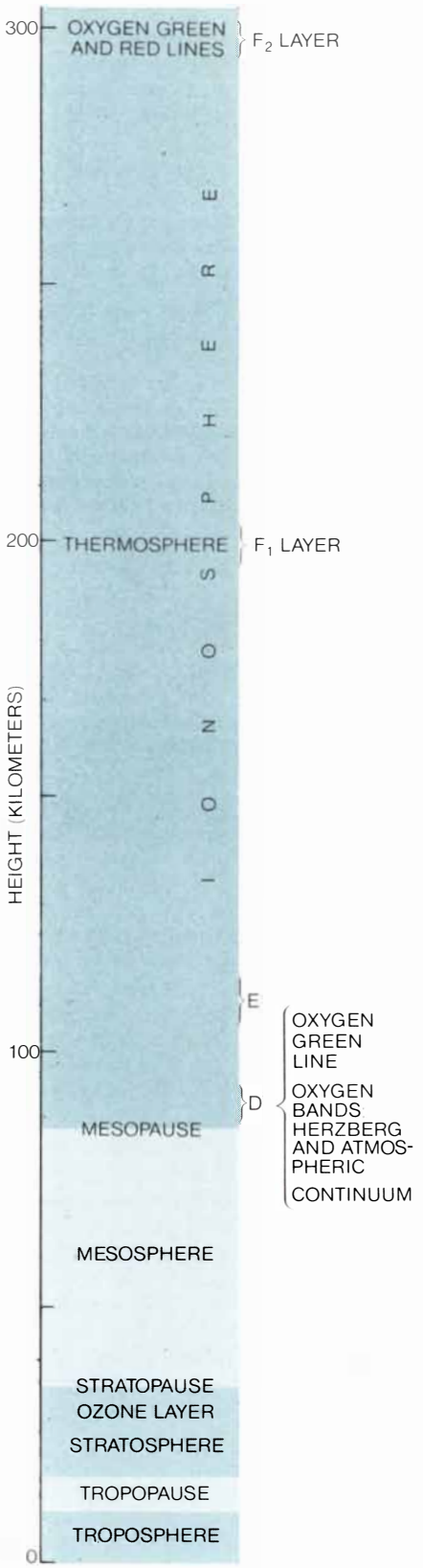
Most of the nightglow emissions are present also in the dayglow, usually with an intensity ranging from 10 to 100 times greater. An important exception is the OH emission, which decreases in intensity during morning twilight and then regains its nighttime value during the day. Also peculiar to the dayglow are a number of emissions in the ultraviolet and the infrared; they are due to atomic oxygen and molecular nitrogen. These emissions result from collisions between the atoms and molecules concerned and electrons separated from other atomic and molecular species by the sun's ultraviolet radiation.

From the spectrum of the airglow a great deal can be deduced about the composition of the upper atmosphere and about the physical conditions at various heights [*see illustration on page 84*]. Until the advent of sounding rockets the heights of the emitting layers were known only roughly. The method then used was developed by the Dutch astronomer P. J. van Rhijn; it depends on the fact that most airglow emissions are produced in relatively thin layers of the atmosphere where the physical conditions allow the particular photochemical processes concerned to proceed. Since an observer looking toward the horizon will see through a greater thickness of the emitting layer than he would if he were looking toward the zenith (which is also why the airglow is brighter near the horizon), the way the brightness of a particular emission varies with the angular distance from the horizon would serve to determine the height of the emitting layer, provided that emission was uniform throughout the layer. Unfortunately the emission is patchy rather than uniform. Moreover, as I pointed out above, the correction for scattering in the lower atmosphere is difficult to make accurately. As a result the "van Rhijn heights" obtained by different observers vary widely. For example, the height of the layer where the green oxygen line is emitted has been reported to be as low as 85 kilometers and as high as 400 kilometers. For OH the range is from 70 to 900 kilometers!

Direct measurements made with rocket-borne photometers are much more reliable. One way to make such a measurement is to have the instrument view the sky at a right angle to the axis of the vehicle and approximately in the horizontal plane. When the rocket passes through the emitting layer, the photometer records a maximum in the intensity. A typical measurement places the height of the layer where sodium is emitting at 89 kilometers and indicates that the layer is about nine kilometers thick if one limits the boundaries to the

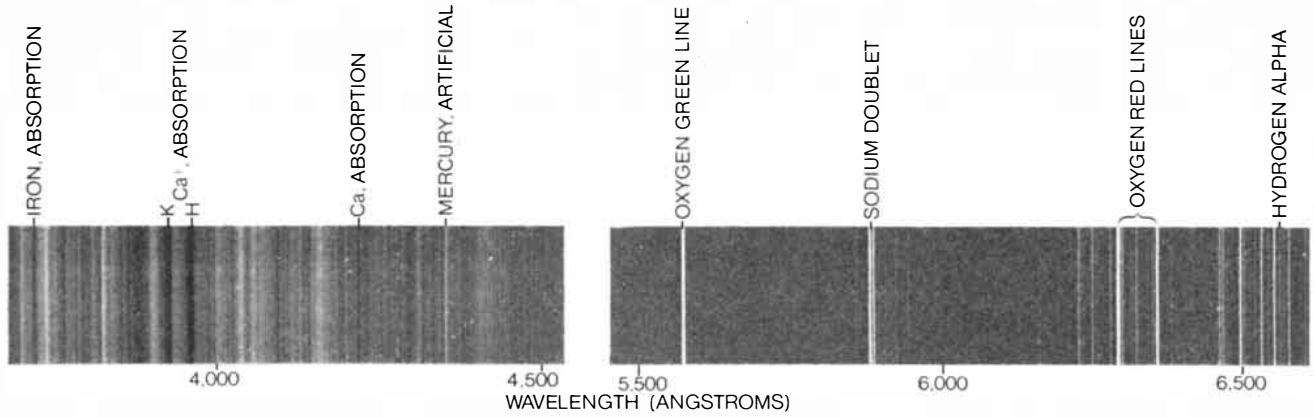


the airglow differ in each case. The emitting region begins about 90 kilometers above the earth and extends to about 300 kilometers.

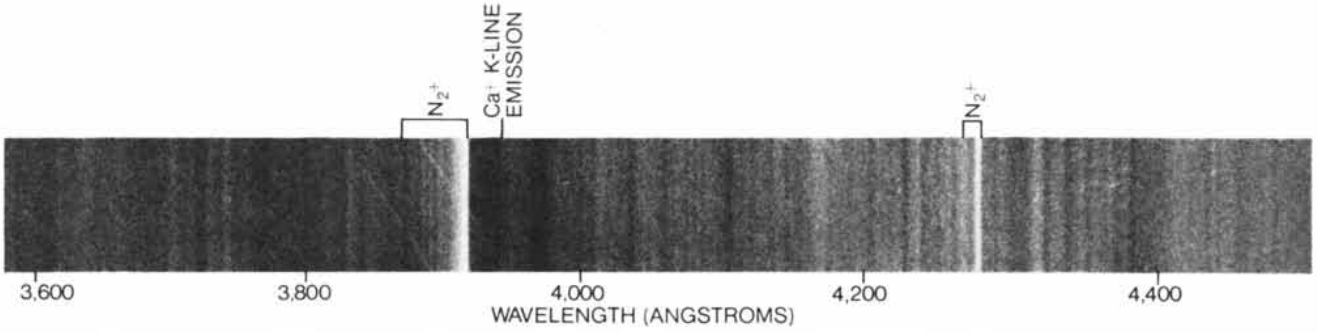


NIGHTGLOW EMISSIONS occur at heights shown on diagram of atmosphere's upper layers. Several emissions, listed at lower right, occur at about 100 kilometers; oxygen emissions producing green and red lines in a nightglow spectrum are at 300 kilometers.

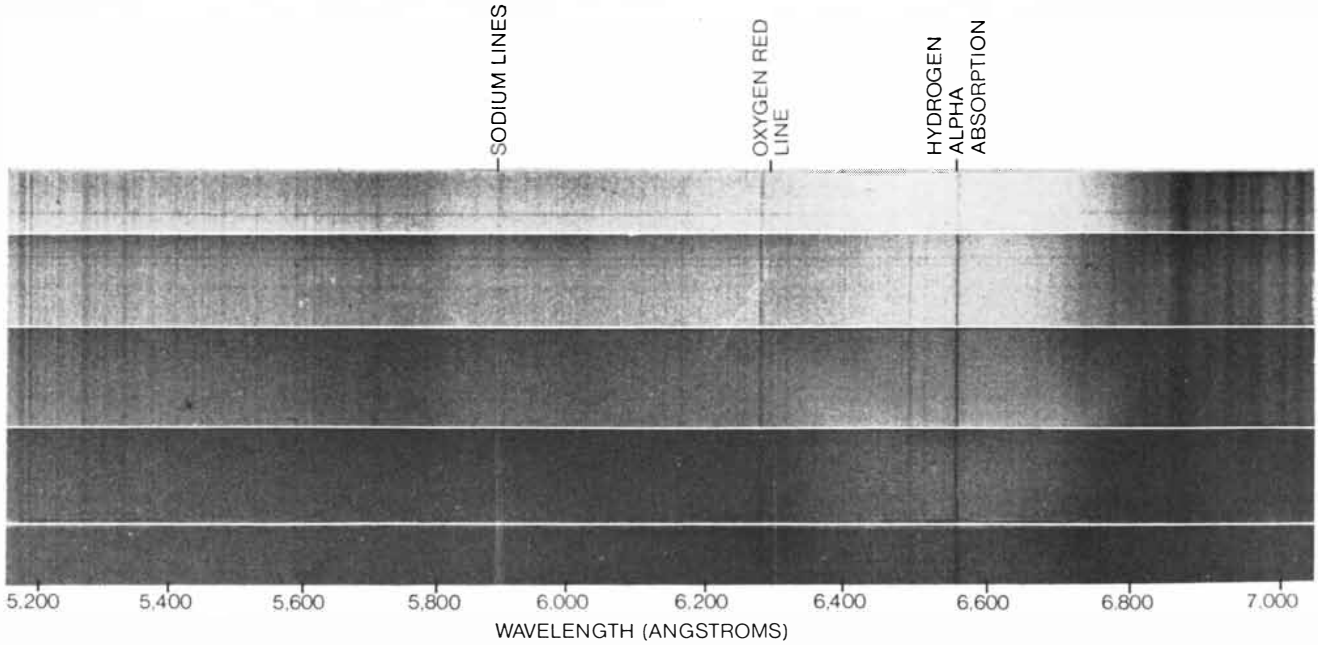
© 1971 SCIENTIFIC AMERICAN, INC



SPECTRUM OF NIGHTGLOW is shown for regions of interest between 3,700 and 6,600 angstroms, with a gap in the middle of the band where there are no nightglow features. In the green, yellow and red region at right the lines are fairly intense. The blue and violet end of the spectrum at left is more complicated. It includes a number of bands that begin sharply toward the violet, or short-wavelength, side and become weaker and more spread out toward the red side. Many belong to Herzberg system of diatomic oxygen.



FEATURES OF TWILIGHTGLOW appear in a spectrum that results primarily from the zodiacal light. They include bands of the ionized nitrogen molecule N_2^+, which begin sharply toward the red end of the spectrum and become degraded toward violet end.



FIVE SPECTRA were made in sequence at three-minute intervals as the sky was darkening at the end of a day. The earliest one is at top, the latest at bottom. The sodium doublet can be seen changing from absorption in the first spectrum, when scattered sunlight in the sky was still strong enough to overcome the emission of light by sodium, to emission in the last spectrum. The brighter of the red oxygen lines can be seen emerging similarly at 6,300 angstroms. Hydrogen alpha appears as an absorption line in these spectra.

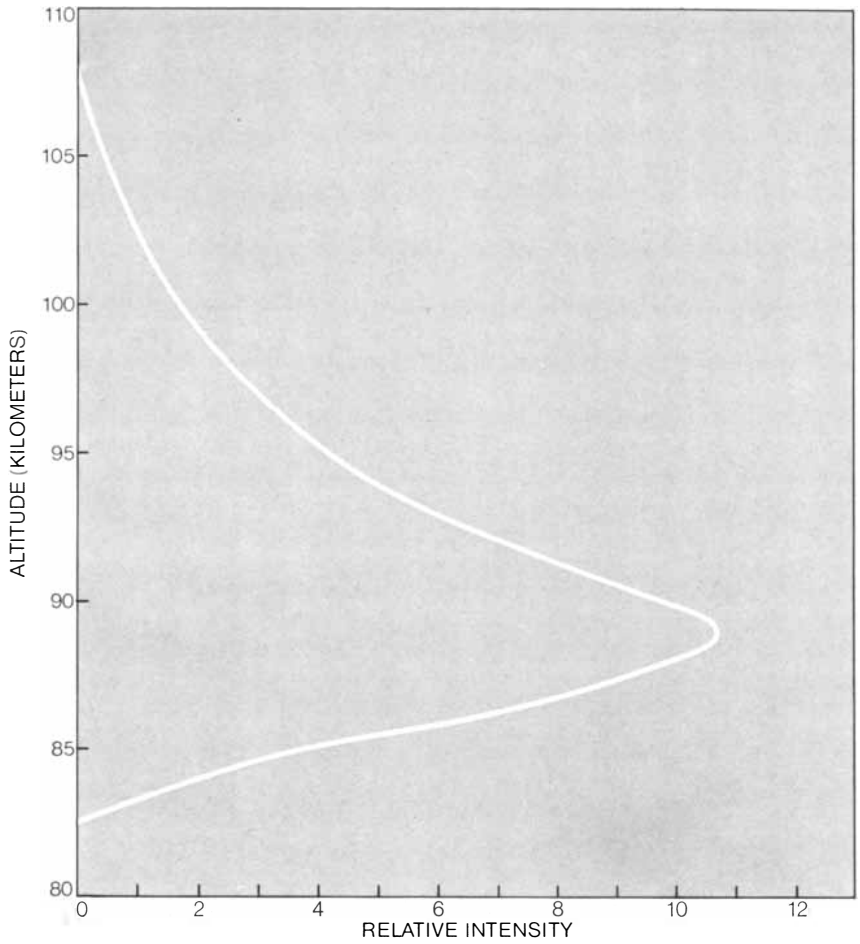
© 1971 SCIENTIFIC AMERICAN, INC

region where the relative intensity of emission is at least half of the maximum [*see illustration at right*]. By similar procedures the probable heights of all the main airglow emissions have been obtained.

The problem of how all the various airglow emissions are produced is far from being satisfactorily resolved. All that is possible at present is to give a general indication of some of the processes thought to be responsible. The difficulties in the way of more complete explanations include lack of precise knowledge of the physical conditions and chemical abundances in the upper atmosphere, uncertainty about the details of some of the reactions and the fact that in some cases more than one reaction seems possible.

The processes of excitation, whereby an atom or a molecule absorbs energy that is subsequently emitted as radiation, are broadly of five kinds. The first is impacts from energetic particles, not so much the particles from outside the atmosphere that are primarily responsible for the aurora as collisions of atoms and molecules with photoelectrons, that is, electrons that have been removed from other atoms or molecules by solar energy and have thereby absorbed energy. The second process involves ions, which are quick to gain or lose electrons because they are electrically unbalanced. Ionic excitation mechanisms operate in the ionosphere, and chiefly in the high ionospheric layer designated *F*, where ions and free electrons are plentiful. At lower levels one finds emissions resulting from the third and fourth processes: photodissociation and chemical reactions. Two related processes, resonance scattering and fluorescence, form the fifth category. Resonance scattering occurs when an atom absorbs energy and later emits radiation of the same wavelength. Fluorescence means that an atom absorbs light of one wavelength and emits light of a longer wavelength. Resonance scattering and fluorescence operate in the airglow layer during the day, at twilight and at great heights at night where the atmosphere is still lit by the sun. In addition to these five processes another process called quenching must be taken into account. An excited atom or molecule does not always lose its energy by emitting radiation; an event such as a collision with another particle can deexcite it. This is quenching.

Let us now see how these processes work for various emissions. For the green oxygen line originating at 90 kilometers the most probable excitation



INTENSITY OF SODIUM EMISSION is portrayed as it was measured by a photometer carried into the atmosphere by a rocket. The maximum emission is at about 89 kilometers. The band where emission is at least half of the maximum extends for about nine kilometers.

mechanism is a three-body association process wherein two oxygen atoms combine or associate to form an oxygen molecule and a third atom becomes excited. The process can be depicted as follows: $O + O + O \rightarrow O_2 + O^*$, where the asterisk indicates an excited atom. Oxygen atoms are particularly plentiful near the 100-kilometer level because of the dissociation of oxygen molecules by the sun's ultraviolet radiation, and a vast amount of chemical energy is stored in them.

The red oxygen lines are emitted at about 300 kilometers, in the *F* region of the ionosphere, which is also the upper limit of the green-line emission. (There is no red-line emission at 90 kilometers because of quenching.) At 300 kilometers the process thought to be responsible for both red and green emissions is atom-ion interchange, where an oxygen ion and one of the atoms forming a molecule of oxygen change places: $O^+ + O_2 \rightarrow O_2^+ + O$, where the superscript plus sign denotes a positive ion. This reaction is followed by dissociative recombination, in which the ionized molecule regains an electron and in so doing dissociates into two oxygen atoms, one or both of which may be excited: $O_2^+ + e \rightarrow O^* + O^*$, with e representing an electron.

In twilight the sodium lines result from resonance scattering. Sunlight of wavelength 5,893 angstroms is absorbed by the sodium atoms and light of the same wavelength is reemitted. In the nightglow photochemical reactions are responsible. Two suggestions have been put forward, both of the atom-exchange type: $NaH + O \rightarrow Na^* + OH$ and $NaH + H \rightarrow Na^* + H_2$.

Where the sodium comes from is another puzzle. Some investigators believe it comes from the earth as salt from ocean spray or volcanic dust carried aloft by atmospheric turbulence; others think interplanetary material or meteors are the source. At any rate, the amount involved is minute: in the upper atmosphere only one atom in 10,000 million is an atom of free sodium.

It is worth noting that sodium is not

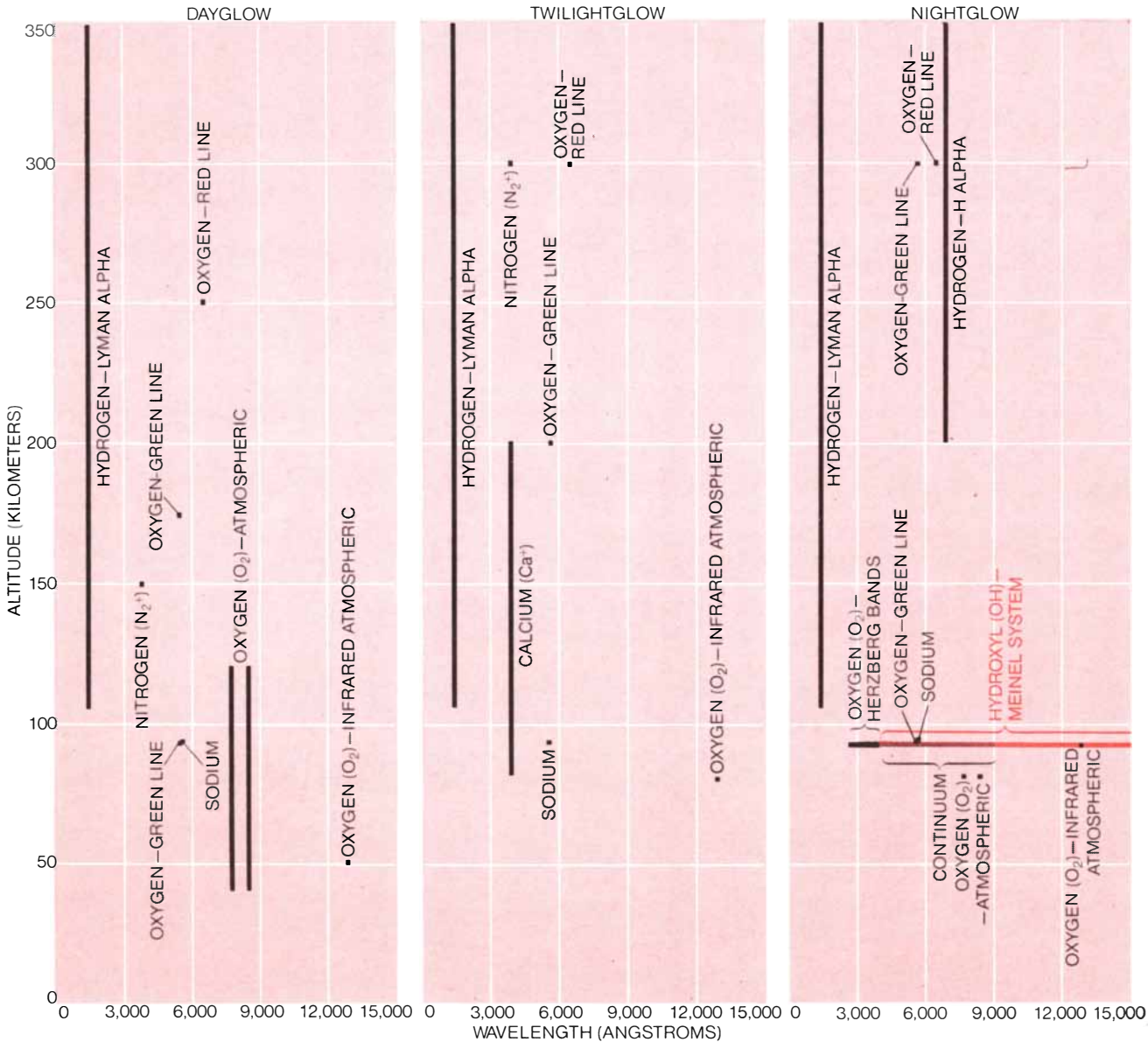
© 1971 SCIENTIFIC AMERICAN, INC

the only metal in the upper atmosphere. The *H* and *K* lines of ionized calcium and two lines due to lithium and potassium are observed in the twilightglow, and a line of ionized magnesium appears in the dayglow. The calcium and lithium lines are quite variable and have a tendency to appear at certain seasons of the year and even on specific dates. The most probable explanation is that the atoms are deposited in the atmosphere by meteor streams that the earth encounters at certain times during its yearly revolution around the sun.

I turn now to the two lines of hydrogen, Lyman alpha and H alpha, which originate much higher in the atmosphere. Hydrogen is the lightest of the constituents of the atmosphere and extends to the greatest heights, where it is always illuminated by solar ultraviolet radiation, either directly or as a result of scattering, which transports the radiation into the shadow of the earth.

The strongest line in the ultraviolet spectrum of the sun is Lyman alpha at 1,216 angstroms; Lyman beta, which at 1,026 angstroms is the second line of the Lyman series, is also strong. The Lyman-alpha and H-alpha lines seen in the spectrum of the night sky arise respectively from resonance scattering of solar Lyman alpha and from fluorescence of solar Lyman beta [*see illustration on opposite page*]. These processes involve the first and second excited levels of neutral hydrogen. When such an atom absorbs light of 1,216 angstroms, the atom is excited to the first upward level, whence it returns to the ground state by emitting a quantum of radiation of the same wavelength. In other words, resonance scattering takes place. When Lyman-beta radiation is absorbed by an atom of neutral hydrogen, the atom is excited to the second level, whence it can return to the ground state either directly with the reemission of a quantum of Lyman beta (resonance scattering) or by way of the first excited level, emitting first a quantum of H alpha and then a quantum of Lyman alpha (fluorescence).

Fluorescence is also generated in the upper atmosphere by cosmic X rays of

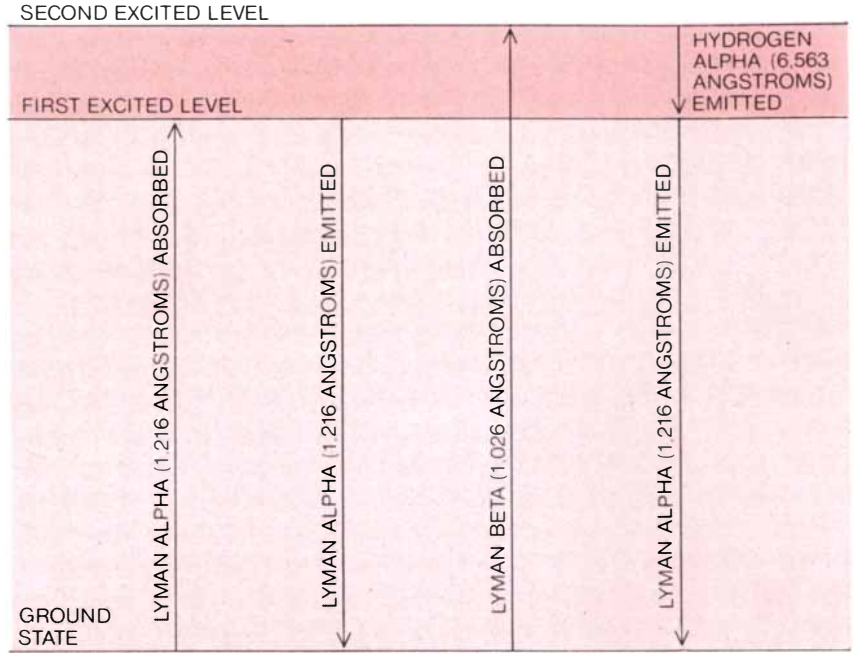


COMPONENTS OF AIRGLOW are shown by emitting atom or molecule and wavelength and probable altitude of emission. For example, the Lyman-alpha line of hydrogen appears at 1,216 angstroms in the dayglow, the twilightglow and the nightglow, and the emissions occur at altitudes from 100 kilometers upward. Some emissions, such as those of infrared atmospheric oxygen, occur over a range of wavelengths and are therefore bands. In the nightglow the Herzberg bands and the continuum band overlap somewhat.

© 1971 SCIENTIFIC AMERICAN, INC

solar or stellar origin. They liberate electrons from atmospheric atoms and molecules at an altitude of about 80 kilometers, and the electrons in turn ionize and excite molecular nitrogen to form N_2^+, which emits the band at 3,914 angstroms that is such a strong feature of the twilightglow. (Observations of this fluorescence have been used to detect the detonation of nuclear weapons in space, since the weapons also generate X rays. Similar observations might serve to record cosmic events, such as supernova explosions, that generate X rays in large quantities.)

Let us now consider the molecular bands. The Herzberg bands of O_2 and the atmospheric infrared bands probably both owe their origin to three-body association: $O + O + X \rightarrow O_2^* + X$, where X, the third atom, is necessary for the association of the two oxygen atoms and is unchanged in the process except that it may acquire some of the energy that is released. As for the Meinel bands, the excitation of OH that produces them is probably the result of two succeeding reactions: $OH + O \rightarrow H + O_2$, followed by $H + O_3 \rightarrow O_2^* + OH^*$. The first reaction supplies the atomic hydrogen. The ozone (O_3) is transported upward from the lower levels where it is produced.



HYDROGEN EMISSIONS of the airglow are depicted schematically. Lyman-alpha emission can result if an atom of hydrogen absorbs sunlight of wavelength 1,216 angstroms. The energy of the atom is thereby raised to the first excited level; the atom returns to the ground state by emitting a quantum of radiation of the same wavelength. The phenomenon is called resonance scattering. An example of fluorescence, which means that light of one wavelength is absorbed and the atom returns to the ground state by stages, emitting light of longer wavelength, is provided by Lyman beta. The atom absorbs light of 1,026 angstroms and is raised in energy to the second excited level. The atom can return to the ground state by emitting first a quantum of hydrogen-alpha radiation and then one of Lyman-alpha radiation.

In addition to the emission lines and bands, the airglow exhibits a faint continuous spectrum that appears to extend from about 4,000 angstroms into the infrared. It shows up in the gaps between emission lines as a weak contribution to the spectrum. Little is known about this emission because precise observations are difficult to make. Study of the phenomenon serves to emphasize the problem of separating the components of the light of the night sky. When a strong emission line is being observed, the background of starlight and the zodiacal light is a relatively small part of the total intensity, but the continuous spectrum compares in brightness to the other two components. The difficulty of assigning an accurate value to the continuous spectrum is twofold: the brightness of the zodiacal light is not accurately known in the parts of the sky far from the sun, where it is faint, and the brightness of the integrated starlight has been measured (by counting stars) only in a narrow region of the visible spectrum and not in the infrared.

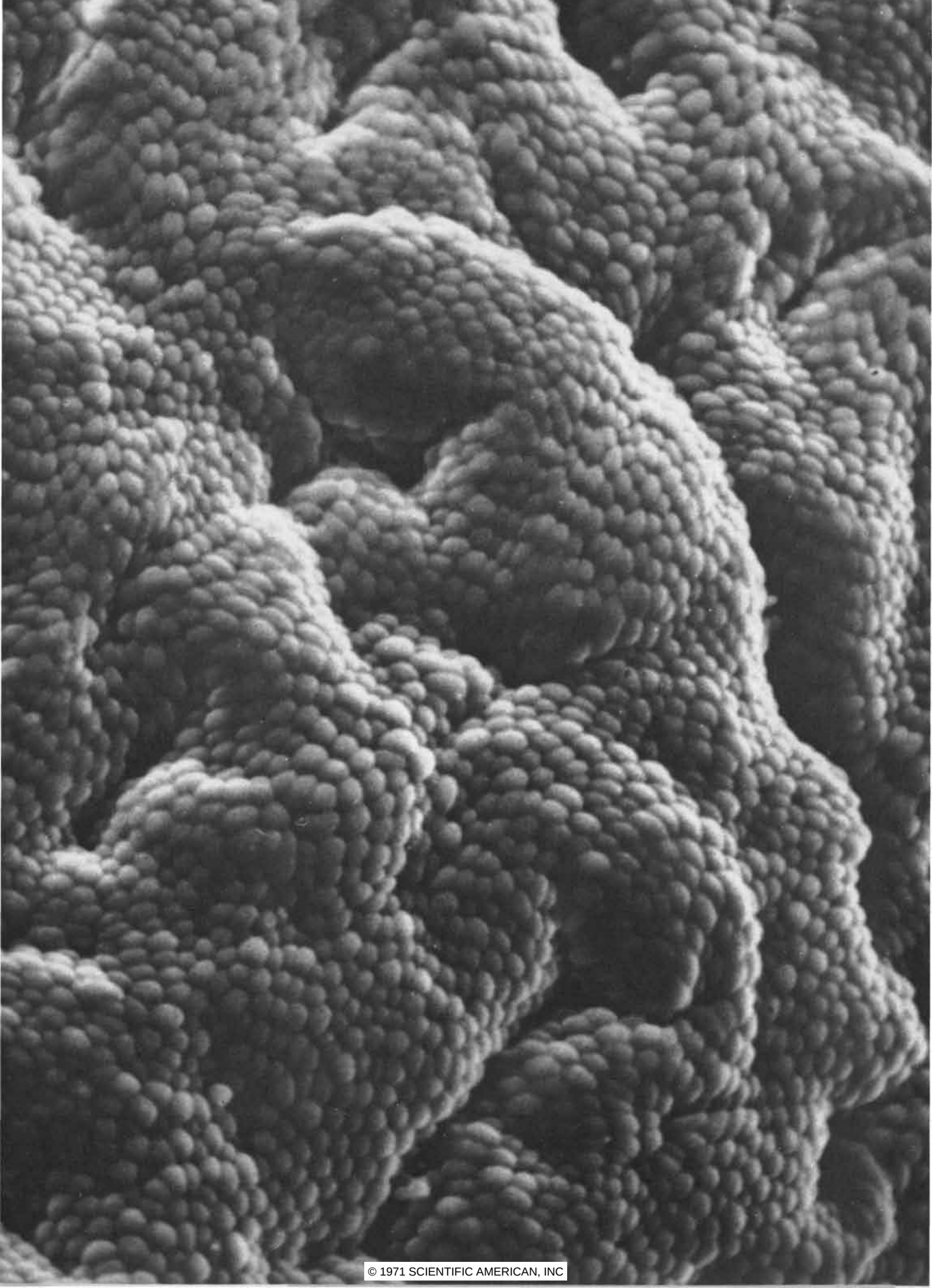
Rocket measurements put the height of the layer of continuous emission at about 90 kilometers. The favored reaction involves nitric oxide: $NO + O \rightarrow NO_2^*$. The reaction is well known in the laboratory for producing a yellow-green afterglow in air. This glow has a continuous spectrum that extends from 3,700 to 9,000 angstroms, with a maximum near 6,500 angstroms. If the airglow continuum could be shown to have a similar form, the identification would be more certain, but reliable measurements have yet to be made.

In contrast to the uncertainties about many features of the airglow, the cause of one phenomenon of particular interest has been discovered. It is the predawn enhancement of the red oxygen lines. In middle latitudes and during the winter months the intensity of the red line at 6,300 angstroms decreases rapidly after the end of twilight. It reaches a minimum soon after local midnight and then begins quite abruptly to increase. The time of this sudden rise is unrelated to the time of the local sunrise; indeed, in midwinter the rise begins several hours before local sunrise.

It turns out that the time of the abrupt increase is the time of sunrise in the ionosphere at the conjugate point in the opposite hemisphere. A conjugate point is one of two ends of any of the lines of force of the earth's magnetic field. If one imagines oneself standing at a point in the Northern Hemisphere with a line of magnetic force starting there and curving into space, the conjugate point is where the line rejoins the earth in the Southern Hemisphere. When it is winter in the Northern Hemisphere, the sun rises earlier in the Southern Hemisphere. When its rays reach the level of the ionosphere that is about 300 kilometers above the earth, the ultraviolet radiation ionizes the atoms there. The liberated photoelectrons stream along the magnetic lines of force and enter the conjugate hemisphere from above. There they excite the oxygen atoms and cause them to radiate the red lines. The sudden enhancement of the radiation, once begun, continues for perhaps an hour and a half. Then the rate of increase diminishes and the intensity remains more or less constant until the time of local sunrise approaches.

The airglow is by no means a well-behaved phenomenon. Its brightness is not uniform over the sky. Each emission varies in its own characteristic way with the time of day or night, with the season of the year, with latitude and with magnetic activity. The causes of these variations serve to illustrate the complexity of the various processes taking place in the upper atmosphere and the extent to which they are interrelated.

© 1971 SCIENTIFIC AMERICAN, INC



© 1971 SCIENTIFIC AMERICAN, INC

Why the Stomach Does Not Digest Itself

The organ secretes a strong acid that can dissolve metals and kill living cells. How it avoids digesting itself and what happens when the safety mechanisms fail are explained

by Horace W. Davenport

Ever since René Antoine Ferchault de Réaumur, the 18th-century man of many sciences, showed that juice secreted by the stomach could digest meat, laymen and physiologists alike have been puzzled by the question: Why doesn't the stomach digest itself? One answer, of course, is that it sometimes does. Under certain circumstances the gastric juice can produce ulcers and even destroy most of the stomach lining. Normally, however, the stomach wall staunchly resists attack; as Claude Bernard observed, it behaves as if it were made of porcelain.

The gastric juice contains hydrochloric acid, one of the most corrosive acids known. At the concentration secreted by the stomach lining the acid is capable of dissolving zinc and is deadly to cells. Yet in the stomach the hydrochloric acid ordinarily acts only to perform the useful functions of killing bacteria in the ingested food and drink, softening fibrous foods and promoting formation of the digestive enzyme pepsin. The corrosive juice is prevented from attacking the stomach wall by a complex physical-chemical barrier that is not yet fully understood.

Let us examine the events in the stomach. When food is delivered from the esophagus, it is deposited temporarily in the stomach's large main body, where it lies undisturbed as in a hopper. The main body's muscular wall is weak; it stretches to accommodate the volume of the meal and its peristaltic contractions are feeble. Gradually the food passes along to the antrum, the smaller end of the stomach pouch, where the principal gastric action takes place. The antrum has a strong muscular wall that contracts vigorously, kneading the food, mixing it thoroughly with the stomach's digestive juices and then delivering it to the duodenum of the small intestine [*see illustration on next page*].

SURFACE OF HUMAN STOMACH'S inner lining, which is the glandular mucosa, is seen enlarged some 700 diameters in this scanning electron micrograph made by Jeanne M. Riddle of the Wayne State University School of Medicine. The view shows the tops of epithelial cells, the gastric pits and characteristic folds of a normal stomach.

The juices are secreted by a tissue called the glandular mucosa, which forms the lining of the stomach wall. The mucosa contains "oxyntic" cells (named from the Greek root *oxy,* meaning sharp or acid), which secrete hydrochloric acid, and the "chief" cells, which secrete pepsinogen, which is then converted into pepsin by the hydrochloric acid. The juices issue from tubules deep in the mucosa and reach the stomach cavity through small pits.

Conceivably the pepsin, which cleaves the chains of the proteins in food, could take part in attacks on the stomach wall, but there is no good evidence that it does so. The hazard to the stomach undoubtedly lies predominantly in the hydrochloric acid, specifically in its hydrogen ions. When the acid is completely ionized in the lumen, or cavity, of the stomach, there is a steep gradient for diffusion of the hydrogen ions from the lumen into the stomach wall, whose cells and interstitial fluid, as elsewhere in the body, are neutral in that the concentration of hydrogen ions within them is very low. Consequently hydrogen ions would rush back from the stomach, where their concentration is high, into the mucosa that secreted them if they were not held off by a barrier. That barrier resides in the surface layer of the mucosa's structure, which is composed of tall, columnar epithelial cells.

The cells are fused together at their apexes, forming tight junctions. Few hydrogen ions manage to get through these junctions and penetrate into the mucosa. Similarly, the barrier blocks the movement of ions from the mucosa into the stomach cavity. The interstitial fluid bathing the cells in the mucosa contains ions of various substances, notably sodium, because the ions move freely into the fluid from the capillary blood vessels nourishing the cells. Therefore the existence and performance of the barrier can be studied in terms of its resistance to the passage of hydrogen ions into the mucosa and to the passage of sodium ions out of the mucosa. The nature of the barrier is now being investigated experimentally by examining its behavior in response to a variety of treatments.

A standard procedure has been developed for observing the performance of the stomach mucosa in dogs. Part of the stomach is formed surgically into a separate pouch that drains out through a tube in the abdominal wall. (Dogs fitted with such pouches can live in a healthy, cheerful state for many years.) In an experiment to test the mucosal barrier the first step is to establish its "normal" performance. The pouch is irrigated for 30 minutes with a known volume of a solution matching the normal gastric juice. From measurements of the content of hydrogen and sodium ions in the pouch fluid at the end of that time one can calculate the rate at which hydrogen ions have crossed the barrier into the mucosa and sodium ions have crossed from the mucosa into the stomach. These rates are taken to be the norm. The pouch is then irrigated for 30 minutes with the material whose effect on the barrier is to be investigated. Finally,

© 1971 SCIENTIFIC AMERICAN, INC

the effect is measured by supplying the pouch again with the normal solution to see how much change, if any, has been brought about in the barrier's performance by the treatment with the experimental material.

Similar tests, without surgery, can be conducted with human subjects and have been done with volunteers. The subject's stomach is washed out and filled for a certain period with a carefully measured solution resembling the human gastric juice. At the end of the period the solution is removed from the stomach by means of a tube for analysis and for determination of the barrier's normal performance, and the material or condition whose effect on the barrier is to be tested is administered to the subject. In general the effects of various treatments have proved to be much the same in man as they are in dogs, so that it appears the barrier functions about the same way in both species.

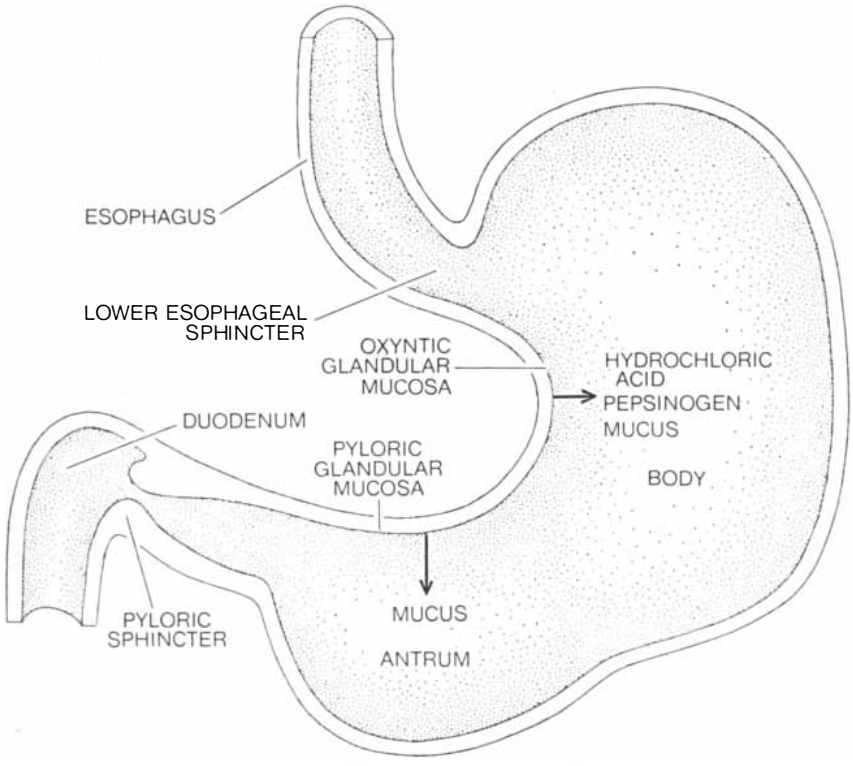
What have these investigations told us about the character and mode of operation of the barrier? The answers provide some interesting new insights in physiology and biochemistry.

The experimental treatments applied to the stomach mucosa were based on the assumption, which proved to be well founded, that the membrane of the cells covering the surface of the mucosa is like that of other cells of the body in construction and properties. Cell membranes generally consist largely of an ordered array of lipids forming a layer that is traversed by a few water-filled pores, which admit the passage of small, water-soluble molecules. Now, it is well known that a detergent such as is used in washing dishes, which removes grease from a plate by forming a water-soluble complex with the fat molecules, can destroy a cell membrane by breaking up its array of lipids. Hence if the barrier in the stomach mucosa is formed by sheets of fat molecules, a detergent should destroy the barrier. Experiments with lauryl sulfate, a common component of commercial detergents, showed that this substance would indeed break the mucosal barrier. (A pointed reminder that one should not drink dishwater!)

One naturally asks next: Does the body contain natural detergents that could gain access to the stomach mucosa? The answer is obvious. The bile, one of whose main functions is to emulsify fats in the food so that they can be digested, must contain such substances. The bile does in fact supply detergents in two forms. One consists of the bile salts, which are flat molecules that are polar and soluble in water on one side and nonpolar and soluble in fats on the other. By combining with water on one side and a fat molecule on the other, bile-salt molecules can act to break up a lipid structure and disperse its molecules. The other detergent in bile is lysolecithin, also a polar-nonpolar molecule. Lecithin, like the bile salts, is synthesized by the liver, and after it is secreted into the small intestine an enzyme secreted by the pancreas converts lecithin to lysolecithin. This substance breaks up cells by removing the lipids from their membranes.

Experiments demonstrated that either a bile salt or lysolecithin could produce the predictable effect on the stomach wall. A solution of one of these detergents, when applied to the mucosa of the pouch formed from a dog's stomach, broke the mucosal barrier. As a result hydrogen ions diffused rapidly from the lumen into the mucosa, and sodium ions similarly diffused from the mucosa into the lumen.

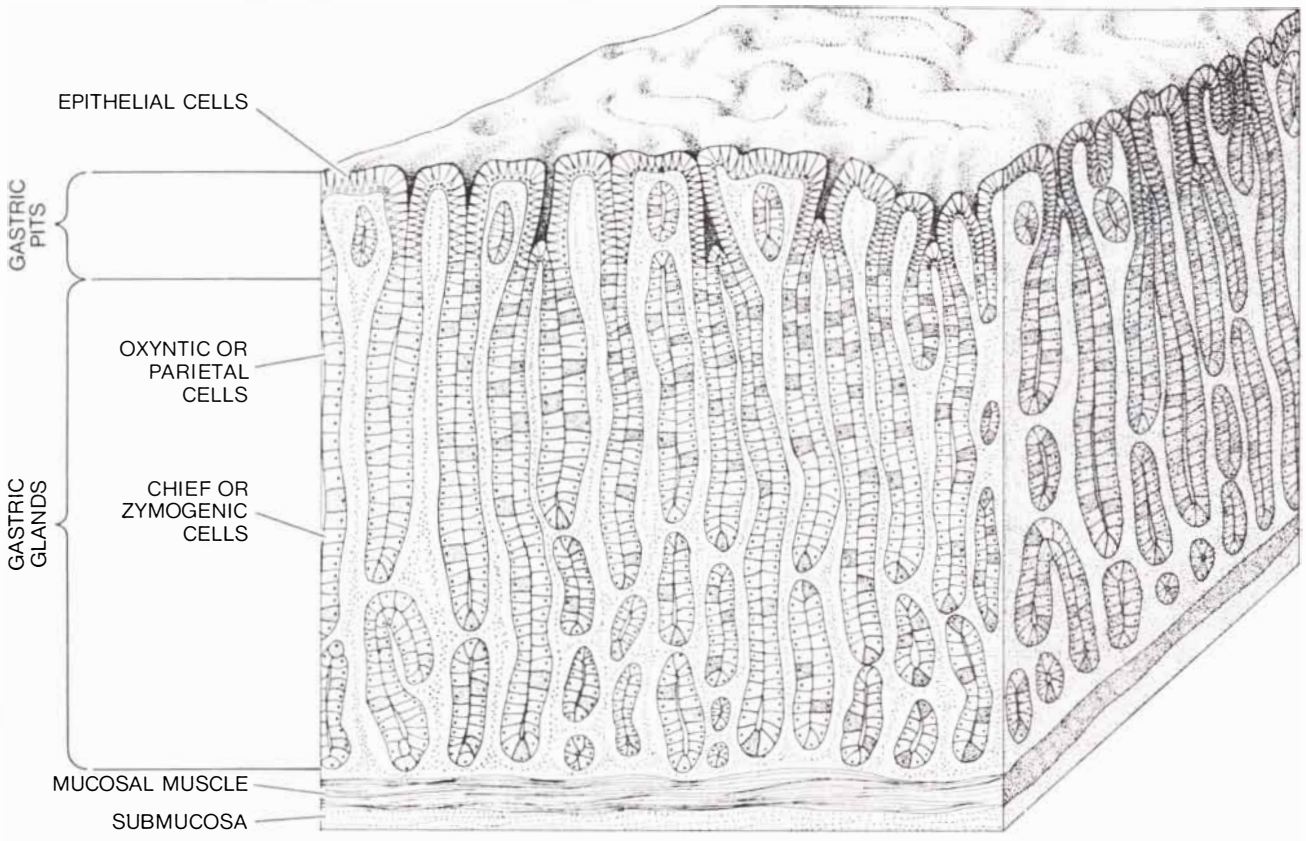
Bile is secreted into the small intestine, and ordinarily the motility of the small intestine carries it down and away from the stomach. The bile may, however, be carried back into the stomach by regurgitation from the small intestine of food undergoing digestion. In fact, patients with gastric ulcers are frequently found to have bile salts in the stomach. Many gastroenterologists now believe that regurgitated bile salts, which attack the gastric mucosal barrier by detergent action, may be a common cause of stomach ulcers.



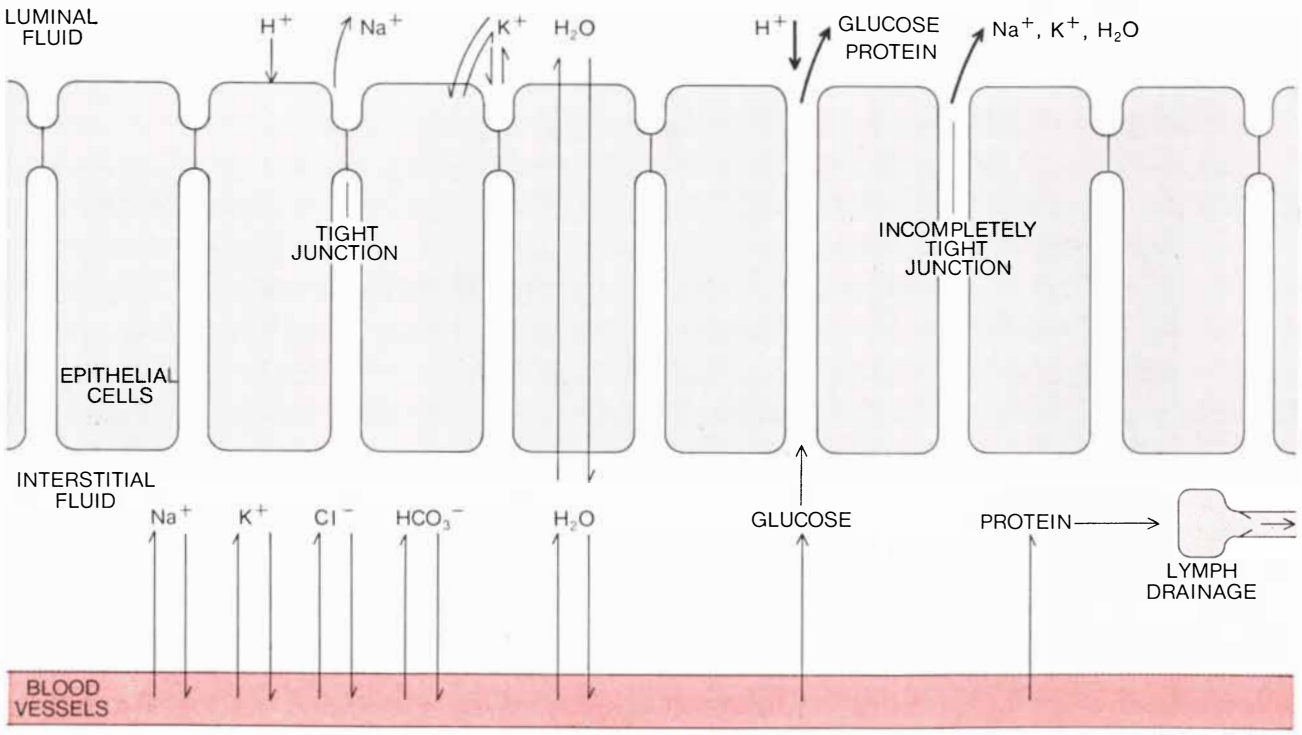
HUMAN STOMACH has two main parts: the body, which acts as a hopper for food, and the antrum, where food is mixed with digestive juices. The food then passes through the pyloric sphincter into the first part of the small intestine, the duodenum. Digestive juices, hydrochloric acid and pepsinogen, are secreted by cells in the mucosa of the stomach body.

Before considering external factors that may contribute to breaking down the barrier, let us examine briefly the effects of such a breakdown and the flooding of hydrogen ions into the mucosa. The invasion of acid stimulates the nerves within the stomach wall, thus stimulating its muscular contractions, and these strong contractions apparently account for the pangs of pain in a patient with peptic ulcer. The acid attack also releases histamine from stores within the mucosa and increases the rate of synthesis of histamine. This powerful substance, as is well known, produces active responses to attacks on tissues; it is responsible, for example, for inflammation and swelling. In the breached stomach mucosa histamine stimulates an increase of the mucosa's secretion of acid. Histamine dilates the precapillary

© 1971 SCIENTIFIC AMERICAN, INC

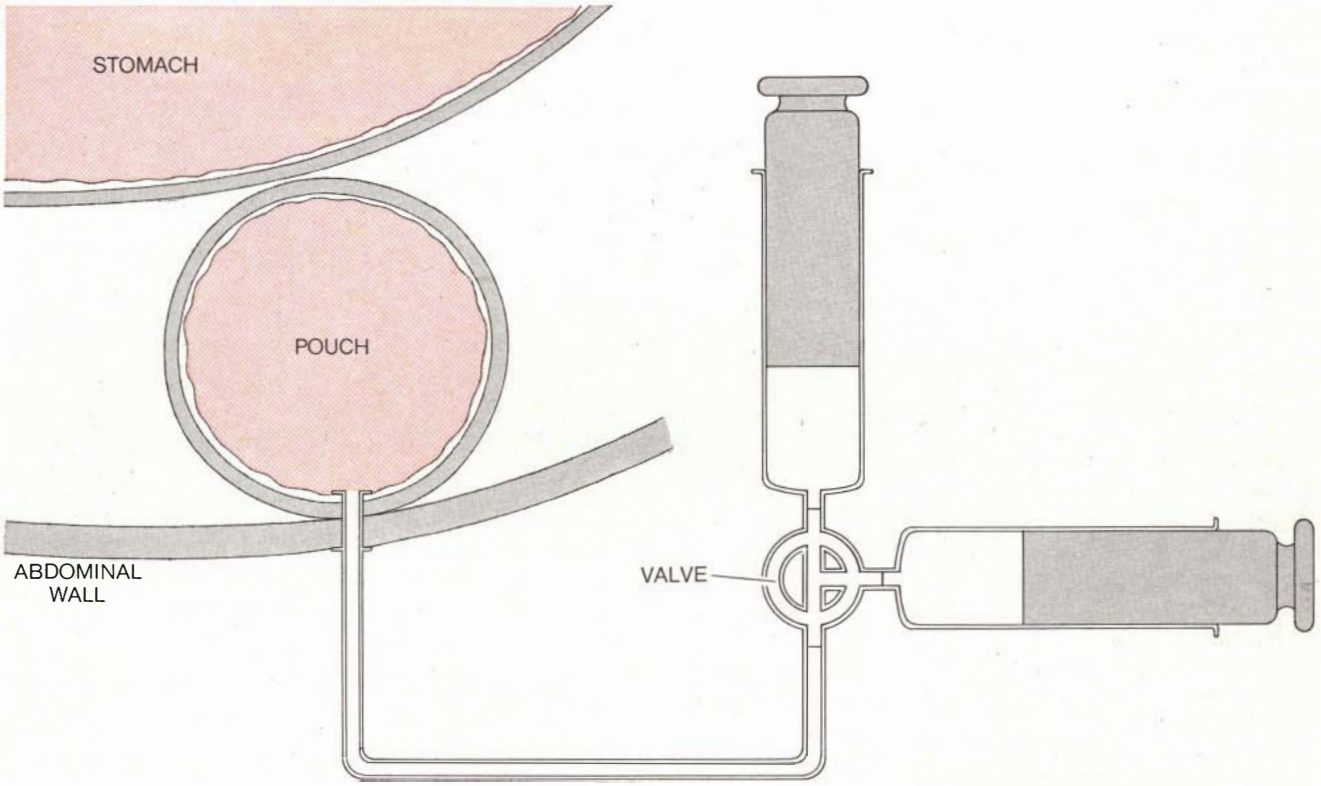


CROSS SECTION OF MUCOSA in the stomach body shows arrangement of gastric cells. Epithelial cells, which secrete mucus, cover the surface of the stomach and also line the gastric pits. The oxyntic (or parietal) cells produce hydrochloric acid; the chief (or zymogenic) cells produce pepsinogen. These cells lie in deep tubules; their secretions reach the surface through the gastric pits.

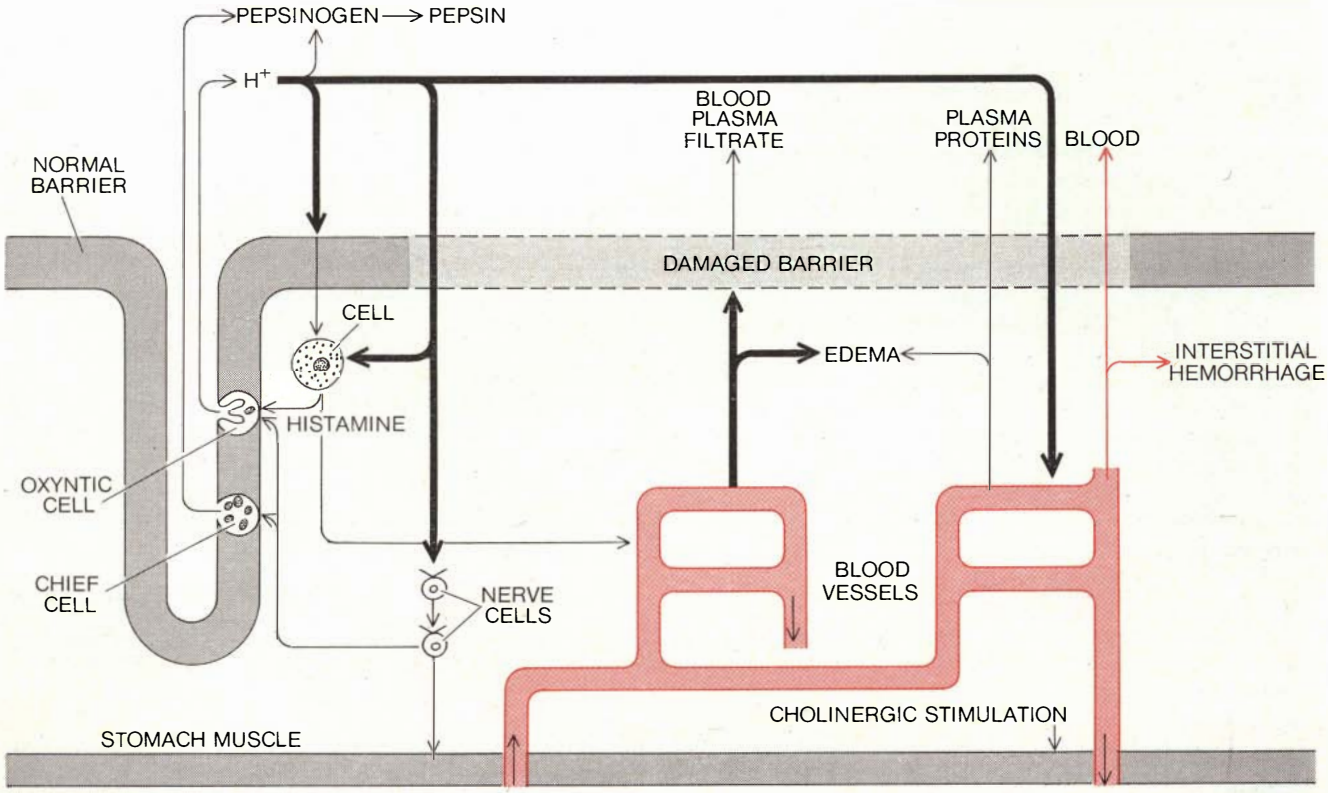


TIGHT JUNCTIONS formed by fused membranes near the tips of the epithelial cells provide a barrier to diffusion. Only small fluxes of ions pass through intact junctions. Water fluxes are large but equal in both directions. Broken junction results in increased ion fluxes and leakage of interstitial fluid and its constituents into the lumen. Blood vessels are freely permeable and nourish the cells.

© 1971 SCIENTIFIC AMERICAN, INC



GASTRIC POUCH in dogs is used to measure effects of various substances on gastric secretions. The pouch, about 50 milliliters in volume, is formed surgically under anesthesia. It is sealed off from the stomach and can be filled and drained through a tube in the abdominal wall. An arrangement for irrigating the pouch with specific solutions, consisting of two syringes and a valve, is shown.



EFFECT OF ACID on normal and broken mucosal barrier is portrayed schematically. Hydrogen ions from hydrochloric acid cause no harm as long as tbe barrier is intact. When tbe barrier is broken, the hydrogen ions rapidly penetrate the underlying cells and stimulate them to release histamine, which in turn stimulates the secretion of more acid and causes the blood vessels in the mucosa to dilate and become more permeable. Hydrogen ions also attack the blood vessels. Concurrent cholinergic stimulation raises blood pressure and hemorrhaging occurs. Nerve cells in the stomach wall stimulated by acid cause tbe stomach muscles to contract.

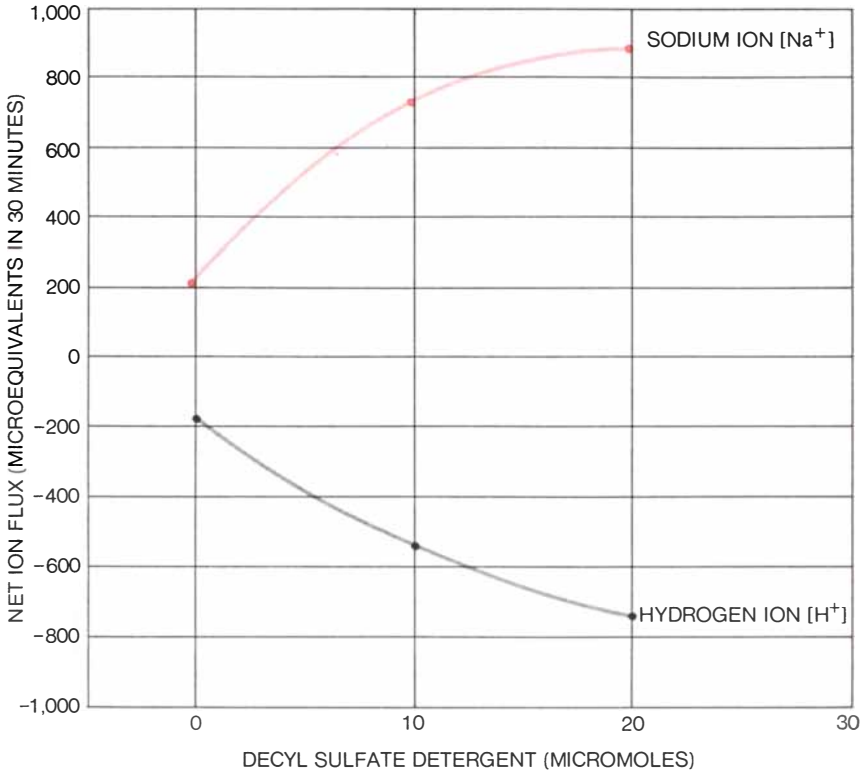
© 1971 SCIENTIFIC AMERICAN, INC

sphincters and capillaries of the mucosa, increasing the capillary blood pressure and blood flow. Furthermore, histamine makes the vessels more permeable, and as a result plasma proteins and fluid are forced out of the capillaries into the interstitial fluid around the mucosal cells, thereby producing edema. Rapid diffusion of acid into the mucosa can break down the capillary walls, and this is the cause of the drastic hemorrhaging that identifies severe cases of stomach ulceration. The stomach is particularly likely to bleed when rapid back-diffusion of acid into the mucosa is combined with vigorous contractions of the muscular wall.

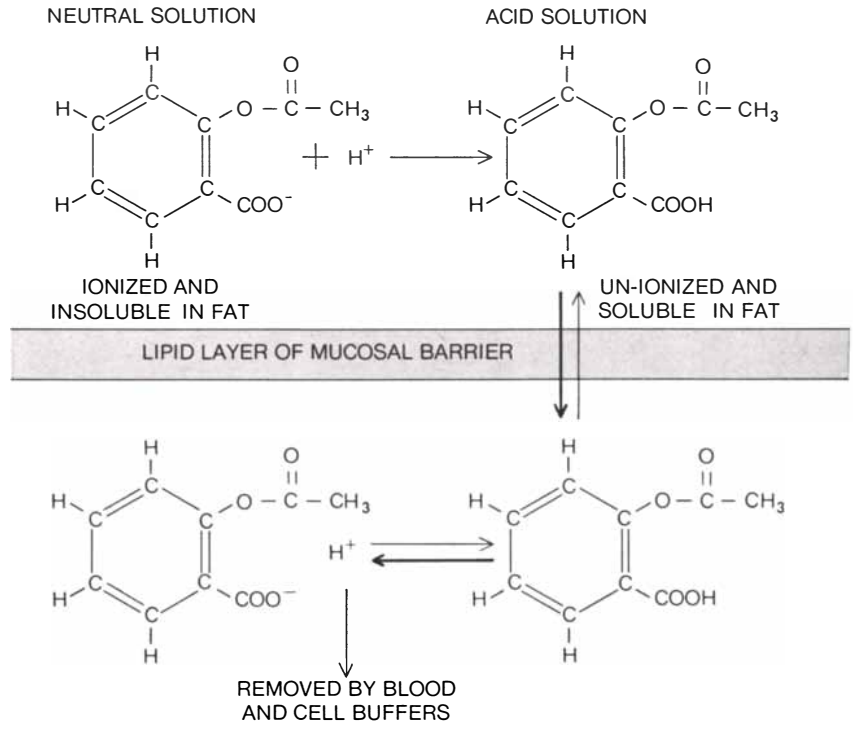
Fortunately the mucosa has a second safeguard against self-digestion in addition to the normal barrier to acid invasion. The mature cells of the stomach mucosa (and the mucosa of the rest of the digestive tract) are continuously desquamating, or flaking off, from the mucosal surface and are replaced by new cells produced within the tissue. Indeed, the human stomach normally sheds about half a million cells per minute. Thus the surface lining of the stomach is completely renewed every three days. By virtue of this rapid renewal the stomach wall can repair even severe damage of the mucosal barrier in a matter of hours or days. This ability to recover from insult accounts in large part for the fact that the stomach seldom does digest itself.

Let us look at some of the potential insults from external sources. The stomach mucosa is generally impervious to penetration by the ingredients of food and drink. By stimulating the secretion of hydrochloric acid a meal does bring about a slight increase in the mucosa's reabsorption of hydrogen ions, which results in tiny, superficial hemorrhages during digestion of the meal; these, however, are transient and are quickly repaired. The only commonly ingested substance that can readily get through the mucosal barrier is ethyl alcohol, a small molecule that is soluble in fat as well as in water. Although alcohol itself does not usually damage the barrier, in combination with certain other substances it can have a damaging effect, and this casts some light on the nature of the barrier.

Recent experiments have shown that salicylic acid or acetylsalicylic acid (aspirin) can breach the mucosal barrier and produce bleeding. This has been demonstrated by labeling red blood cells in the body with radioactive chromium (Cr-51); measurement of the radioactiv-

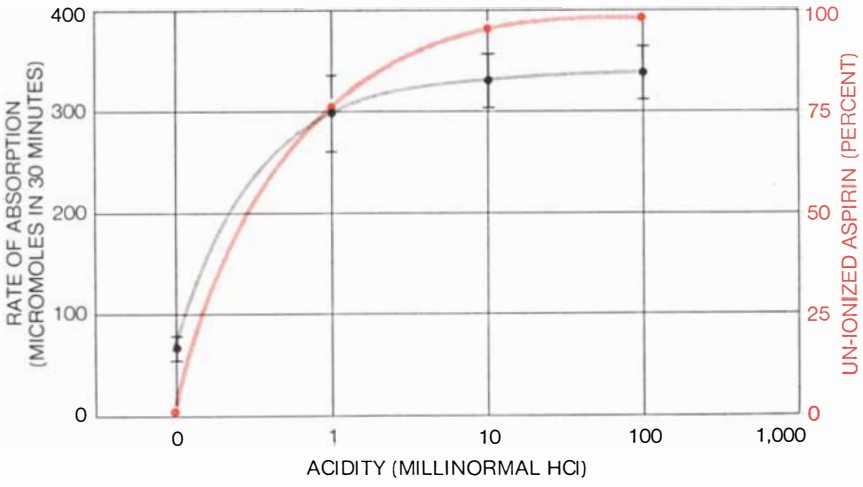


EFFECT OF DETERGENT on mucosal barrier was determined by washing a gastric pouch in a dog with a detergent and measuring the rate of diffusion of hydrogen ions and sodium ions. The diffusion rates indicate the extent to which the mucosal barrier has been damaged.

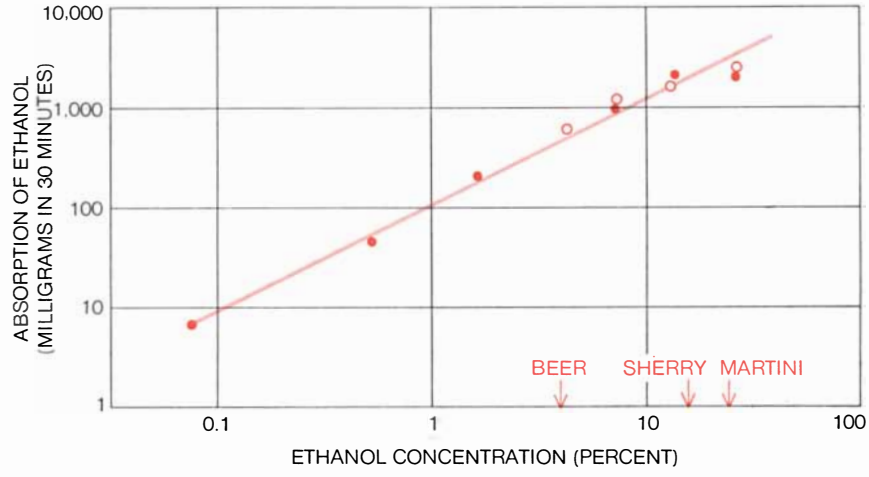


ASPIRIN (acetylsalicylic acid) dissolved in water or a neutral solution becomes ionized and is insoluble in fats. When aspirin is dissolved in the strongly acidic solution of the stomach, it becomes fat-soluble and passes through the mucosal barrier, damaging it in the process. Once past the barrier the aspirin molecules are ionized and become trapped.

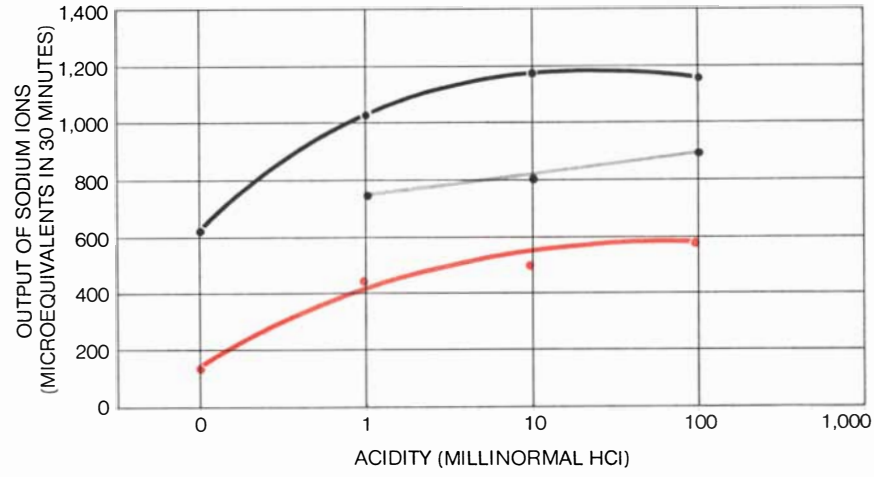
© 1971 SCIENTIFIC AMERICAN, INC



ABSORPTION OF ASPIRIN in the stomach is directly related to the acidity of the gastric juices. As acidity increases, aspirin becomes un-ionized and penetrates the mucosal barrier.



ETHANOL, or ethyl alcohol, which is soluble in fat and in water, can pass through the mucosal barrier and the rate at which it is absorbed is a linear function of its concentration. Ethanol in neutral solution (*open circles*) is absorbed at the same rate as ethanol in acid solution (*solid circles*). High concentrations of ethanol damage the mucosal barrier.



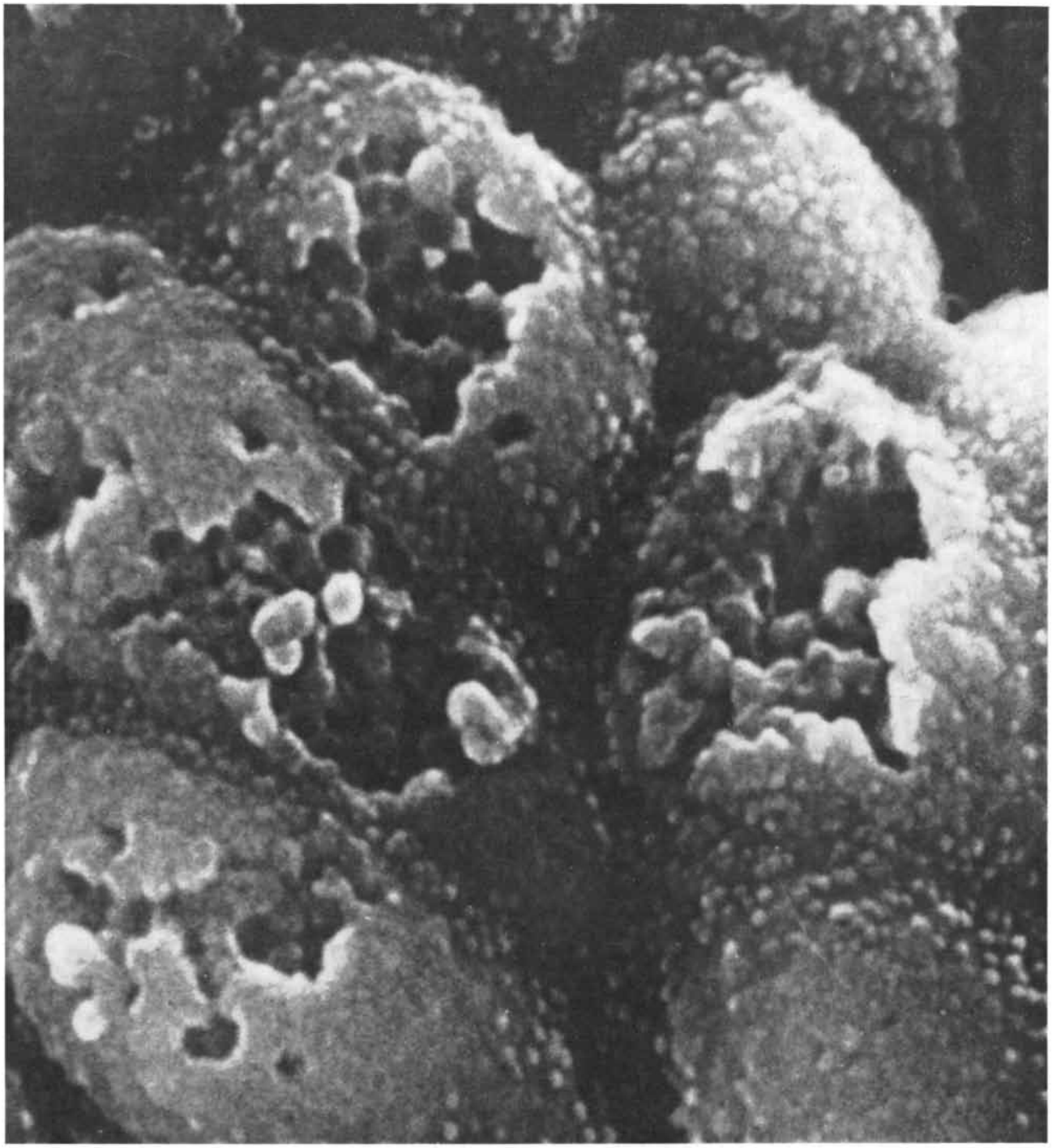
SYNERGISM OF ASPIRIN AND ALCOHOL in damaging mucosal barrier is shown by increased flow of sodium ions through the broken barrier. Aspirin (*colored curve*) in acid solution causes damage. A 4 percent solution of ethanol added to aspirin (*gray curve*) increases damage. Even greater damage is caused by aspirin and 9 percent ethanol (*black*).

ity in the feces then gives a measure of the amount of blood that has been lost by breakdown of the mucosal barrier. By this means it has been found that for most persons the bleeding produced by aspirin is trivial: the blood loss after taking two five-grain tablets usually amounts to only something between half a milliliter and two milliliters. Some susceptible people, however, may lose hundreds of milliliters of blood as a reaction to aspirin. Physicians have found that most patients coming to the emergency room of a hospital with massive bleeding of the upper gastrointestinal tract have taken salicylates within the preceding 24 hours. There have been cases in which habitual users of salicylates have developed profound anemia by losing blood at a rate higher than the body's production of red blood cells.

The effect of a salicylate in producing bleeding can be highly variable in different people or at different times. Why should this be? The present understanding of the stomach mucosal barrier provides the beginning of an explanation. The solubility of a salicylate, such as aspirin, in fat depends on the acidity of the medium in which it is dissolved. In a neutral solution aspirin becomes ionized, the hydrogen being removed from its carboxyl group (COOH) so that the group is left with a negative charge (COO^-). In that state the molecule is relatively insoluble in fat; consequently it does not readily penetrate the lipid barrier of the mucosa. In an acid solution, on the other hand, the carboxyl group remains un-ionized and aspirin is then soluble in fat and can diffuse rapidly through the mucosal barrier [*see bottom illustration on preceding page*]. Once in the mucosa, the salicylate is immediately ionized and prevented from crossing the barrier in the opposite direction. Thus a steep gradient for diffusion of the salicylate into the mucosa builds up. Salicylate pours into the mucosa at a rate that depends in the first instance on the acidity of the contents of the stomach cavity [*see top illustration at left*]. Experiments also show that if alcohol is present in the stomach, the ability of a salicylate to break the mucosal barrier is enhanced, even when the acidity of the stomach contents is low. These results have been found in human volunteers as well as in dogs with the experimental pouch. Cocktails and aspirin would therefore appear to be an insalutary combination.

The penetration of the salicylate into the mucosa has two damaging effects: (1) it breaks the barrier (by killing cells

© 1971 SCIENTIFIC AMERICAN, INC



DAMAGED EPITHELIAL CELLS on the surface of the inner wall of a human stomach are magnified about 12,000 diameters in this scanning electron micrograph by Jeanne Riddle. The tops of some cells have been destroyed, leaving holes in the mucosal barrier.

at the mucosal surface), thereby opening the gates for further influx of acid, and (2) it acts as a weapon to destroy the mucosa (by promoting bleeding).

I do not want to leave the impression that we now know all about why the stomach does not digest itself. As is always true in every report of a scientific investigation, we must conclude by saying we know only part of the story and "much more needs to be discovered." The stomach mucosal barrier can be broken in a number of different ways. Some people seem to have naturally weak barriers. Even when the barrier is strong, it may break down under the stress of severe illness or traumatic accident. It seems well established that nervous stress itself can bring about the barrier's destruction. All these matters need much more detailed investigation.

It is encouraging that we are beginning to learn the physical and chemical nature of the mechanisms by which the stomach avoids digesting itself. With further information along these lines it should become possible to find effective medical means for averting breakdown of the mucosal barrier, for protecting a mucosa that is unable to protect itself and for assisting the mucosa in its remarkable function of self-repair.

© 1971 SCIENTIFIC AMERICAN, INC

Population Genetics and Human Origins

Amid the bewildering array of early fossil hominoids, is there one whose morphology marks it as man's hominid ancestor? If the factor of genetic variability is considered, the answer appears to be no

by Robert B. Eckhardt

In the century or so that man has been examining his own immediate ancestry, almost every new fossil discovery has led to the revision, or at least the refinement, of earlier hypotheses about human evolution. One is reminded of G. K. Chesterton's fictional detective, Father Brown, and his search for the missing Earl of Glengyle. As Chesterton relates in "The Honour of Israel Gow," the unaccountable absence of the Earl from Glengyle Castle was by no means the only puzzle in the case. In the castle itself were piles of precious stones without any setting, loose snuff heaped on mantelpieces and tables, odd collections of tiny metal wheels, gears and springs, and numerous candles but not a single candlestick.

Challenged by the police to make sense of the situation, Father Brown offered three hypotheses. Could the Earl have detested the French Revolution and therefore been reenacting the family life of the last Bourbons? Snuff was taken by 18th-century aristocrats, and candles were the customary illumination of the time. Perhaps the springs and gears had something to do with the fact that Louis XVI was an amateur locksmith and the gems were somehow associated with Marie Antoinette. Alternatively the Earl might have been a secret burglar. The candles would be for a lantern, diamonds and metal wheels for cutting the glass out of windows and the snuff for throwing in the face of pursuers. Father Brown's third hypothesis was the least elaborate but the most shocking: the Earl had been dabbling in black magic.

Asked if he really believed any of this, the priest replied that he certainly did not but that he wanted to show how several different explanations could take all the evidence into account with equal consistency. "Ten false philosophies will fit the universe," Father Brown observes. "Ten false theories will fit Glengyle Castle. But we want the *real* explanation of the castle and the universe."

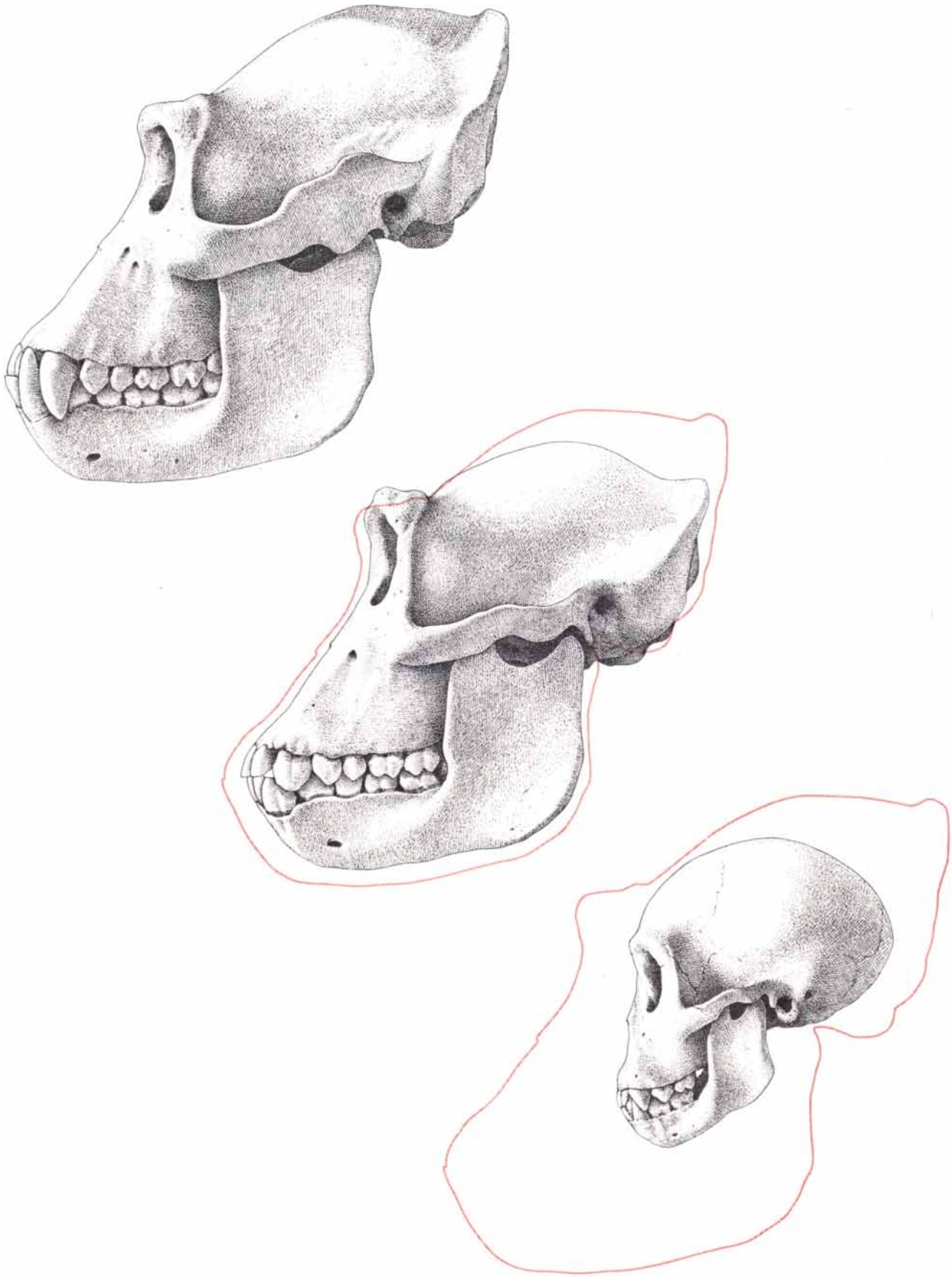
So it is with the origins of man: many explanations will fit the evidence but only one can be wholly correct. The general picture is clear enough. As certainly as Father Brown and the police knew that the Earl of Glengyle was missing, we know that man's ancestry lies among the Old World primates of Miocene and Pliocene times, a period extending from 20 million to three or four million years ago. That much was known to Charles Darwin and Thomas Huxley in the middle of the 19th century. There was paleontological evidence in the form of a primate lower jaw and associated fragments of limb bone that had been discovered in France by Alfred Fontan in 1856. The fossils were described by Édouard Lartet, who gave the chimpanzee-sized animal the generic name *Dryopithecus* ("forest ape") and the specific name *fontani* (in honor of Fontan). The geological formation that held the bones had been laid down some 14 million years ago in late Miocene times. *Dryopithecus* was clearly old enough to be ancestral both to man and to man's closest living relatives: the great apes [*see top illustration on page 96*].

Since that time fossils of *Dryopithecus* have been found in other deposits of Miocene and Pliocene age not only in Europe but also in Africa, the Middle East and the Far East. These numerous finds have had a twofold result. First, they amply verify the general conclusion that Darwin and Huxley and their contemporaries had reached: that the hominoids (members of the taxonomic superfamily that includes both man and the apes) of Miocene and Pliocene times were ancestral to the hominids (members of the genus *Homo,* of whom man is the only living representative). The second effect has been not to clarify but to confuse. Not all dryopithecine fossils are identical. The remains of some individuals bear a closer resemblance to those of man than the remains of others do. As a result students of human evolution have attempted to pick and choose from among a wide array of fossil hominoids certain ones that might have a claim to being the earliest true hominids.

One of the first such candidates for a place in the ranks of early man was put forward in 1910 by Guy Pilgrim of the Geological Survey of India. This was *Sivapithecus indicus.* (It was the fashion at the time to give the fossil apes of India names drawn from the Hindu pantheon.) The genus and species had originally been established on the basis of a single tooth—a right lower third molar—discovered in the Siwalik Hills, a fossil-rich area north of New Delhi. When a partial lower jaw was found there several years later, it gave Pilgrim a basis for a remarkably manlike reconstruction [*see illustration on page 98*]. The American paleontologist William K. Gregory demonstrated in 1916, however, that the jaw fragment could be interpreted much more conservatively. Gregory's alternative reconstruction made it evident that *Sivapithecus* was another of the widely distributed dryopithecines and apparently not a hominid.

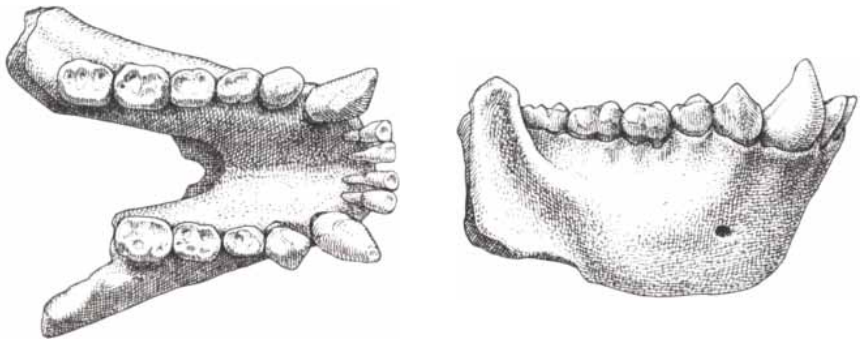
From time to time since then charter membership in the family of man has been extended to various other Miocene and Pliocene hominoids. Among them is *Oreopithecus bambolii,* a species first unearthed in a lignite mine on Monte Bamboli in central Italy just a century ago. The discovery of a number of well-preserved *Oreopithecus* fossils since that time, including a nearly complete skeleton unearthed in 1958, has led Johannes

© 1971 SCIENTIFIC AMERICAN, INC



TWO EXAMPLES of morphological variability are presented by the three primate skulls shown here. These are the variabilities that result from differences in sex and in age. The skull of an adult male gorilla (*top*) is larger and more robust than the skull of a female gorilla of equal age (*middle*); the contrast becomes obvious when an outline of the male's skull (*color*) is superposed. The immature gorilla (*bottom*) is unlike either adult. In spite of these morphological extremes all three are members of the same species.

© 1971 SCIENTIFIC AMERICAN, INC



ANCESTRAL HOMINOID *Dryopithecus* was first found by Alfred Fontan in France in 1856; the type specimen, *D. fontani*, consisted of a chimpanzee-sized lower jaw, shown two-thirds natural size, and some limb-bone fragments. Dryopithecines have since been found in fossil deposits scattered throughout the Old World (*see illustration at bottom of page*).

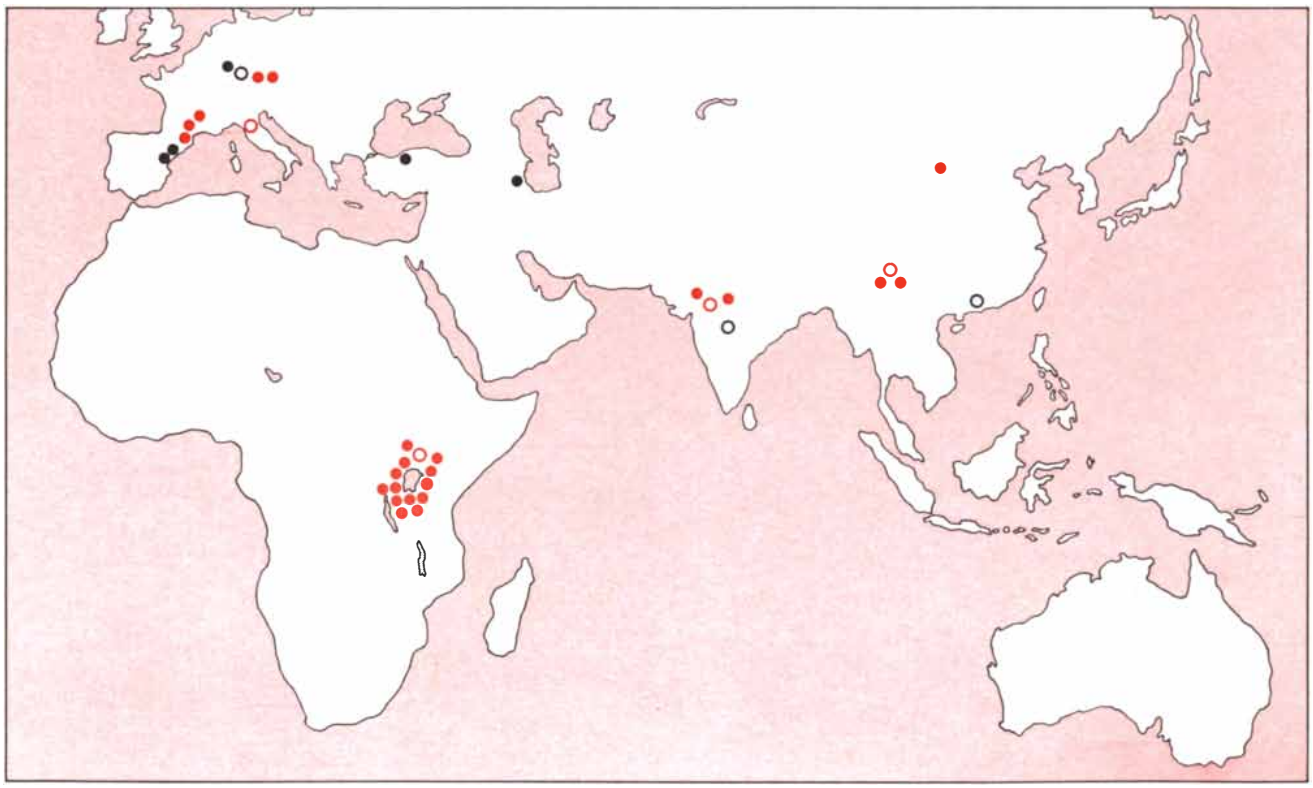
Hürzeler of the natural history museum in Basel to propose hominid status for this primate. The animal's flattened face and some features of its pelvis provide the main reasons for its designation as a hominid, but Hürzeler's proposal has not gone unchallenged [see "The Early Relatives of Man," by Elwyn L. Simons; SCIENTIFIC AMERICAN, July, 1964].

More recent nominations concern certain fossils found in India and in Africa. The Indian finds are also from the Siwalik Hills; they were assigned to the genus *Ramapithecus* by G. Edward Lewis in the 1930's. The African ones are believed by L. S. B. Leakey to represent two species within a genus he has named *Kenyapithecus*. It is perhaps typical of the problems that beset primate paleontology that these individuals have been deemed to represent as many as three distinct species of early hominids even though there is no certainty that the hominid line had yet diverged from the dryopithecine stock in general.

The problem arises primarily because the taxonomic position of any new fossil is determined on the basis of exquisitely detailed morphological studies of isolated specimens. The analysis is almost always conducted on fossil jaws and teeth because these dense and durable parts of the body are the most likely to be preserved. What this procedure tends to ignore is that among such living hominoids as chimpanzees the jaws and teeth exhibit a high degree of morphological variability. There is no reason to believe the same was not true of hominoids in Miocene and Pliocene times.

Indeed, are there any grounds for assuming that morphological evidence alone makes it possible to draw a valid distinction between the majority of these early hominoids and some ancestral hominid that may be concealed among them? In view of the morphological variability among living hominoids, I think not. It would appear that, by scarcely considering the role of variability in evolution, paleontologists have emphasized taxonomic "splitting" at a time when the record of man's animal ancestry might better be clarified by considerable "lumping."

What are the causes of variability? One of them is age. Among primates, as among most mammals, mature animals are inevitably more robust than



OLD WORLD HOMINOIDS of the late Miocene (*color*) and early Pliocene (*black*) are found in fossil strata from eastern Spain to northern China. The solid circles represent sites where various dryopithecine fossils have been unearthed. The open circles show genera recently proposed as ancestral hominids. These include, from left, a German specimen of *Ramapithecus*, the Italian primate *Oreopithecus*, *Kenyapithecus* of East Africa and finally specimens of *Ramapithecus* and *Gigantopithecus* found in India and China.

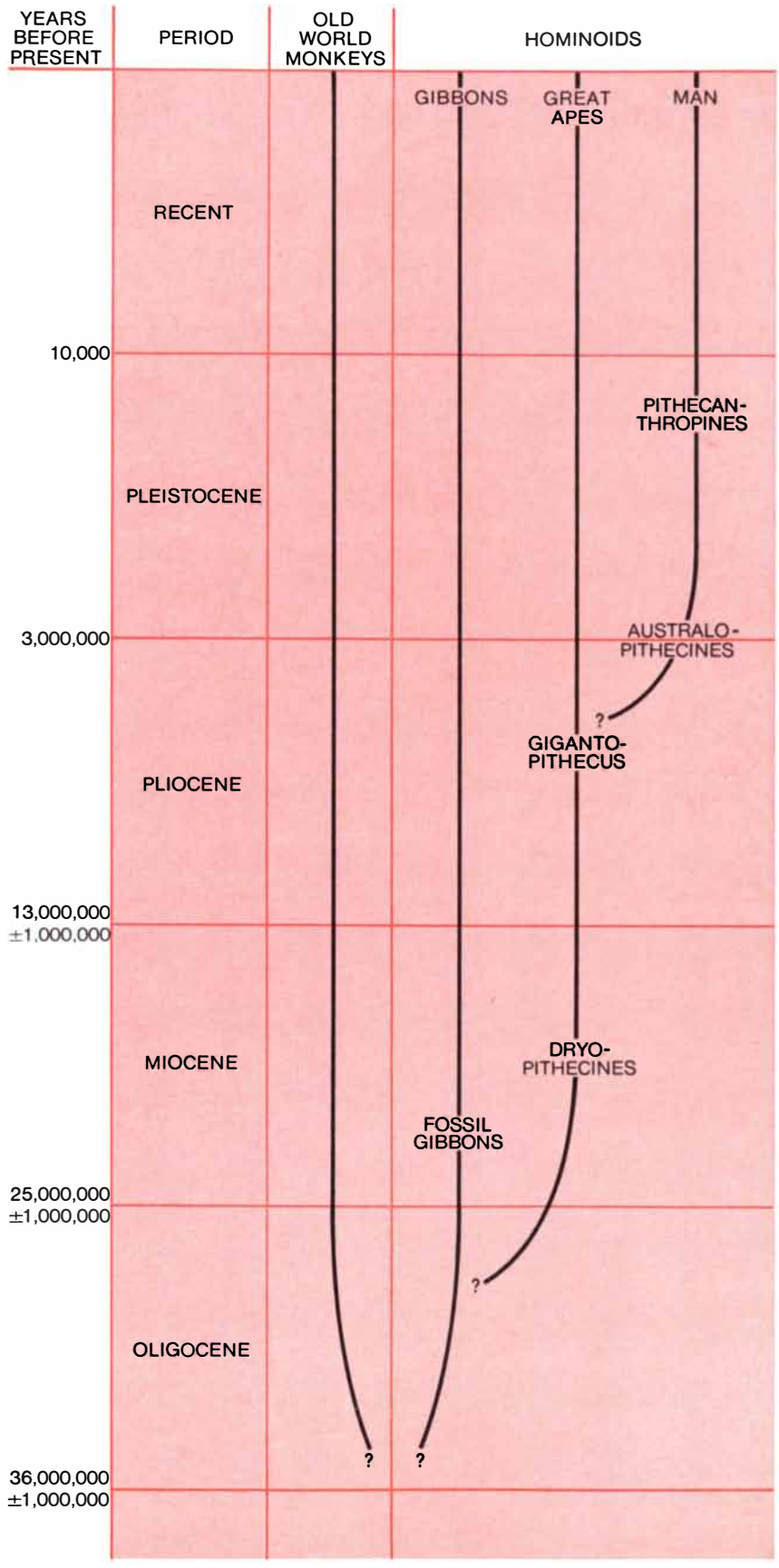
© 1971 SCIENTIFIC AMERICAN, INC

young ones. Another is sex: the males of most primate species are noticeably larger and more rugged than the females. A third cause is geographical. Populations of wide-ranging species such as chimpanzees and gorillas show specific adaptations to the differing ecological conditions encountered in different regions. One may cite the differences that exist in height of face, shape of mouth, length of trunk and limb and thickness of coat between the mountain gorilla and the lowland gorilla.

To see what this means with respect to man's early ancestry, let us begin with the assumption that no species of fossil hominoid exhibited any less variability attributable to causes such as these than any ape species does today. The variability that can be observed within populations of chimpanzees, gorillas and orangutans will therefore serve as a minimum standard of variability. If the variability among an assemblage of hominoid fossils does not exceed this minimum, we should be reluctant to conclude that the assemblage represents more than a single species.

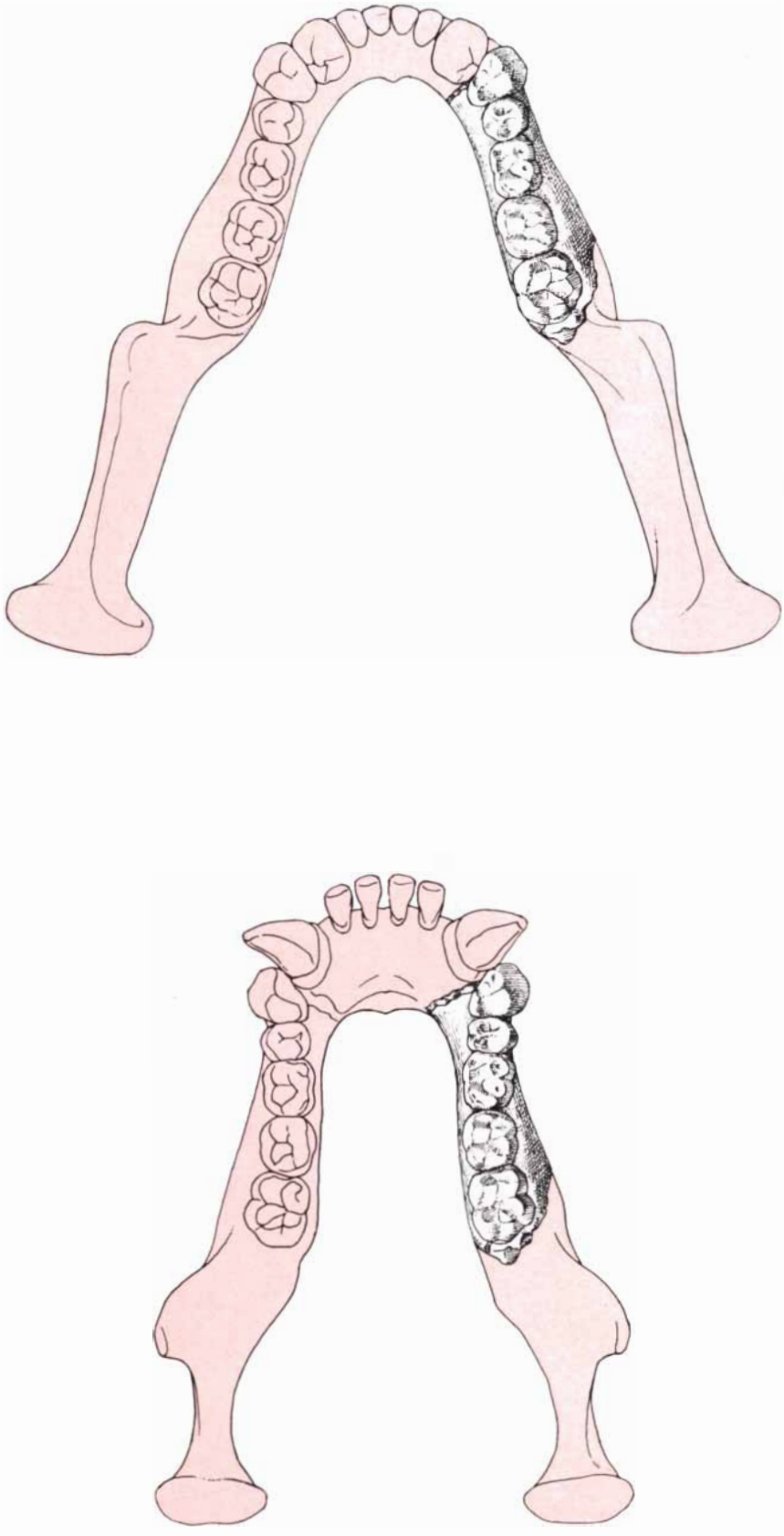
What if the fossils exhibit a greater variability? Then a fourth factor must be taken into account: time. As long as this element is ignored, an important source of inherited variation is in danger of being overlooked. Rarely, if ever, does a collection of fossils come from one point in time, so that all its individual members could have been born in the same generation. On the contrary, the individuals often belong to populations that are separated by as much as a million years. Even when all the specimens in a collection constitute a single line of descent, the likelihood is that the genetic changes over many generations will produce a range of variation among the related fossils that greatly exceeds the range encountered within any one generation of a living species. Does the factor of genetic change therefore make it impossible to judge whether or not a collection of fossils represents more than one evolving line? I believe not. It does, however, demand some careful reasoning and quantitative analysis by the investigator.

A change in gene frequency—the phenomenon that ultimately underlies a change in some measurable character such as tooth size—may result from any one of several causes. One is the process of preferential survival and reproduction of different genetic constitutions that we call natural selection. Another is random genetic drift, a factor that makes for change within relatively small populations. A third is gene flow: the transmis-



HOMINOID LINES probably split off from the line that led to the Old World monkeys at some point more than 36 million years ago. It is uncertain just when the hominid line, which includes two accepted precursors of modern man, the australopithecines and pithecanthropines, split off from the lineage of the great apes in terms of behavior and ecological position rather than morphology. The question marks denote additional uncertainties.

© 1971 SCIENTIFIC AMERICAN, INC



EARLY RUNNER in the race for status as an ancestral hominid, *Sivapithecus indicus* has a very manlike appearance in a reconstruction by Guy Pilgrim (*top*); the fragment of fossil jaw appears in black and the reconstruction in color. The same fossil fragment was reconstructed more conservatively by William K. Gregory, who considered it to be simply one more dryopithecine (*bottom*). The range of variation in the size of tooth crowns among dryopithecine fossils in general seldom exceeds the range found in living chimpanzees.

sion of genetic material between populations that are not reproductively isolated. A fourth is mutation; like genetic drift, it is a random factor. Of all these factors by far the most important as a source of adaptive change is natural selection. Therefore if we declare that the trends discussed here are exclusively the consequence of natural selection, our loss in precision will be slight.

The number and variety of generic and specific names that have been assigned to fossil hominoids are confusing, to say the least. One of the more misleading aspects of this taxonomic nightmare is the assignment of different names to earlier and later representatives of a single phyletic line. The difference implies that the specimens come from populations that were reproductively isolated, whereas the transmission of genetic material from the earlier population to the later was actually uninterrupted.

Even more confusion arises from the failure of many to give adequate weight to the imperfections inherent in the hominoid fossil record. Most of the specimens are less than precisely dated both in absolute and in relative terms. As a consequence individuals representing nothing more than successive stages of a single phyletic line have in a number of cases been assigned distinctive taxonomic names, as if they had actually been contemporaries and neighbors. Here we shall avoid these possible sources of confusion by asking with respect to any fossil collection how many phyletic lines it represents rather than how many species.

Specifically, the question can be phrased as follows: Given the known morphological variability in a fossil collection and making appropriate allowances for variability due to sex, age, environment and genetic change with time, must we conclude that the collection represents more than one phyletic line? To answer the question we must determine two sets of values. The first is the per-generation rate of measurable change that would be needed to produce the total change observed within a single phyletic line through the action of genetic factors. The second is the magnitude of the selection differential that would be required to produce such a rate of change. Once the two values are known they can be compared with the equivalent values in living populations. If the values for the fossil collection are higher than those for the living samples, we are forced to conclude that the fossil collection does indeed represent more than one phyletic line. If, however, the fossil

© 1971 SCIENTIFIC AMERICAN, INC

values are less than or equal to those known for living populations, then we cannot ignore the possibility that only a single evolving line is represented in the collection.

The first value—the rate of change per generation—is found as follows. One initially determines the amount by which the observed range of measurable variation in the fossil collection exceeds the range of variation in a living hominoid population. This figure is then divided by the estimated number of generations separating the earliest specimens in the fossil collection from the latest. The length of a generation must also be estimated. For example, among chimpanzees females reach sexual maturity between the ages of six and 10 years and males between seven and eight. For convenience in calculation one may therefore say that the length of a chimpanzee generation is approximately 10 years. In my calculations I have assumed that the length of a generation among the hominoids of Miocene and Pliocene times was about the same.

The determination of the second value—the selection differential required to change the mean of the measurements by a given amount from one generation to the next—depends on learning the heritability of the character, that is, the fraction of the total observed variation that is genetic in origin. This is necessary because the amount of mean change in one generation, commonly termed the "response to selection," is a product of the selection differential and the heritability of the character.

Heritability estimates are expressed in units that extend from 0, meaning that no variation is attributable to genetic factors, to 1, meaning that all variation is attributable to genetic factors. Heritability estimates, which play a vital role in the selective breeding of plants and animals, are based on the degree of resemblance observed between relatives. Such resemblances are of course due both to genetic and to environmental factors. This is to say that the phenotype—an organism's observable appearance—is the consequence of interaction between the environment and the organism's genetic constitution, the genotype. But only genes and not total phenotypes are transmitted from parents to offspring, and only genes are subject to change in frequency due to selection.

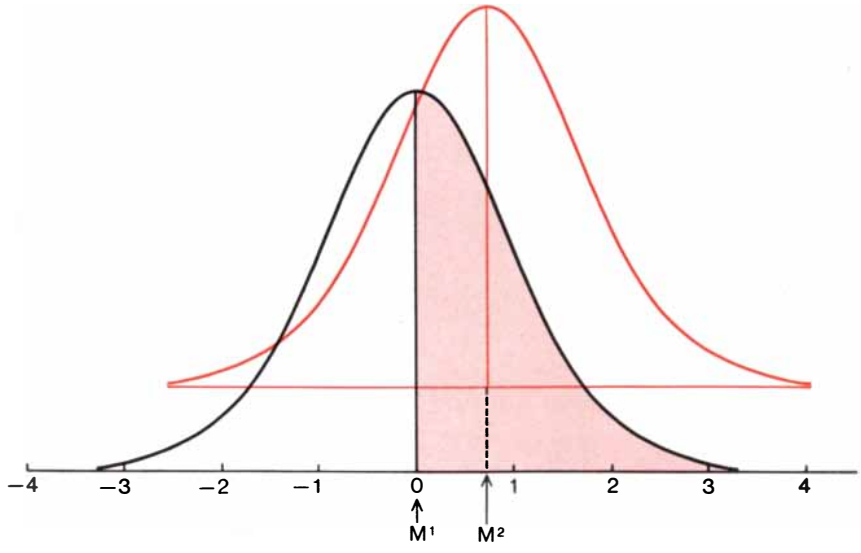
I took as my primary sample of living hominoids the small population of chimpanzees housed at the Yerkes Regional Primate Research Center in Atlanta, Ga. The population includes several generations. My measurements of the animals' teeth enabled me to estimate the heritability of tooth-crown length and width. Averaged over all the teeth in the upper and lower jaws, the heritability of crown length was approximately .1 and crown width approximately .2.

In comparison with what is known about the heritability of tooth size in man, both of these values are quite low. Moreover, due to the small size of the sample and the fact that heritability estimates depend on the particular balance of genetic and environmental factors in the population under examination, my estimates may not be applicable to chimpanzees in general. If there is error, however, it is preferable to err on the low side, because inferences based on low heritability estimates will necessarily be conservative ones. In any case I assumed that the heritability of tooth size among fossil hominoids was no higher than these figures.

I was now in a position to compare variability in tooth dimensions among living and fossil hominoids. The fossil collection I chose consisted of upper and lower teeth of various Miocene and Pliocene primates unearthed in India. The number of teeth in the collection ranged from a maximum of 16 lower second molars to no teeth at all in the case of certain upper and lower incisors. The entire collection consisted of 36 upper teeth and 60 lower teeth.

In addition to the estimates concerning generation length and tooth-size heritability, it was necessary to make certain subsidiary assumptions with respect to the fossil collection. The first of these is that the rate of change in tooth size had been approximately constant in successive generations over the six-million-year span that separates the earliest specimens from the latest. This assumption, which implies that the evolutionary forces affecting tooth size acted in a consistent direction at a fixed rate, cannot be entirely correct. During some parts of this long time interval selection may quite possibly have reversed or at least slowed the long-term trend. Therefore a logical alternative assumption might be that the rate of change did fluctuate. Such an assumption implies, however, that during some parts of the time interval tooth size must have changed in the net preferred direction even more rapidly. Thus the assumption of a constant rate of change appears to be a reasonable approximation of the actual situation, barring the least likely assumption of all, namely that the sum total of the variations observed in the sample is due to evolution before late Miocene and early Pliocene times and none of it is due to evolution that took place during this six-million-year interval. The fact is that we have little knowledge of specific events that might have influenced hominoid evolution during the interval, so that we have no way of gauging the magnitude or duration of fluctuations in the rate of phenotypic change or selection differential.

A second assumption is that the measurable changes that occurred over a period of time were due primarily to the



SELECTION DIFFERENTIAL that brings about evolutionary change affects a succeeding generation in the manner illustrated in this graph. If all the first-generation parents are drawn only from the half of a population that possesses some particular trait to a greater than average degree (*color*), the mean value of the trait in the next generation will be shifted from M^1 to M^2, a quantity equal to eight-tenths of one standard deviation. It is assumed in this exaggerated example that all parental couples produce the same number of offspring.

© 1971 SCIENTIFIC AMERICAN, INC

action of systematic evolutionary pressures rather than entirely random factors. Natural selection should have acted directionally to change the dimensions of a given tooth from one size to another, perhaps in response to climatic changes that had altered the nature of the food supply. Random factors, on the other hand, would merely have caused unpredictable fluctuations from generation to generation.

A further assumption is that variation within the fossil species stayed approximately constant from generation to generation. This is reasonable because among fairly large natural populations there are rarely any absolute barriers to the input of new variation and consequently little fixation attributable to genetic drift. We would not expect genetic variability to be depleted by the action of natural selection, since gene flow from neighboring populations could constantly renew it. This assumption has further significance, as we shall see. If the genetic variability in a population is not steadily depleted by selection and if environmental components of variation also remain unchanged, it follows that heritability—the ratio of genotype variance to total phenotype variance—should also remain approximately constant.

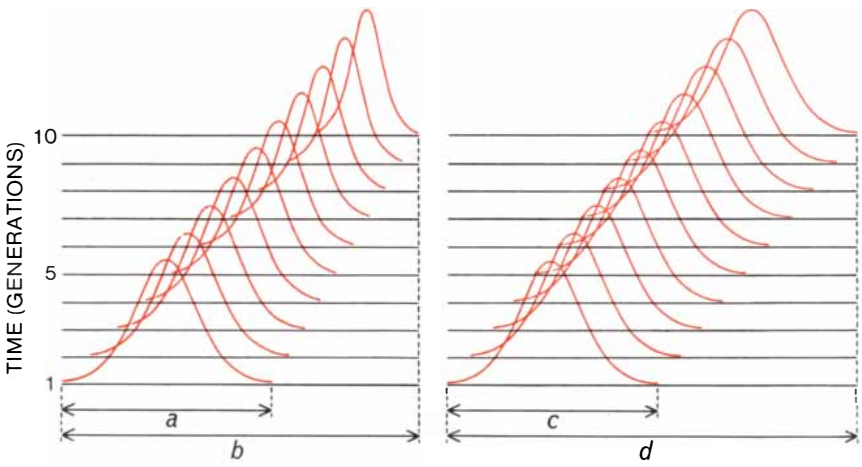
The fossil collection I analyzed contained the teeth of individuals believed to represent two *Dryopithecus* species, *D. sivalensis* and *D. indicus*, and one candidate for hominid status, *Ramapithecus punjabicus*. The collection consists of 16 fragments of upper jaws or loose upper teeth and 29 lower-jaw fragments or teeth. Only one upper central incisor is included and no lower central or lateral incisors. I compared the range of variation in tooth-crown length and width of the fossil molar, premolar and canine teeth with similar measurements taken from the chimpanzee population at the Yerkes Research Center and from a larger sample of wild chimpanzees in Liberia. A total of 24 measurements, 12 of tooth-crown length and 12 of tooth-crown width, were compared. Only nine of the fossil measurements (specifically the width of the upper canine, the length of the upper third molar, the width of the lower first premolar, the length and the width of the lower second premolar, the widths of the lower first and second molars and the length and the width of the lower third molar) showed a range of variation greater than the range of variation among the chimpanzees. In one instance (the length of the lower second molar) the fossil range of variation was exactly the same as the chimpanzee range. In the remaining 14 measurements the fossil range of variation was less than the chimpanzee range.

I then calculated the amount of change per generation that would have been required to account for the nine significantly larger fossil variations, assuming that the greater range among the fossils was due entirely to a shift in the mean within a single phyletic line of evolving hominoids and not to the presence in the collection of two or more lines. Expressed both as a specific measurable difference and as a percent change in the mean, the amount of change per generation over the six-million-year period proved to be very slight.

The greatest variation in tooth-crown length, represented by fossil lower third molars with an eight-millimeter range of variation as compared with a 3.2-millimeter range among living chimpanzees, would be accommodated by a measurable change per generation of only eight millionths of a millimeter, or .057 thousandth of 1 percent, in the mean length of the crown. The smallest variation—a difference of range in variation with respect to lower first molar width of little more than half a millimeter—would be accommodated by a change per generation of one millionth of a millimeter, or a little less than 100 thousandths of 1 percent. The selection differential per generation needed to effect both changes would not have exceeded .08 micron and .005 micron respectively. To achieve these small shifts in the means would have required very little selection pressure in each generation.

A significant comparison can be made between these calculations and the results of a 100-generation experiment conducted with fruit flies in the 1950's by the Italian geneticist Adriano Buzzati-Traverso. He maintained four populations of *Drosophila* in the laboratory over a 48-month period. Each population consisted initially of some 700 flies. No artificial selection pressures were applied, but as generation succeeded generation it became apparent that natural selection was leading to a number of morphological changes. One of them was a shortening of the wings in all four populations. By the 55th generation the mean wing length of the population most severely affected had become 1.31 percent shorter than the mean wing length in the smaller of the two parental strains responsible for the initial population.

On the basis of this difference it was calculated that the average rate of change in wing length per generation was somewhat greater than .24 thousandth of 1 percent. This was several orders of magnitude less than the rate of change per generation in laboratory populations exposed to artificial selection pressures. It is, however, nearly a full order of magnitude greater than the maximum rate of change per generation I have calculated from tooth-dimension variations among the Indian fossil hominoids. This strongly suggests that the calculation respecting the fossils is plausible and therefore that the two supposed genera and three supposed species



LONG-RANGE CONSEQUENCES of selection pressure are illustrated in exaggerated form with respect to a population (*left*) that is subject to depletion of genetic variability and a second population (*right*) that has its variability kept constant by gene flow from a neighboring population. In the first instance the range of dimensional variation (*a*) is less in successive generations, whereas in the second the range (*c*) remains constant. In both, however, the range of variation over many successive generations (*b, d*) greatly exceeds the range within any one. Because related fossil primates are usually individuals separated by thousands of generations, the range of variation within a single evolving line can be extreme.

© 1971 SCIENTIFIC AMERICAN, INC

		MEAN (MILLIMETERS)		RANGE OF VARIATION (MILLIMETERS)		CHANGE PER GENERATION				SELECTION DIFFERENTIAL PER GENERATION (MICRONS)	
						ONE MILLIONTH OF A MILLIMETER		ONE THOUSANDTH OF 1 PERCENT			
		LENGTH	WIDTH	LENGTH	WIDTH	LENGTH	WIDTH	LENGTH	WIDTH	LENGTH	WIDTH
UPPER CANINE	CHIMPANZEE	13.79	10.42	12.8	5.5						
	FOSSIL	15.88	13.12	7.6	7.2		2.8		.018		.014
UPPER MOLAR 3	CHIMPANZEE	9.36	10.79	4.5	4.0						
	FOSSIL	11.43	12.17	5.5	2.5	1.7		.015		.017	
LOWER PREMOLAR 1	CHIMPANZEE	10.80	8.07	3.1	4.2						
	FOSSIL	10.52	9.38	2.6	8.7		7.5		.080		.038
LOWER PREMOLAR 2	CHIMPANZEE	8.24	9.04	3.0	3.8						
	FOSSIL	8.13	9.64	4.0	5.1	1.7	2.2	.021	.023	.017	.011
LOWER MOLAR 1	CHIMPANZEE	11.24	10.36	3.3	2.8						
	FOSSIL	11.26	10.25	2.4	3.4		1.0		.0098		.005
LOWER MOLAR 2	CHIMPANZEE	11.43	11.40	5.1	3.1						
	FOSSIL	12.62	11.77	5.1	6.0		4.8		.041		.024
LOWER MOLAR 3	CHIMPANZEE	10.73	10.06	3.2	3.3						
	FOSSIL	14.06	12.22	8.0	6.3	8.0	5.3	.057	.043	.08	.027

COMPARISON OF VARIABILITY in length and width of tooth crowns among living chimpanzees and among hominoid fossils from the late Miocene and early Pliocene strata in India reveals that of the 14 categories of upper and lower teeth represented, only nine fossil categories (*color*) display variations in size that exceed the variations found among living chimpanzees. Attaining such extreme measurements through the mechanism of genetic selection calls for a percent change in dimension in successive generations that amounts to only fractions of a thousandth of 1 percent. This is nearly an order of magnitude less than the change in dimension per generation observed with respect to reduced wing length in 100 consecutive generations of flies reared in reproductive isolation.

represented in the collection might well belong to a single phyletic line.

Calculations similar to those presented here have been done on the crown dimensions of fossil teeth believed to represent the European and African dryopithecines *D. fontani, D. laietanus, D. africanus, D. nyanzae* and *D. major,* and also such purported hominids as *Oreopithecus* and *Kenyapithecus.* The results in all cases were consistent with those obtained with the Indian fossil material. Neither among the European and African fossils nor among *Dryopithecus* and *Ramapithecus* specimens from the rest of Asia is the fossil range of variation in tooth size so much greater than the modern chimpanzee's range, when related to the time available for evolutionary change, as to force the assumption that more than one evolving line of dryopithecine hominoids existed in any of these areas before middle Pliocene times.

On the basis of these tooth-size calculations, at least, there would appear to be little evidence to suggest that several different hominoid species are represented among the Old World dryopithecine fossils of late Miocene and early Pliocene times. Neither is there compelling evidence for the existence of any distinct hominid species during this interval, unless the designation "hominid" means simply any individual ape that happens to have small teeth and a correspondingly small face. Fossil hominoids such as *Ramapithecus* may well be ancestral to the hominid line in the sense that they were individual members of an evolving phyletic line from which the hominids later diverged. They themselves nevertheless seem to have been apes—morphologically, ecologically and behaviorally. They were apes morphologically because the variation they display, when balanced against the length of time available for evolutionary change, allows the simple—and therefore logically preferable—conclusion that only a single phyletic line of great apes was evolving at that time. Ecologically and behaviorally they were apes because they lived in an era when no ecological imperative demanded an evolutionary change in the hominid direction. Climatic conditions in Miocene and early Pliocene times allowed tropical forest to flourish over vast areas of Eurasia and Africa. Such an environment favors apes but does not require the development of hominid behavior patterns.

It is reasonable to suppose that some ape populations of this period were nonetheless developing what may be called behavioral preadaptations to the hominid way of life, just as modern chimpanzees display rudimentary stages of toolmaking. Such behavioral patterns would not assume evolutionary importance, however, until the preadapted apes found it impossible to continue living the undemanding forest existence to which they were accustomed.

Such an ecological challenge appar-

© 1971 SCIENTIFIC AMERICAN, INC

ently took place during the later Pliocene. As a result of increasing aridity tropical forests began to shrink and give way to open grassland on a large scale. In parts of the Old World (portions of Africa and of Southeast Asia in particular) tropical forests remained, and so did apes. Where the forests vanished the apes did too. They must in most instances simply have perished. By making relatively rapid and radical behavioral adjustments, however, at least one preadapted ape population survived and by so doing came face to face with new selective pressures. These pressures in turn modified gene frequencies and so brought about the appearance of those morphological characteristics that exemplify the hominid line.

Once it is recognized that ecological and behavioral changes must have preceded morphological changes, it is possible to renew the quest for a specific ancestor of the hominid line from a different viewpoint. The search should be not for anatomical advance but for a population of hominoids, still predominantly apelike in appearance, that managed to survive under altered ecological conditions approximating those experienced by later hominids. Is there such a group?

There is, although its members have generally been excluded from any claim to ancestral position in the hominid line because of their great size. This is the population represented by fossil discoveries in both China and India (many teeth and a few jawbones) assigned to the genus *Gigantopithecus* [see "Gigantopithecus," by Elwyn L. Simons and Peter C. Ettel; SCIENTIFIC AMERICAN, January, 1970]. In some of its morphological characteristics this hominoid resembles later hominids to the degree that not only its discoverer, G. H. R. von Koenigswald, but also one of the principal students of *Gigantopithecus* teeth, the late Franz Weidenreich, have argued for its affinity with the hominid line. The animals may have been genuinely gigantic, but this is uncertain; estimates of their height when they stood upright range from five to 12 feet.

At the time of Weidenreich's studies the oldest generally recognized hominid fossils were the specimens called *Pithecanthropus erectus* (now usually referred to as *Homo erectus*) discovered in Java toward the end of the 19th century. It was Weidenreich's view that the hominid line had evolved by reduction in size, beginning with *Gigantopithecus*, which he believed had been twice as large as a gorilla, and progressing through forms of intermediate size, *Pithecanthropus* included, to *Homo sapiens*. Today, of course, room must be found in any such scheme for the undoubted hominid *Australopithecus*.

When the difference in size between the teeth of *Gigantopithecus* and *Australopithecus* is considered in the light of possible change over a period of time, it appears that even if no more than 100,000 years separated the two forms, the greatest change per generation necessary to unite the mean measurements is quite small. In either tooth-crown length or tooth-crown width the required change would be no more than one micron per generation. This could be accomplished by a selection differential of at most five microns, which is no great amount. Moreover, certain *Gigantopithecus* specimens show a surprising resemblance to some of the earliest-known *Australopithecus* fossils, although others are substantially different in appearance [*see illustration on opposite page*].

In this connection it is at least suggestive that one intermediate form placed by Weidenreich between *Gigantopithecus* and *Homo erectus* was the then enigmatic Javanese fossil known as *Meganthropus*, which quite probably represents an East Indian population of australopithecines. Comparisons such as these make it clear that interpretations of the hominoid fossil record must be based not on individual specimens but on populations. The interpretations must in addition take into consideration the time available for evolutionary changes that can bridge the morphological gaps between specimens.

The fact that changes in tooth dimensions over a period of time would allow *Gigantopithecus* to be placed in an ancestral relationship to *Australopithecus* does not, of course, prove that *Gigantopithecus* was a hominid. At the very least, however, it clears the way for consideration of such a possibility. For example, the fossil fauna found in association with *Gigantopithecus* both in India and in China suggests an open-savanna environment rather than a forest one. Perhaps this primate, like *Australopithecus*, was adapted to foraging for its food in open country. Going further, it is even possible to relate this kind of adaptive pattern to the development of tool use in foraging.

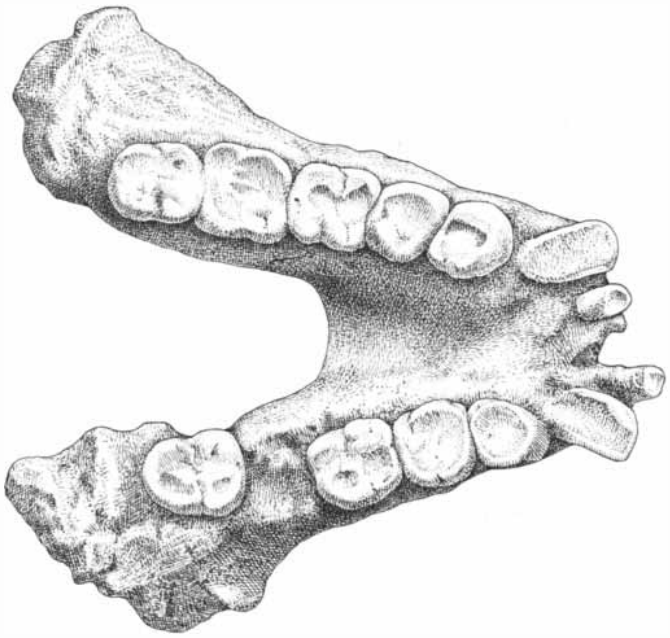
A large and presumably powerful primate such as *Gigantopithecus* could no doubt have easily killed many animals merely with a blow of its hand. The use of any hand-held object—stick, bone or stone—would have further enhanced this killing power. Simultaneously, the use of a weapon would have allowed a reduction in the user's size without any effective loss of the ability to kill. Further, a decrease in size would have meant an increase in speed and agility and, because of a corresponding reduction in the daily caloric demand, would also have been a favorable factor during any period of food scarcity.

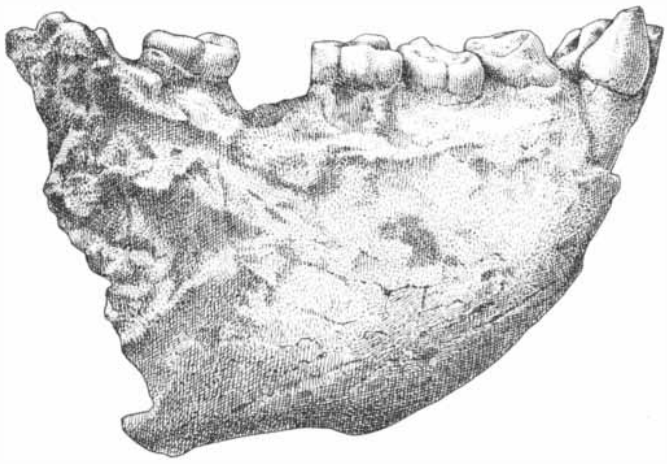
The mechanism suggested here need not be considered the only one that might have brought about such a transformation. Indeed, few now consider *Gigantopithecus* to be ancestral to the hominid line, in spite of Weidenreich's and von Koenigswald's earlier suggestion of such a possibility. But if *Gigantopithecus* should eventually be accorded ancestral status, at least one paradox in hominoid evolutionary history—the disappearance of an animal with numerous dental characteristics that soon thereafter reappear in early hominids—would be eliminated.

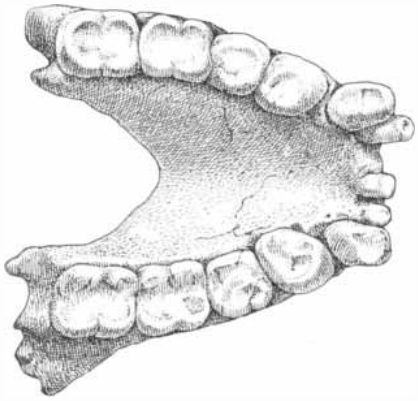
In Chesterton's story the puzzles of Glengyle Castle were solved when Father Brown realized that each piece of evidence was in fact only half a piece. In every instance what was missing was something made of gold: the settings for the gems, the boxes for the snuff, the watchcases that had held the wheels, gears and springs, and an entire collection of gold candlesticks. The explanation was the insane literalness of an old servant who had once been told that he was entitled to the Glengyle gold and, on the Earl's disappearance, had taken all of it but nothing else. With the fossil record too what is seen—the morphological similarities and dissimilarities between man and certain individual Miocene-Pliocene hominoid specimens—is only half of the evidence. When viewed together with the missing half, the evidence of population variation and change over a large span of time, the fossil record must eventually yield its true meaning.

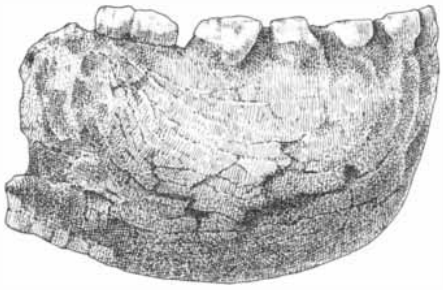
LATER RUNNER in the race for status as an ancestral hominid is *Gigantopithecus*, whose claim stems from the probability that changes in ecological conditions and behavior patterns preceded morphological change in the hominid direction. Few *Gigantopithecus* specimens fall within the hominid range of sizes; the jaw of an old male from China (*top on opposite page*), for example, is substantially more massive than a gorilla's. The much less massive jaw of a female specimen of *Gigantopithecus*, also from China (*middle*), however, appears to bear a close resemblance to one of the oldest-known specimens of *Australopithecus* (*bottom*), a jaw unearthed in the Omo Valley of Ethiopia.

© 1971 SCIENTIFIC AMERICAN, INC









© 1971 SCIENTIFIC AMERICAN, INC

MATHEMATICAL GAMES

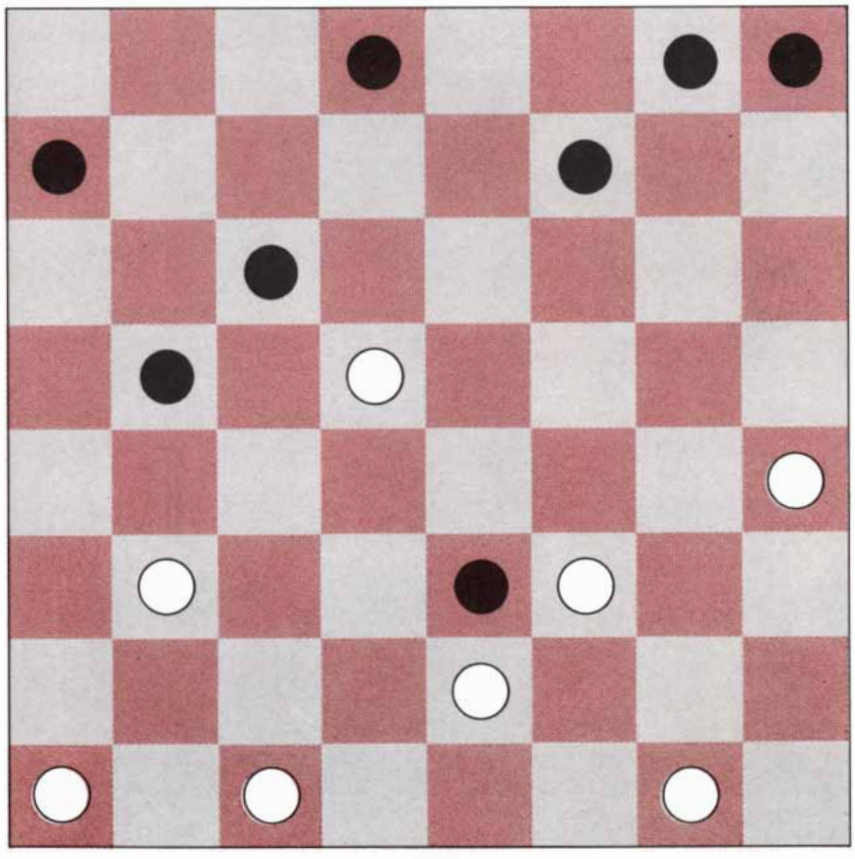
How to triumph at nim by playing safe, and John Horton Conway's game "Hackenbush"

by Martin Gardner

"The good humour is to steal. . . ."
—WILLIAM SHAKESPEARE,
Corporal Nym in
The Merry Wives of Windsor

During the past decade a great deal of significant theoretical work has been done on a type of two-person game that so far has no agreed-on name. Sometimes these games are called "nim-like games," "take-away games" or "disjunctive games." All begin with a finite set of elements that can be almost anything: counters, pebbles, empty cells of a board, lines on a graph, and so on. Players alternately remove a positive number of these elements in accordance with the game's rules. Since the elements diminish in number with each move, the game must eventually end. None of the moves is dictated by chance; there is "complete information" in that each player knows what his opponent does. Usually the last player to move wins.

The game must also be "impartial." This means that permissible moves depend solely on the pattern of elements prior to the move and not on who plays or on what the preceding moves were. A game in which each player has his own subset of the elements is not impartial. Chess, for example, is partial because a player is not allowed to move an opponent's piece. It follows from the above conditions that every pattern of elements is a certain win for either the first or the second player if the game is played rationally. A pattern is called "safe" (or some equivalent term) if the person who plays next is the loser and "unsafe" if the person who plays next is the winner. Every unsafe pattern can be made safe by at least one move, and every safe pattern becomes unsafe through *any* move. Otherwise it is easy to prove the contradictory result that both players could force a win. The winner's strategy is playing so that every unsafe position left by the loser becomes a safe one.

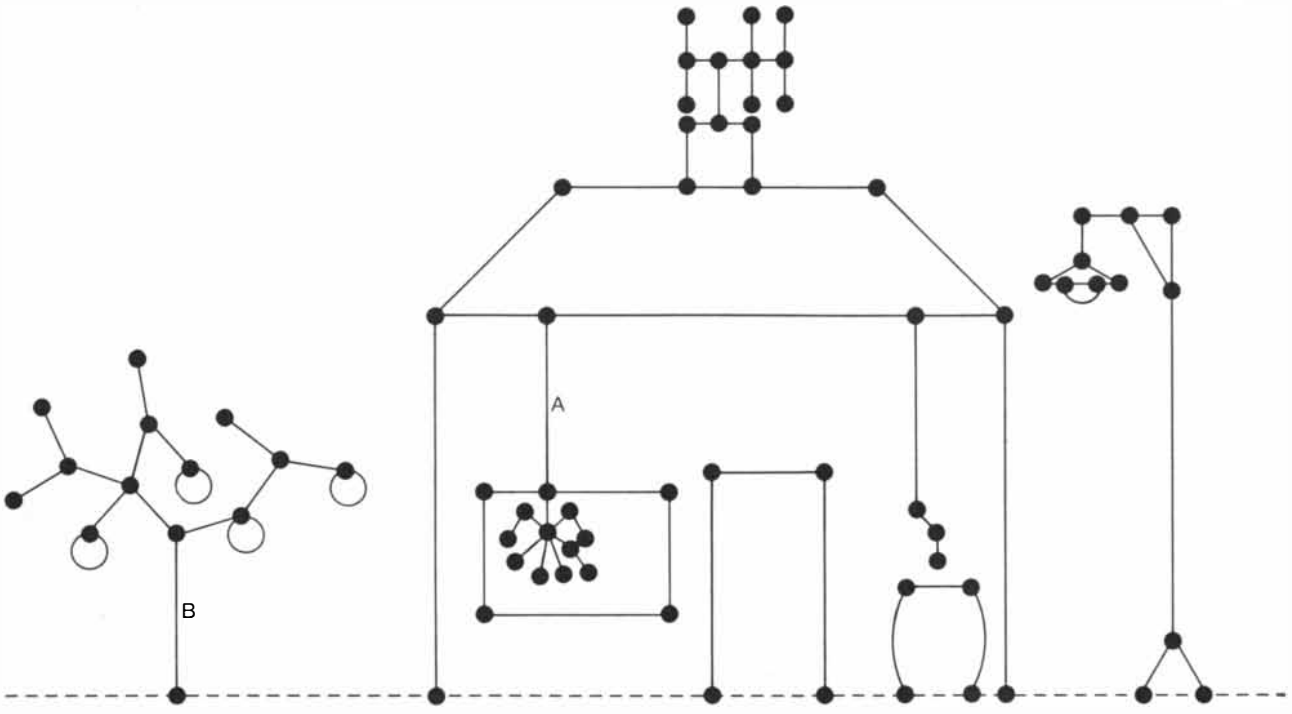


A nim game on a chessboard

The best-known example of such a game is nim. The word was coined by the Harvard mathematician Charles L. Bouton when he published the first analysis of the game in 1901. He did not explain why he chose the name, so that we can only guess at its origin. Did he have in mind the German *nimm* (the imperative of *nehmen*, "to take") or the archaic English "nim" ("take"), which became a slang word for "steal"? A letter to *The New Scientist* pointed out that John Gay's *Beggar's Opera* of 1727 speaks of a snuffbox "nimm'd by Filch," and that Shakespeare probably had "nim" in mind when he named one of Falstaff's thieving attendants Corporal Nym. Others have noticed that NIM becomes WIN when it is inverted.

Nim begins with any number of piles (or rows) of objects with an arbitrary number in each pile. A move consists in taking away as many objects as one wishes, but only from one pile. At least one object must be taken, and it is permissible to take the entire pile. The player who takes the last object wins. Bouton's method of determining whether a nim position is safe or unsafe is to express the pile numbers in binary notation, then add them without carrying. If and only if each column adds to an even number (zero is even) is the pattern safe. An equivalent but much easier way to identify the pattern (with practice one can do it in one's head) is to express each pile number as a sum of distinct powers of 2, eliminate all pairs of like powers and add the powers that remain. The final sum is the nim sum of the pattern. In current parlance this is called the "Grundy number" or "Sprague-Grundy" number of the pattern, after Roland Sprague and P. M. Grundy, who independently worked out a general theory of take-away games based on assigning (by techniques that vary with different games) single numbers to each state of the game.

© 1971 SCIENTIFIC AMERICAN, INC



The Hackenbush Homestead

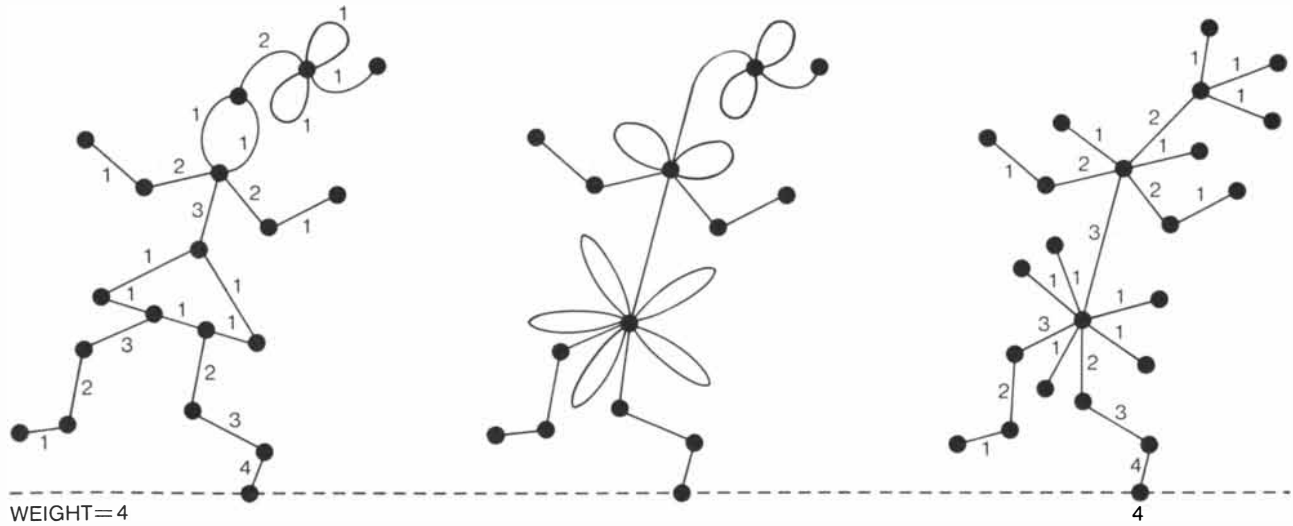
For example, assume that a game of nim begins with three piles of three, five and seven counters.

$$3 = \not{2} + \not{1}$$
$$5 = \not{4} + \not{1}$$
$$7 = \not{4} + \not{2} + 1$$

Pairs of 4's, 2's and 1's are crossed out as shown. The sum of what remains is 1. This is the nim sum of the pattern. If and only if the nim sum is zero is the pattern safe, otherwise it is unsafe (as it is here). If you play an unsafe pattern, you win by changing it to safe. Here removing one counter from any pile will lower the nim sum to zero. In three-pile nim, with no pile exceeding seven counters, the safe nim patterns are 0–*n*–*n*, where *n* in the first triplet is any digit from 1 through 7, and 1–2–3, 1–4–5, 1–6–7, 2–4–6, 2–5–7, 3–4–7, 3–5–6. If your opponent plays next, he is sure to leave a pattern with a nonzero nim sum that you can lower to zero again, thereby maintaining your winning strategy.

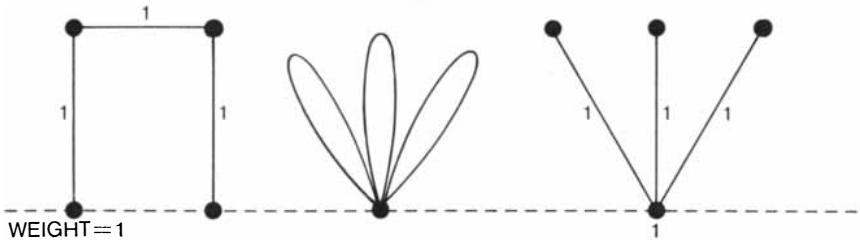
Like all games of this type, nim has its *misère* form, in which the player who takes the last piece is the loser. In many take-away games the strategy of *misère* play is enormously complicated, but in nim only a trivial modification is required at the end of the play. The winner need only play a normal strategy until it is possible to leave an odd number of single-counter piles. This forces his opponent to take the last counter.

Many take-away games seem to demand a strategy different from that of nim but actually do not. Suppose the rules of nim allow a player (if he wishes) to take from a pile, then divide the remaining counters of that pile into two separate piles. (If the counters are in rows, this is the same as taking contigu-



Girl on one foot

© 1971 SCIENTIFIC AMERICAN, INC



Door

ous counters from inside a row and regarding those that remain as being two distinct rows.) One might expect this maneuver to complicate the strategy, but it has no effect whatever. To win, compute the nim sum of a position in the usual way and, if it is unsafe, play a standard move to make it safe. For example, in the 3–5–7 game suppose your first move is taking a counter from the three-pile, leaving the safe 2–5–7. Your opponent removes two counters from the seven-pile and splits the remaining five counters into a two-pile and a three-pile. The pattern is now 2–5–2–3. Its nim sum is six, which you make safe by taking two from the five-pile.

A pleasant counter-moving game on a chessboard is shown in the illustration on page 104. No fewer than two columns may be used. In this example we use all eight columns. Black and white counters are placed on arbitrary squares in each column, black on one side, white on the other. (A randomizing device, such as a die, can be used for the placement.) Players sit on opposite sides and alternate moves. A move consists in advancing one of your counters any desired number of empty cells in its column. It may not leap its opposing counter, so that when two counters meet, neither may move again. The last player to move wins.

An astute reader may see at once that this game is no more than a thinly disguised nim. The "piles" are the empty cells between each pair of opposing counters. In the illustration the piles are 5–1–4–2–0–3–6–3, which has an unsafe nim sum of 4. The first player can win by moving the counter in column one, three or seven forward four spaces. If the game had begun with all the counters in each player's first row, the pattern would have been 6–6–6–6–6–6–6–6, a safe position because its nim sum is zero. The first player must lose. The second player groups the columns into four pairs, then duplicates each of his opponent's moves in the paired column, a strategy that ensures a zero nim sum after every move.

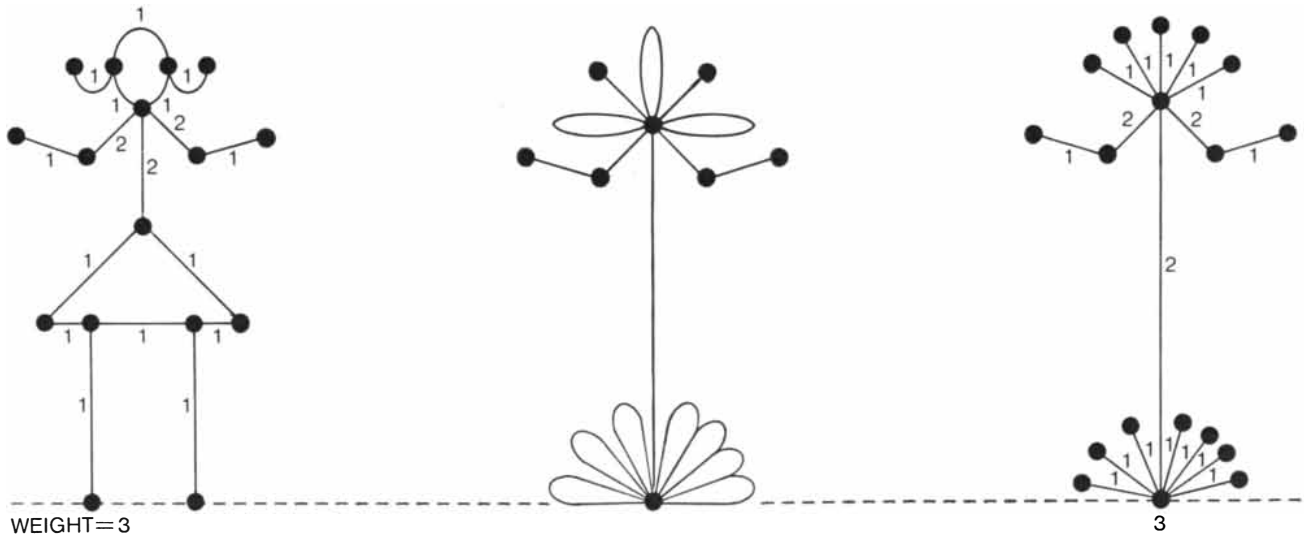
Suppose we complicate the rules by allowing either player to move backward as well as forward. Such a retreat is equivalent, of course, to *adding* counters to a nim pile. How does this affect the winning strategy?

A better-disguised game based on nim addition is a delightful pencil-and-paper game recently invented by John Horton Conway, the University of Cambridge mathematician frequently mentioned in this department in connection with his cellular-automaton game "life." Conway calls the new game Hackenbush, but it has also been called Graph and Chopper, Lizzie Borden's Nim and other names. The following description is taken from a chapter by Conway in "Hackenbush, Welter, and Prune," a discussion of several nimlike games that will appear in an exciting book-in-progress by Elwyn R. Berlekamp, Conway and Richard K. Guy.

The initial pattern is a set of disconnected graphs, such as the Hackenbush Homestead as drawn by Conway [*see top illustration on preceding page*]. An "edge" is any line joining two "nodes" (spots) or one node to itself. In the latter case the edge is a "loop" (for example each apple on the tree). Between two nodes there can be multiple edges (for example the light bulb). Every graph stands on a base line that is not part of the graph. Nodes on the base line, which is shown as a broken line in the illustrations, are called "base nodes."

Two players alternate in removing any single edge. Gravity now enters the game because taking an edge also removes any portion of the graph that is no longer connected to the base line. For instance, removing edge *A* eliminates both the spider and the window since both would fall to the ground, but removing the edge joining the spider to the window removes only the spider. Taking edge *B* chops down the entire apple tree. If one edge of the streetlight's base is taken, the structure still stands, but taking the second edge on a later move topples the entire structure. The person who takes the picture's last edge is the winner.

As in nim, every picture is either safe (second-player win) or unsafe (first-play-



Girl on both feet

© 1971 SCIENTIFIC AMERICAN, INC

er win), and the winner's strategy is to convert every unsafe pattern to safe. To evaluate a picture each graph must be assigned a number measuring the graph's "weight." To arrive at the assignment the first step is to collapse all the "cycles" (closed circuits of two or more edges) to loops, turning the graph into what Conway calls an apple tree, although in many cases the loops are best regarded as being flower petals. To see how it works, consider Conway's girl [*see figure at left in bottom illustration on page 105*]. She incorporates two cycles: her head and her skirt. First the two nodes of her head are brought together and then the two edges are bent into loops. Do the same with the five nodes and five edges of the skirt. The girl is now a flower girl [*middle figure*]. The next step is to change her to an ordinary tree by replacing each loop with a single branch [*figure at right*].

We now calculate this tree's weight. First, label 1 all edges with a terminal node (a node unconnected to another edge) or, to put it differently, all edges that, if removed, cause no other edges to fall off the tree. Label 2 all edges that support only one edge. Each remaining edge is labeled with one more than the nim sum of all the edges it immediately supports. Consider the edge corresponding to the girl's hair between her head and her hair ribbon. It immediately supports 1–1–1. A pair of 1's cancel, giving a nim sum of 1. Add 1 to the nim sum and this edge has a weight of 2. The edge that forms the body above the skirt immediately supports edges of values 2–1–2–1–2. The nim sum is 2. Add 1 and the edge has a weight of 3.

The girl's unraised thigh supports 3–1–1–3–1–1–1, a nim sum of 1, to which 1 is added to give the thigh a value of 2. The calf below it has a value of 3, the foot a value of 4. (In each case we simply add 1 to the value of the single, immediately supported edge.) Since the foot is the only support of the entire graph, the girl has a weight of 4. All edge values are now transferred to corresponding edges on the original girl.

With practice, edge values can be computed directly on the original graph, but it requires great care. For example, the girl's five skirt edges, raised thigh and body are all "immediately" supported by her unraised thigh. This is clear in the tree graph but is not so obvious in the original graph because many of the immediately supported edges are not close to the thigh.

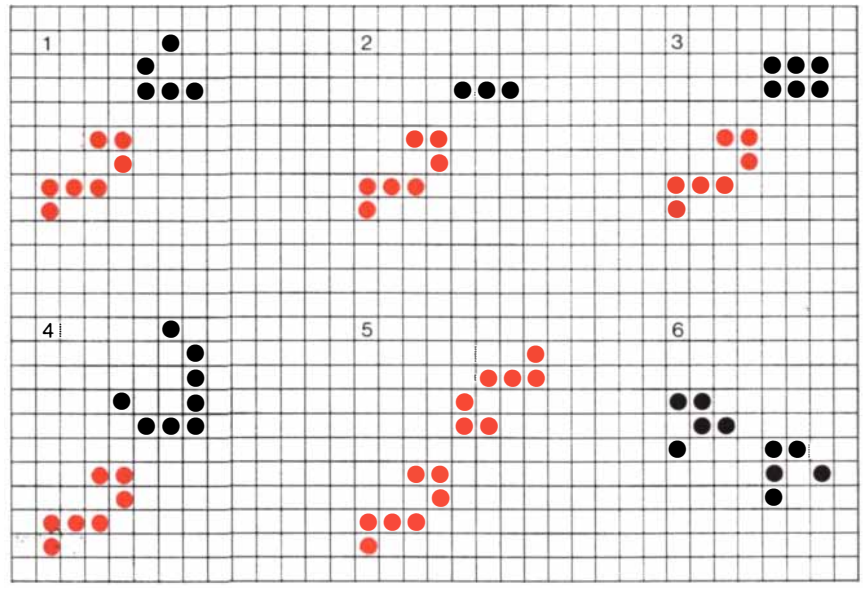
If a graph has more than one base node, such as the door, barrel and lamp in the Homestead, collapse the base cycle into loops, remembering that the broken line segment between a pair of base nodes is not part of the graph. The door's transformations are shown in the top illustration on the opposite page. Since the nim sum of 1–1–1 is 1, the door's weight is 1. A girl standing on both feet [*see bottom illustration on opposite page*] has a weight of 3. Note how the two cycles formed by her skirt and legs collapse into seven loops. A winning move, for a game played with her alone, is taking the top of her head or one of her hairs. This lowers the value of her head to zero, her body to 1 and her weight to zero. In this manner a weight can be assigned to each of the five graphs that make up the Hackenbush Homestead: the apple tree, house (including window, spider, chimney, television antenna and drainpipe), door, barrel and streetlight.

If Hackenbush is played with only the girl on one foot, the game is as trivial as playing nim with only one pile. The first player can win at once by taking the supporting foot. The poor girl collapses and he acquires all her edges. In the case of a figure with more than one base node, such as the door, we must remember to take an edge so that the remaining nim sum is zero. A first player can do this only by taking the door's top edge, leaving two graphs of weight 1 each, or a combined nim sum of zero. Taking either side leaves only one graph (of weight 2), which can be taken entirely by the second player.

A picture consisting of n graphs, such as the five graphs of the Homestead, is treated exactly like five piles in nim. The nim sum of all the weights is the total Grundy number. If and only if this number is zero is the picture safe and the second player assured of winning. As in nim, the winning strategy is to play so that the nim sum of what remains is always zero.

The reader is invited to determine the weight of each graph in the Hackenbush Homestead and verify that the Homestead's nim sum is 10. Since this is not zero, the first player can win. It turns out (of course Conway designed it that way) that there is only one edge the first player can take that will guarantee a win by lowering the nim sum to zero. Which edge is it? Conway's answer will be given next month along with the answer to the chessboard nim-game problem.

I have written only a brief, popular account of Hackenbush. Conway's chapter will go deep into the game's theory, proving a variety of unusual theorems that are too technical to discuss here.



The eater **(color)** ***and some of its prey***

A remarkable seven-bit stable figure in Conway's game "life" has been discovered by R. William Gosper, Jr., and his associates at the Massachusetts Institute of Technology. Called "the eater," it is capable of eating a large variety of "life" forms and then quickly repairing itself. The first four patterns in the illustration above show the eater about to ingest a glider, a blinker, a prebeehive and a lightweight spaceship. In the fifth pattern two eaters are poised to eat each other. This is prevented by their amazing self-repairing ability, creating what is probably the smallest oscillating pattern, with a rare period of 3. The last pattern shows how two gliders collide to produce an eater in 13 moves.

© 1971 SCIENTIFIC AMERICAN, INC

THE AMATEUR SCIENTIST

Groups are organized to observe the eclipse of stars by the moon

Conducted by C. L. Stong

If you should be driving along a back road one moonlit night, you might encounter the following scene. About every 500 feet for a mile or so you would see someone looking through a small telescope pointed at the moon. If you stopped to inquire, you would discover that the observers were timing the disappearance and reappearance of stars eclipsed by the moon as it moves across the sky. The people engaged in this task will inform you that they are observing lunar occultations. They are participating in an activity, combining the pleasures of observational astronomy with the opportunity of gathering useful scientific data, that is attracting the interest of a growing number of amateurs. One of them is Trudy E. Bell, formerly a history of science major at the University of California at Santa Cruz and now a member of the editorial staff of this magazine. She discusses her avocation as follows:

"Everyone knows that the moon and the stars rise in the east and set in the west, an apparent motion caused by the eastward rotation of the earth. Those who watch the night sky carefully also observe that the stars move westward faster than the moon does. A star that rises with the moon disappears below the western horizon several minutes before the moon does. In other words, the moon appears to drift slowly eastward through the pattern of fixed stars. The relative motion is not apparent at a glance because the moon drifts eastward at a rate of about one lunar diameter per hour.

"As the moon proceeds across the sky it occasionally encounters a fixed star that lies in its path. The star disappears behind the eastern edge of the moon and eventually emerges from the western edge. During the encounter the moon intercepts rays of light emitted by the star, creating a shadow that sweeps across the earth.

"If the star is the sun, the result is a spectacular eclipse. Twenty solar eclipses will be seen somewhere on the earth during the remainder of this century. Lunar shadows cast in the light of more distant stars sweep the earth somewhere every night, so that the phenomenon can be seen many times each year. These lunar occultations are less dramatic than solar eclipses because the shadows can rarely be detected by the unaided eye. Nonetheless, the eclipses can be eerily beautiful, particularly when the edge of the moon just grazes the star.

"To observe an occultation you look directly at a bright star that lies in the path of the moon. The star will wink out when it is covered by the advancing edge of the moon. It will reappear in less than two hours. Not all stars that are occulted can be seen by the unaided eye. Some are too dim. The occultation of many dim stars, however, can be observed with a small telescope.

"The shadow of the moon, as cast in the parallel light rays of a distant star, turns out to be quite interesting and useful even though it cannot be seen by the unaided eye. It is an exact silhouette of the moon, distorted only by the curvature of the earth. Its jagged edges reproduce the profile of the moon's mountains, valleys and plains. In effect the silhouette enables the observer to examine part of the moon's topography without making a voyage into space. To examine the profile of the lunar mountains, however, the observer must station himself at a strategic point on the earth.

"Assume that the shadow sweeps across the earth from west to east. All observers in its path may see a star wink out or another star suddenly pop from behind the moon. In either case the observer is seeing a total occultation.

"The few observers who are fortunate enough to be stationed at the northern and southern edges of the shadow, which is 2,000 miles wide, get a special treat. They see what is termed a grazing occultation. At these locations the jagged edge of the moon causes the star to blink off and on successively as lunar crags and valleys pass in review. An observer who is just outside the edge of the shadow sees a near-miss: one edge of the moon comes extremely close to the star but does not quite touch it. An observer who is a few miles away but within the shadow sees the star wink off and five minutes later reappear slightly west of the moon.

"Observers at intermediate points see the best show. Some catch the starlight at the bottom of the lunar valleys. Observers who are stationed progressively closer to the edge of the shadow see the lunar mountains at intermediate elevations. The observer who is barely inside the edge of the shadow sees a momentary blink as the summit of the highest lunar crag drifts across the ray of starlight. Observers only 50 or 100 feet apart may see different details on the lunar features, thereby obtaining a high accuracy in locating the features. The point at which the ray of starlight is tangent to the edge of the moon can also be identified. This point can be used for determining the position of the moon itself. During total occultations the observer cannot easily identify the exact points on the moon's edge at which the star disappears and reappears.

"Astronomers can mathematically predict the approximate position of the moon, in relation to the earth and the fixed stars, at any time in the future. They would like to be able to predict its future position exactly. Although such precision may be unattainable, several projects have been undertaken in recent years to reduce the uncertainty. For example, astronauts have placed reflectors on the moon so that a laser beam from the earth can be reflected back along the same path. This scheme yields an improved measure of the distance between the moon and the earth.

"In addition a program of observing grazing occultations on a worldwide basis has been established to serve as a check on the accuracy of the laser observations. The program may also substantially reduce present uncertainties

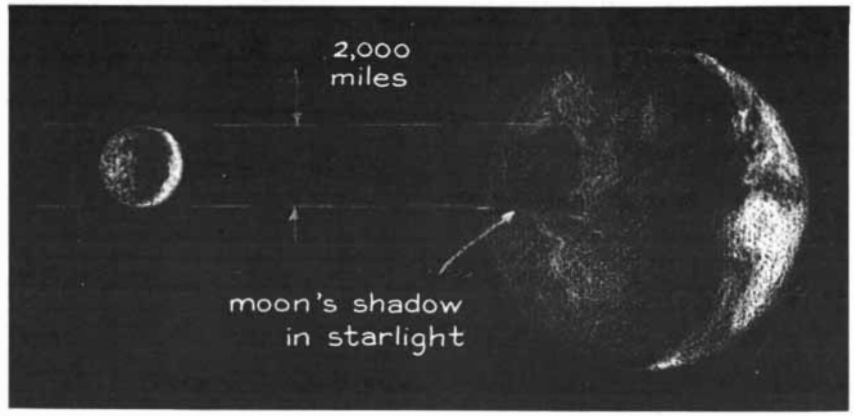
© 1971 SCIENTIFIC AMERICAN, INC

in the position of the moon with respect to the fixed stars. Responsibility for organizing and conducting this program has been accepted by Her Majesty's Nautical Almanac Office at Herstmonceux Castle in England. Astronomers there routinely predict as accurately as possible the paths where grazing occultations of certain stars can be seen. The office encourages teams of amateurs to use the predictions for making timed observations of grazes. The resulting observations are collected by the office for computing improved predictions, which amateurs can use for making more observations that lead to still better predictions and also to improved mathematical procedures for making predictions.

"The observation of grazing occultations can generate other useful information. For example, although the rough terrain at the edge of the moon has been mapped as carefully as possible from the earth, data collected by observing grazing occultations can establish accurate points of reference for the study of systematic errors in the charts and thus can lead to their improvement. Moreover, since grazing occultations are often visible over intercontinental distances, observations of them at widely separated locations can be used to establish the relations between independent surveying networks. In general, then, by knowing exactly where the moon is at any time one has a reference point for determining the relative location of other points in the universe, including points on the earth's surface.

"The technique of observing grazes can be easily mastered by laymen of all ages. The bulk of the work is currently being done by amateur astronomers. Professional astronomers record total occultations, but most of the observations of grazes are made by laymen because the requirements of the work are ideally suited to the resources of the amateur.

"Observers of grazing occultations work with fairly simple apparatus. Each observer should have a small telescope with an objective lens or a mirror at least two and a half inches in diameter. Many observers use six-inch reflecting telescopes that they make at home. The optimum size of the instrument, in terms of its light-gathering power, depends on the brightness of the star to be observed and the extent to which the moon is illuminated by the sun. A bright, fourth-magnitude star occulted on the dark side of a lunar crescent that is 40 percent illuminated can be observed successfully with a good pair of binoculars if they are rigidly mounted. At the other extreme an eight-inch reflecting telescope would be barely adequate for observing a star of the eighth magnitude if the star appeared near the cusp of a gibbous moon. Glare from the illuminated portion of the moon increases the difficulty of making observations. For this reason instruments of large focal ratio (*f*/8 or more) are preferred. A small refracting telescope is somewhat preferable to a larger reflector. Clean optics are important.

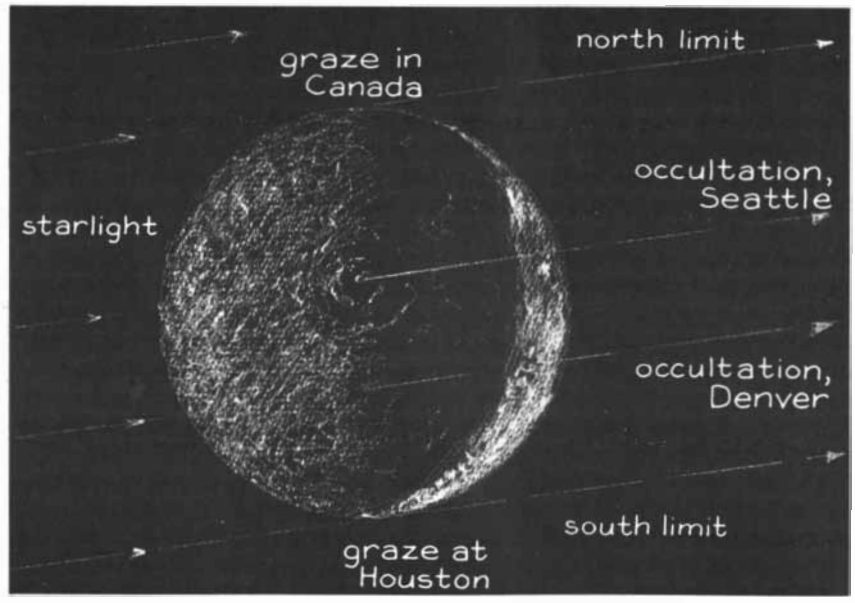


Shadow of the moon cast on the earth when a star is occulted

"A clock drive, which compensates for the rotation of the earth and keeps the telescope trained on a selected star, is convenient but not essential. If a clock drive is used, it must be of the type that operates from a battery or some other portable source of power. Many sites from which grazing occultations are visible lie along back roads where electric power is not conveniently available.

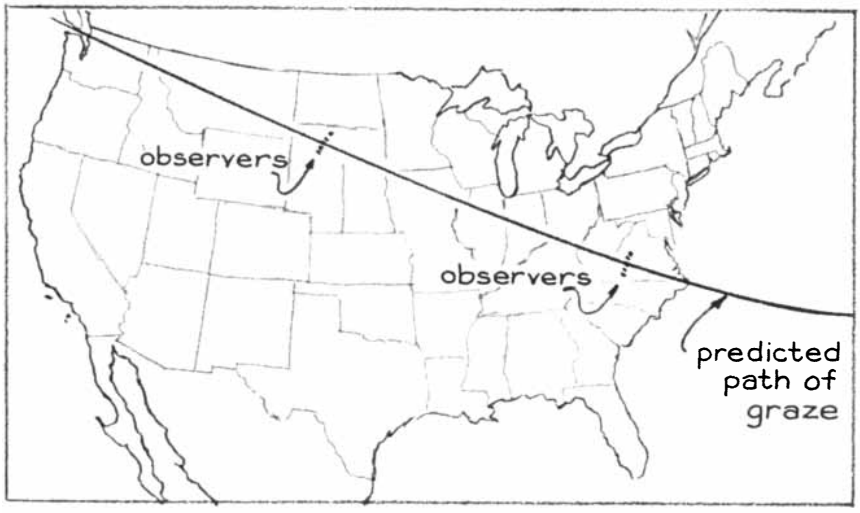
"The primary purpose of observing a grazing occultation is to obtain accurate timings of the momentary disappearances and reappearances of the star behind the lunar crags. There are two main methods of recording the events. One is for the observer to use a portable tape recorder and a shortwave radio receiver. The radio brings in accurate time signals broadcast either by station WWV (2.5, 5, 10, 15, 20 and 25 megacycles), which is operated by the National Bureau of Standards at Fort Collins, Colo., or by the equivalent Canadian station CHU (best heard on the East Coast at 3.33, 7.335 and 14.67 megacycles). These time signals are recorded on tape along with the observer's shouts of the events he sees through the telescope. Battery-powered equipment is easiest to set up and entails the least cost. Depending on the ingenuity and mechanical inclination of the individual, an observer can make valuable contributions to astronomy for less than $100.

"In the second method timings from



Relation of observer's position and type of occultation seen

© 1971 SCIENTIFIC AMERICAN, INC



Wide separation of two teams of observers

many observers are recorded on one tape with a long cable that connects all the stations. Each observer has a small oscillator that generates a tone of distinctive pitch. Sections of telephone wire typically 500 feet long are joined to form the cable, and an oscillator is connected to each of the junctions. When the observer sees the star disappear or reappear at the edge of the moon, he presses a button on the oscillator in a certain sequence. The inputs from all the oscillators are recorded on the tape at a central station, together with time signals from WWV or CHU. At least one group in the U.S. acquired a paper-chart recorder with multiple pens. Inputs from the various cable stations are thus recorded on paper instead of on magnetic tape.

"Observing teams operate under the direction of an expedition leader who is usually an experienced amateur. He is responsible for selecting the site for observing the graze; if a cable is used, he supervises the installation and testing of the apparatus. At the conclusion of the expedition he analyzes the data and forwards them to Her Majesty's Nautical Almanac Office.

"The leader receives information about future grazes from any one of half a dozen individuals, called computors, scattered over the U.S. These people generate the predictions with the aid of large electronic computers and special programs that take into account lunar theory, star positions and the motion of the moon to yield the predicted limit of a grazing occultation: the path traveled by the extreme northern or southern edge of the shadow of the moon cast by the occulted star. Depicted on a map of the earth, the predicted limit is an arc that can be several thousand miles long. Groups of observers set up their instruments at right angles to the arc at any convenient point along it [*see illustration above*]. Grazes can be observed within a mile or two of the limit. Any number of groups can observe the graze sequentially by setting up their equipment along the predicted limit.

"Such necessary information as the position of the limit, the brightness of the star and the date and time of graze is sent to the expedition leader on computer printout sheets; he then chooses the particular observing site for his group. He must be familiar with the geography of the area, since he is constrained by a number of factors. The ideal graze site should be not more than an hour's drive away, out along flat country roads with easily identified landmarks and few interfering lights. The leader relies on special large-scale maps published by the U.S. Geological Survey and the Canada Department of Energy, Mines and Resources to determine the best location.

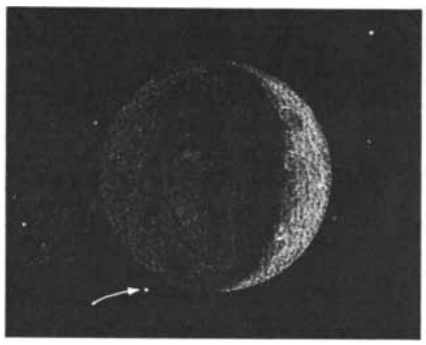
"After the expedition leader selects the site he must decide where to put the observers. What he is interested in is the profile of the moon's edge at the time of central graze and the place on the moon where the graze occurs. The shadow of the moon cast by the occulted star is a silhouette of the lunar limb. Fortunately there are contour charts of the moon's limb, drawn from photographs made between 1927 and 1956 by Chester B. Watts at the U.S. Naval Observatory in Washington and at two other observatories. With Watts's charts the expedition leader, or another person designated as a profile plotter, can draw the profile. Knowing the geography of the moon at the point of graze, the leader can tell which locations north and south of the predicted limit will yield the most information about the lunar limb.

"The expedition leader generally arrives at the graze site several hours before the observers do—often in daylight—and surveys the area to make sure that the station locations selected from the maps are satisfactory. If a cable is to be used, it is unrolled along the road. Usually each length of cable is stored on its own spool and distances are measured along the ground, so that when all the lengths are simultaneously unrolled they fit together evenly. In this way any number of people can assist in laying the entire cable. For an experienced team the operation may take less than an hour.

"The expedition leader identifies for the observers the star that will be occulted and helps beginners to center it in the field of view of their telescopes. Part of the computer data the leader receives tells where the star will be with respect to the limb of the moon at the time of graze. Even the crescent moon scatters enough light to render invisible all but some fairly bright stars (eighth magnitude and above). Only a few such stars will be in the field of view, and seldom is more than one in the proper location for the graze at the proper time.

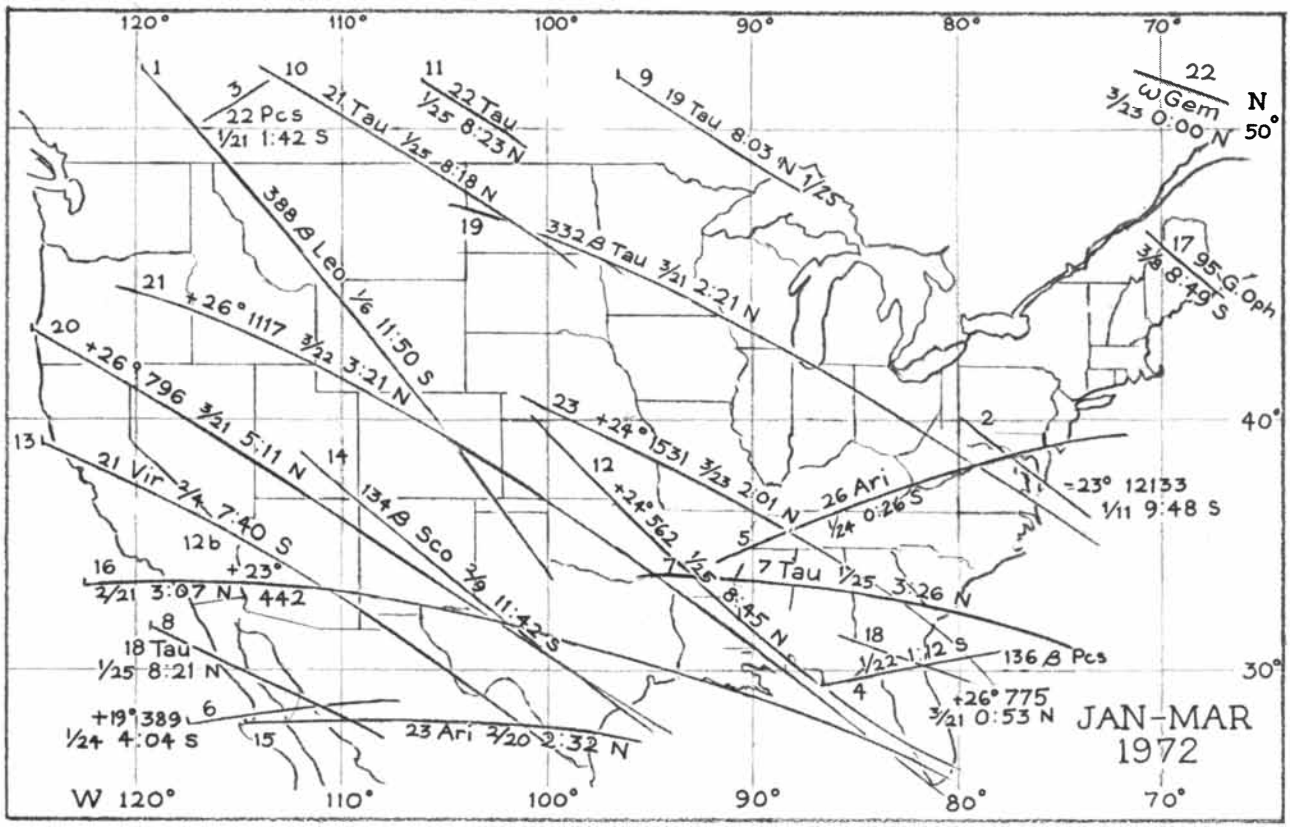
"Observers take note of the seeing conditions and how they affect the steadiness of the image of the star, so that during the graze a temporary dimming will not be mistaken for a disappearance. The seeing conditions are recorded on paper or tape. During this preliminary period observers also select the magnification that yields the clearest image of the star. Optimum magnification usually ranges between 90 and 150 diameters.

"As the moment of the graze approaches, all observers keep their eye glued to the eyepiece of their telescope, and anyone at a station who is not observing should maintain strict silence. If tape recorders are used at individual stations, each observer calls out a clear, sharp 'D' into the microphone when the star disappears behind the moon's edge and a sharp 'R' when it reappears. For short disappearances (those on the order



Star (arrow) *before grazing occultation*

© 1971 SCIENTIFIC AMERICAN, INC



Paths of grazing occultations visible in the U.S. through March of this year

of a tenth of a second) it is customary to call 'Blink.' A similarly short reappearance is recorded by calling 'Flash.' When oscillators are being used, the button is pressed once to signal a disappearance and twice in rapid succession to signal a reappearance. Some groups prefer to signal the events with the toy noisemaker known as a cricket, which emits a click when a metal tab is depressed. The cricket has the advantage of emitting a sharp sound that is easier to time than a fuzzy-edged voice syllable.

"Following the graze the group breaks camp. The event can occur at any hour of the night. It may be followed by coffee, a night's sleep or breakfast. Tapes or data that have been recorded by hand must be analyzed; the useful information from them is transcribed on report forms that are available from the expedition leader. If individual recorders are used at each station and are of the type having a capstan to drive the tape, the chances are good that the sounds can be reproduced and successfully timed with a stopwatch. Then the analysis can be made after the observer has returned home. All pertinent data are sent to the leader as soon as possible, but never later than 10 days after the expedition.

"Having turned in his observations to the expedition leader, the observer can relax until the date of the next graze. He is spared the difficulty of employing his data to improve the body of mathematics known as lunar theory. This work is performed at Her Majesty's Nautical Almanac Office with the aid of computers.

"A number of astronomical societies around the country have established grazing-occultation programs and welcome amateurs who would like to participate in the expeditions. An active group may go 'grazing' on an average of once a month. The accompanying map [*above*] shows some of the brighter grazes that will be visible in the U.S. between January and March of this year. They are numbered in order of their occurrence. Each predicted limit path starts in the west at the date and Greenwich Mean Time given beside the path. G.M.T. is used in astronomy to give a single time for an event. Subtract five hours for Eastern Standard Time. *N* and *S* indicate whether the graze is on the northern or the southern limb of the moon. The other numbers and letters are the names of the stars.

"Those who are interested in joining a group can get a foretaste of the fun by observing a few total occultations. These events can be observed anywhere, even from a rooftop or a backyard. Whereas grazing occultations can be seen only along a narrow strip a few miles wide, total occultations appear everywhere within the 2,000-mile width of the shadow. Train your telescope on the moon for a few evenings. Carefully examine the star pattern at the eastern edge of the moon. Eventually you will see one or more stars directly in the lunar path and within a diameter or so of the approaching disk. You can estimate when the moon will overtake the star by bearing in mind that the moon drifts a distance equal to its diameter in about an hour. By the same technique you can estimate when the occulted star will reappear, taking into account the width of the lunar disk at the point where the star disappeared.

"Beginners can also get a feel for occultation observing, without the responsibility of actually timing the events, by accompanying an expedition and just watching a graze. It is a beautiful spectacle and one of the few astronomical phenomena that can be viewed from start to finish within a reasonable period of time. Those who decide to plunge in can get more information about this avocation, together with the addresses of nearby occultation observers, from the man who started it all: David W. Dunham, Apartment C, 4771 South Spring Street, St. Louis, Mo. 63116."

© 1971 SCIENTIFIC AMERICAN, INC

WHY MISS OUT?

The important books you want to read and own are yours... in

The Library of Science

How many times have you passed up the books you wanted—only to discover they're impossible to find when you decided to get them? Let The Library of Science bring you the important books you want for reading or reference NOW—books that cover *every* scientific discipline, and all at worthwhile savings. Bonus books, too—*any* book on our list becomes a bonus book when you've bought four selections. Why miss out?

INTRODUCTORY OFFER:

Take any 3 books all only $3.95 (values to $47.95)

if you will join now and agree to take 3 more selections or alternates in the next twelve months

(retail prices shown)

79940. THE SPECTRA AND STRUCTURES OF SIMPLE FREE RADICALS. *Gerhard Herzberg.* Spectroscopy is the most powerful tool for studying these chemical will-o'-the-wisps and this new volume by the master is an important addition to the literature. **$11.00**

80640. THE STORY OF QUANTUM MECHANICS. *Victor Guillemin.* "It is three books rolled into one: the early story of quantum theory, the principal accomplishments of that theory, and the philosophical implications thereof."—*Eugene P. Wigner, Nobel Laureate in Physics.* **$8.95**

64050. THE NERVOUS SYSTEM. *Dr. Peter Nathan.* Neuroanatomy, neurophysiology and behavioral psychology are welded together in this engrossing presentation of nature's most complex wonders—the brain and spinal chord. **$7.95**

81800. SYMMETRY, ORBITALS, AND SPECTRA. *Milton Orchin and H. H. Jaffé.* The unity of modern chemistry stems directly from the integrating concepts of symmetry, molecular orbital theory, and absorption spectroscopy. Here, two renowned chemists blend the three into a total approach to chemical theory. **$16.50**

63860. THE NATURE AND GROWTH OF MODERN MATHEMATICS. *Edna E. Kramer.* Ingenious new approach to the history of developing mathematical concepts. Lucid, brilliantly organized, even dramatic, this is a real find. *Counts as 2 of your 3 books.* **$24.95**

56320. INTRODUCTION TO CALCULUS AND ANALYSIS. *Richard Courant and Fritz John.* An up-to-date introduction, stressing the interaction between mathematical analysis and its applications, as well as the reciprocity between differential and integral calculus. **$11.95**

66020. OPTICS. *Miles V. Klein.* Covers not only traditional topics in geometrical and physical optics but provides a complete introduction to the full range of modern developments including lasers, holography, and coherence. The best book in its field. **$14.95**

44350. ELEMENTS OF CARTOGRAPHY: Third Edition. *Arthur H. Robinson and Randall D. Sale.* Includes a wealth of new material, especially techniques of air-mapping, and the most recent experiments in design and reproduction. **$10.95**

63910. THE NATURE AND FUNCTION OF SCIENTIFIC THEORIES. *Edited by Robert G. Colodny.* Six scholars analyze the contemporary status of scientific theories. Historical antecedents as well as linguistic, logical, epistemic, and ontological elements of theory-building are discussed. **$12.95**

77680. SECRETS OF THE GREAT PYRAMID. *Peter Tompkins.* The history and mysteries of mankind's most monumental structure—the Great Pyramid of Cheops. Is it nothing more than a colossal tomb? Or does much of the ancient world's arcane knowledge lie hidden in its intricacies? **$12.50**

56260. INTRODUCTION TO MATHEMATICAL STATISTICS: Fourth Edition. *Paul G. Hoel.* Extensive rewriting of a great classic. Mathematically rigorous yet filled with concrete problems and practical applications. **$11.50**

42150. CHAMBERS' DICTIONARY OF ELECTRONICS AND NUCLEONICS. *Edited by L. E. C. Hughes, R. W. B. Stephens and L. D. Brown.* With more than 9,000 individual entries, a comprehensive list of abbreviations, acronyms and symbols and over 100 pages of data tables. **$14.50**

36290. BIOPHILOSOPHY. *Bernhard Rensch.* Recent revolutionary discoveries in biology form the basis for an unprecedented natural philosophy of life constructed by the brilliant German biologist-philosopher. **$12.50**

37400. CHALLENGE FOR SURVIVAL: Land, Air and Water for Man in Megalopolis. *Edited by Pierre Dansereau.* Distinguished experts from every field of conservation examine the urban environment and the potential for disaster represented by the twin menaces of overpopulation and unbridled technology. **$7.95**

37560. CHEMICAL APPLICATIONS OF GROUP THEORY: Second Edition. *F. Albert Cotton.* Brilliant revision of a trail-blazing work to include the latest developments in the field. Bridges mathematics and chemistry to provide a coordinated approach to an important analytical tool. **$12.00**

65210. OCEANOGRAPHY: An Introduction to the Marine Environment. *Peter K. Weyl.* Brings light to the mysteries of the ocean—the geology of the ocean basins, the chemistry of sea water, life beneath the sea and fundamental energy processes. Copiously illustrated. **$12.50**

34390. ASIA, EAST BY SOUTH: Second Edition. *Joseph Earle Spencer and William L. Thomas.* The background to today's crises becomes immediately understandable in this revised edition of an unusual geography of the Orient. **$16.50**

70070. PRINCIPLES OF HOLOGRAPHY. *Howard M. Smith.* The first unified and complete treatment of this newest of sciences, holography—lenseless, three-dimensional photography. Covers the history, theory, practice, and applications. **$9.95**

46580. THE EVOLUTION OF MAN AND SOCIETY. *C. D. Darlington.* One of the world's great scientists sums up a lifetime of research and speculation on every aspect of human development. Like Darwin's *The Descent of Man,* it will stand as a landmark in the history of modern thought. **$12.95**

50110. GENETICS OF THE EVOLUTIONARY PROCESS. *Theodosius Dobzhansky.* The definitive work on the modern synthetic theory of evolution by the world's foremost geneticist. **$10.95**

45990. ESSAYS ON CELLULAR AUTOMATA. *Edited by Arthur W. Burks.* Eight experts in the field present a fascinating overview of one of the fastest-growing and most intriguing branches of computer science. **$12.50**

59840. MAN-MACHINE COMMUNICATION. *Charles T. Meadow.* The best general introduction to the subject of man-computer conversation and the ways of getting the most out of both machine and human. **$13.95**

48790. FOSSILS IN AMERICA. *Jay Ellis Ransom.* Convenient guide for the amateur fossil collector. Contains a complete list of fossil sites in the United States plus handy tips, basic working equipment, and an ingeniously simplified fossil identification and classification system. **$8.95**

53870. THE HUDSON RIVER: A Natural and Unnatural History. *Robert H. Boyle.* An eye-opening book and an impassioned call to action that could be about any river—or every river—in the United States. **$7.95**

MEMBERSHIP APPLICATION

2-586

The Library of Science, Riverside, N.J. 08075

Upon acceptance of this order, please enroll me as a member and send the three books I have indicated. Bill me only $3.95 for all three, plus postage and handling. If not delighted, I will return all books within ten days and this membership will be cancelled.

As a member, I need accept only three more selections during the next 12 months at reduced member prices, plus postage and handling. Savings range up to 30% and occasionally even more.

I understand that I will receive free advance Reviews which fully describe each month's Main Selection and Alternates. If I wish the Main Selection, I need do nothing; it will come to me automatically. If I do not wish to receive the Main Selection, or I wish an Alternate Selection, I will advise you by using the convenient reply card provided with the monthly Reviews. I understand that I may choose a bonus book for every four selections purchased. (Introductory offer counts as first selection toward a bonus book.)

Send no money. Members are billed when the books have been received.

(Offer good in Continental U.S. and Canada only.)

3 Books for $3.95. (write in numbers) Some expensive books count as 2 or 3 selections

_____ _____ _____

Name_____

Address_____

City_____ State_____ Zip_____

© 1971 SCIENTIFIC AMERICAN, INC

BOOKS

Eighty years before Columbus a Chinese armada with 37,000 men reached Africa

by N. Sivin

SCIENCE AND CIVILISATION IN CHINA. VOLUME 4: PHYSICS AND PHYSICAL TECHNOLOGY; PART III: CIVIL ENGINEERING AND NAUTICS, by Joseph Needham, with the collaboration of Wang Ling and Lu Gwei-Djen. Cambridge University Press ($55).

In a series of seven expeditions beginning in the year 1405 a great navy under the command of the "Eunuch of the Three Jewels," Chêng Ho, carried the glory of the Chinese emperor throughout the seas to the south and west. By 1417 the fleet had reached the Kenyan town of Malindi, where "some few decades later Vasco da Gama would find his Arab pilot to take his ships across to Calicut, thus opening Asian seas to European penetration." By 1433 it had cast anchor off Mecca. Chêng's armada included, among other ships, 62 nine-masted galleons 450 feet from bow to stern and more than 180 feet across the beam, and it had a company of 37,000 men. The vessels were equipped not only with multiple masts and fore-and-aft sails (which allow beating efficiently to windward) but also with true axially mounted rudders and strong bulkhead-built hulls that divided naturally into watertight compartments. All these developments were known in China by the third century. They did not come together in Europe to make possible large ships capable of sailing near the wind and surviving the rigors of the Atlantic and less familiar oceans until the 15th century. In fact, the watertight compartments and their corollary, pedal-operated bilge pumps (on the principle of the noria used in ancient China to raise water into irrigation channels), did not appear at the western end of Eurasia until the 18th century. At that time their Cathayan origin was no secret.

This free and pacific circulation between continents was nothing new for the Chinese. Chêng Ho was not only a Muslim but also the son of a *hajji*. We do not know whether Chêng's father had also made his pilgrimage to Mecca by sea. It is at least likely that he took the overland Silk Road, leaving paved highway in Kansu to follow the caravan tracks of Central Asia. If we recall that that had been Marco Polo's route to China, and that a millennium earlier still the silk of the Seres traded over this route was draining away the gold of Rome, we are moved once again by the steadiness of human effort in China, clearing and maintaining the innumerable channels of communication, internal and external, that kept the complex social organism alive and open to change over the centuries.

Long after Chêng Ho's time the roads and canals continued to be built, the bridges improved, the irrigation systems expanded. Chêng's argosies, however, were a final blaze of splendor before the extinction of the large and intrepid navy that had been founded 300 years earlier. The political decisions that killed it were part of a decisive turning inward of the civilization. Vasco da Gama, on the other hand, marks practically the beginning of European maritime expansion. By the time Chêng's fleet had left Africa for the last time the Portuguese had hardly begun to explore the west coast of the continent. A generation later they had rounded the Cape of Good Hope. We can picture a meeting at Malindi in our minds readily enough, but it never took place.

Chêng Ho and his navy stand out in relief against the rich panorama of this latest installment of *Science and Civilisation in China* because they exemplify the two questions that practically everything Joseph Needham has written on China is meant to begin answering: How was China able to attain an early scientific and technological superiority over the West, so that until the Renaissance it had more to give than to receive? And why, in spite of this early lead, did that improbable sequence of events, the scientific revolution, happen in Europe and not in China? The second question is bound to be asked by anyone as soon as he learns that the Chinese did serious science and did it well. The first might be called Needham's discovery, since the richness of the Chinese technical tradition across the board became obvious only gradually in the course of his survey.

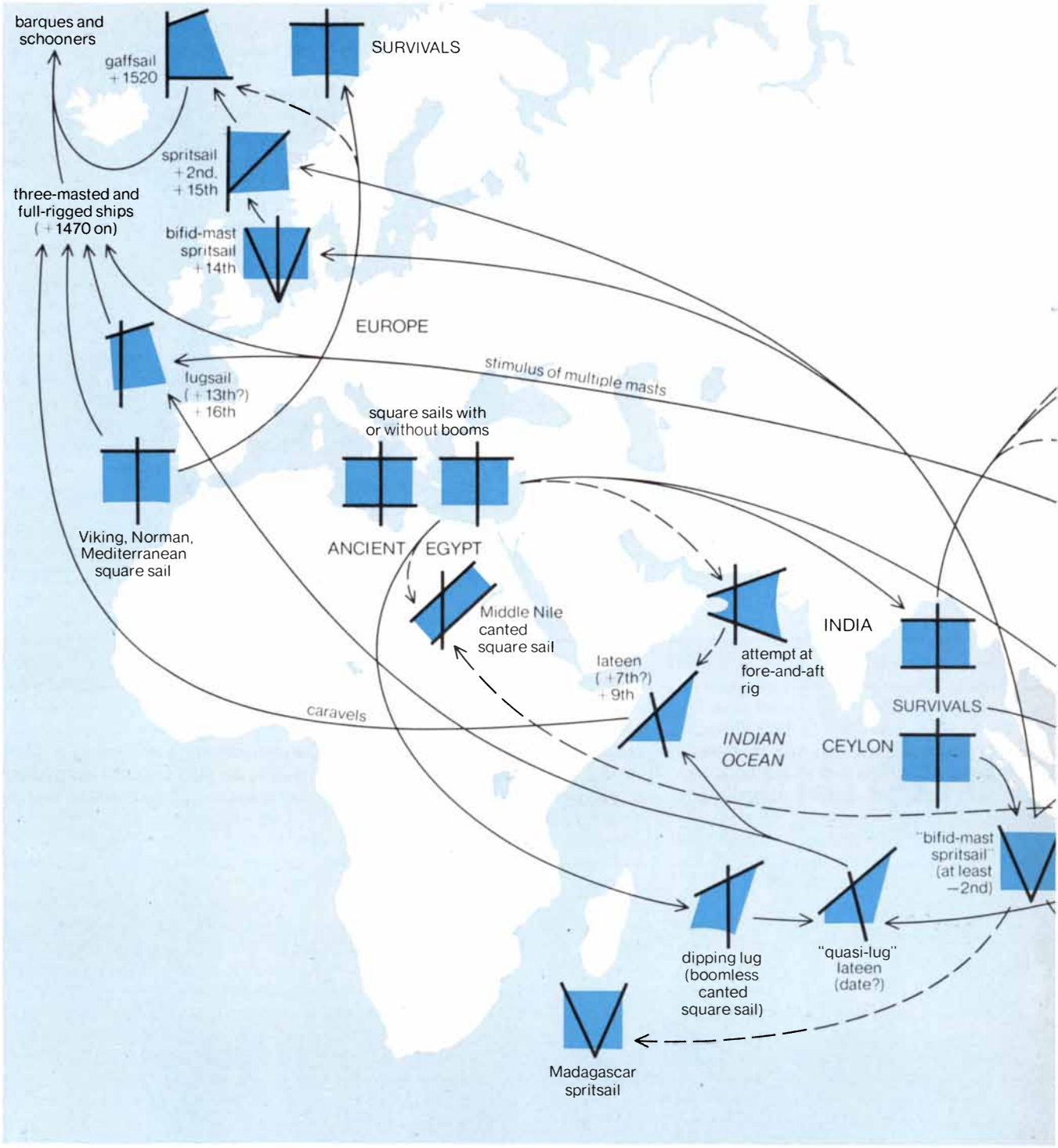
It will be a long time before the myth of a Mysterious East peopled by a race of philosopher-aesthetes dies away. The fact is that a Chinese visitor to England in the year 1400 would have had ample reason to consider it a technologically backward country. Still, not all Needham's colleagues would agree that there is any real significance in generalizations about the superiority of early Chinese *scientific* thought to that of Europe. In spite of interesting and valuable discoveries in every field of knowledge generated by man's attempt to comprehend nature objectively, the Chinese were weak in just those areas that became most crucial to the Galilean and Newtonian syntheses. The Mohists' impressive moves in the direction of logic in the fourth century B.C. found no successors, and so the idea of rigorous demonstration never evolved. There was no more application of exact measurement to planned experimentation in China than in Europe before Galileo, even though Chinese alchemists around the seventh century were working in that direction. An early Chinese strength in numerical and algebraic methods was paid for by a corresponding lack of development in geometry. Hence the ability to compute celestial positions and phenomena remained seriously limited, and responsibility for astronomical predictions was repeatedly delegated by the Chinese court to foreign mathematicians: Indian in the seventh or eighth century, Islamic in the 13th and Jesuit in the 17th.

One addresses this problem not by asking how national character (whatever that may be) dictated mutually exclusive approaches but by seeking to identify style-defining choices growing out of a culture's exploration of the full range of possibilities in scientific thought. If certain limitations are the unpredictable consequences of alternatives chosen ear-

© 1971 SCIENTIFIC AMERICAN, INC

lier, there are others, as Needham emphasizes, no less important that reflect the continuous influence of social values and priorities.

So much for science. Needham's case for the exceptionally high level of Chinese technology, in spite of the tentativeness with which many supporting details have had to be presented in his exploratory survey, needs no such qualification. The innovations of Chinese artisans were no more dependent on the high tradition of theoretical science than those of their European counterparts before modern times (although there is reason to believe inadvertent alchemical explosions led to the discovery of gunpowder). The useful arts of Western classical antiquity and of China appear on the whole to have been quite commensurable, with many clues from both pointing back to a common archaic center of diffusion in Mesopotamia and perhaps another in Egypt. There is no doubt, however, that while European civilization was falling apart the Chinese proceeded apace to one brilliant means

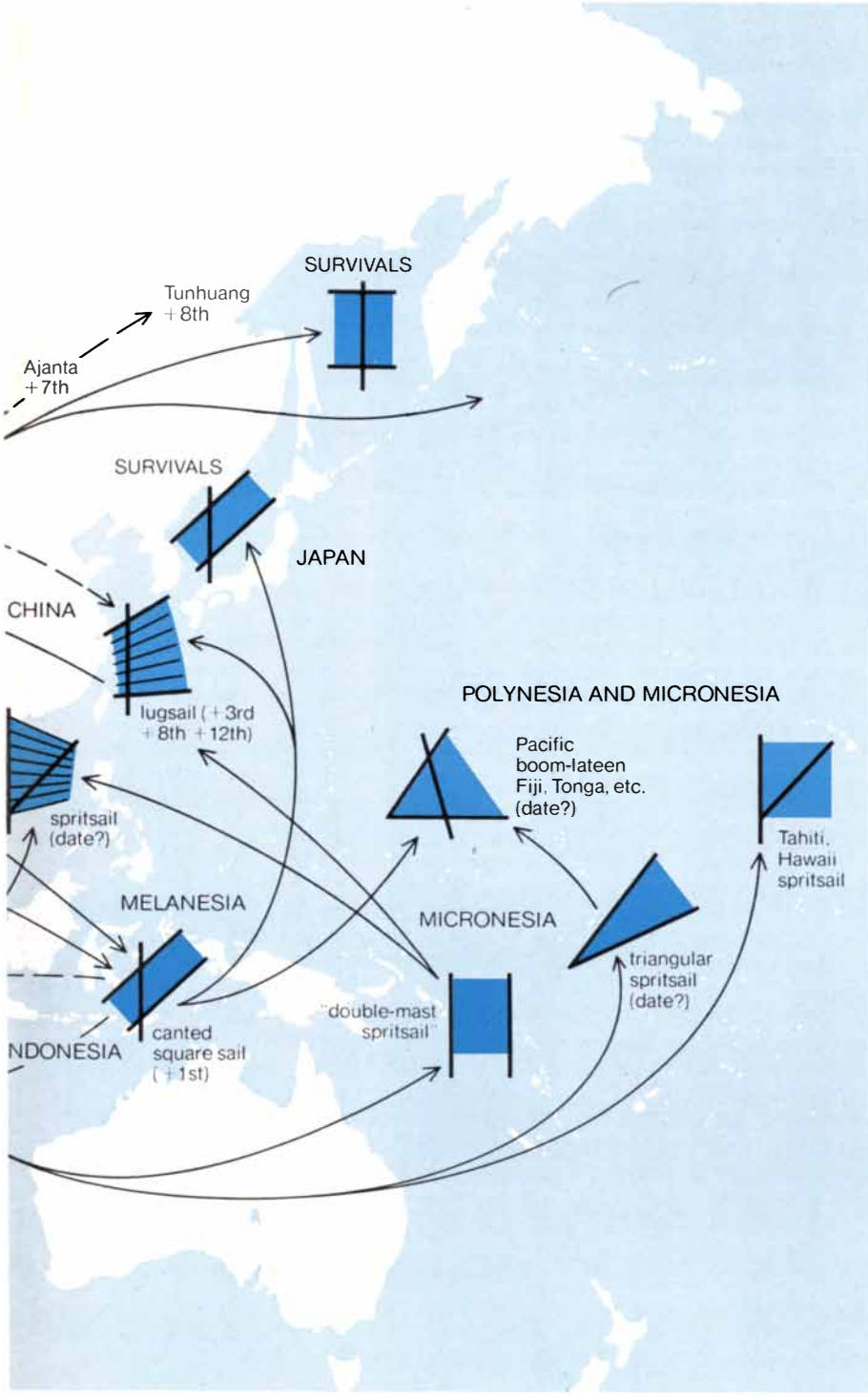


TYPES OF SAILS are superimposed on a map of the Old World in this illustration adapted from Volume 4, Part III, of *Science and Civilisation in China*. The sails themselves are in color; the heavy lines are masts, booms, sprits, gaffs and so on. The arrows

© 1971 SCIENTIFIC AMERICAN, INC

after another of magnifying human effort with machines and replacing it with the motive power of animals, wind and water. By the time Europe had recovered from the material and ideological regression of the Dark Ages, many of the problems of manufacturing, processing and construction that faced it had been solved half a world away.

Needham has long taken the position that, given the constancy of intercourse between East and West, inherent likelihood favors those who postulate transmission over those who argue for independent reinvention (particularly after an interval of centuries). This position is eliciting less skepticism as Needham gradually assembles what he calls clusters of inventions that appeared at about the same time in Europe long after they were well established in China. To the 12th-century cluster of the magnetic compass, paper, deep drilling and other inventions must be added the axial rudder and probably the segmental-arch bridge. (Iron-chain suspension bridges may well go back to A.D. 600 or earlier in China. Descriptions of them aroused much wonder in Europe in the 17th century, but apparently none was built there until well into the 18th.) The 14th-century and early 15th-century combination of printing and projectile artillery, which migrated at a time when the Pax Mongolica made possible relatively easy overland travel the length of Asia for Marco Polo and others like him, can now be expanded to include the pound-lock (which allows the easy uphill progress of boats) and the summit canal.



show the possible genetic derivations of the types of sails, with the broken arrows indicating the more speculative relations. No attempt has been made to show relative sizes of sails.

Needham's search for patterns of strength, precocity and priority—the "grand titration" of one culture against another—both unites the accumulating volumes of *Science and Civilisation in China* thematically and shapes their content. Volume 1 sketches the backdrop: geography, history and the movement of people and ideas in and out of China in early times. Volume 2 is a survey of philosophical ideas, with emphasis on their potential for inhibiting or furthering the development of modern approaches to science. The concepts on which the Chinese scientific tradition was based are analyzed with sophistication. A foundation is thus laid for succinct technical descriptions and characterizations later on, as when Needham remarks in the most recent volume that Confucian architecture was "immanentist, ethical, hierarchical, liturgical, axial, symmetrical." In Volume 3, devoted to mathematics and the sciences of sky and earth, we can see how much less than obvious were some of the consequences of the Chinese strength in numerical algorithms linked with weakness in spatial geometry. For example, in grid mapping the Chinese reached a high level early in spite of the lack not only of the great-circle theorem but also of projective geometry as a whole. There was nothing to inhibit the recurring notion of an infinite universe, and from the beginning we find in mathematical astronomy the use of equatorial coordinates, which were not adopted in Europe until the time of Tycho Brahe. Deep exploration of the methods of the exact sciences and the pressures, internal and external, that determined their character hardly lies within the purview of a work whose intent is explicitly pioneering and extensive. Nonetheless,

© 1971 SCIENTIFIC AMERICAN, INC

TEACHERS:
Bring the excitement of this magazine into your classrooms with
SCIENTIFIC AMERICAN Offprints

The unique intellectual stimulation you find each month in the pages of this magazine can be a part of your standard classroom fare if you regularly include SCIENTIFIC AMERICAN Offprints in your reading assignments. From a constantly growing inventory of more than 800 available articles—reproduced from the magazine with full text, full illustration, and full color—you can select those most suitable for the design of your course and use them as enrichment "chapters" along with your textbook.

Hundreds of thousands of students have already responded to SCIENTIFIC AMERICAN Offprints with fresh enthusiasm for their courses and with a better grasp of the subject.

Why not try a selection of Offprints in *your* next course? They are still only 25¢ each.

For a free descriptive brochure, write to Dept. 7-1

W. H. FREEMAN AND COMPANY

660 Market Street, San Francisco, California 94104
58 Kings Road, Reading RG1 3AA, Berkshire, England

the transparent organization, scrupulous documentation and lavish bibliographies (150 closely printed pages in Volume 4, Part III alone) will much simplify the intensive work of the next generation.

Part I of Volume 4 gathers from scattered sources and exhibits the connections of those Chinese ideas that would interest a modern physicist: knowledge and speculation in such areas as heat, light and sound. The story of magnetism and the compass is told in a way a scientist can take seriously, with an absorbing and high-spirited excursus on the origins of the compass in divination and its ties to the evolution of the world's many variants of chess from a lost cosmological game. Hardly less pleasurable is the section in Part II of Volume 4 on the prehistory of aeronautical engineering: a disquisition on the roots, growth and spread of the kite, and on the helicopter top, the high-speed sailing carriage, the parachute and the balloon. Part II is a thorough digest of what is known (including much that was never known before) about machines and power sources. The ampleness of the fourth volume gives scope for much more consistent depth than the first three volumes were able to provide.

Readers who want a preview of the volumes to come (Volume 5 on chemistry and chemical technology, including warfare, Volume 6 on biology and medicine and Volume 7 on the social relations of science) are well served for the moment by two recent collections of Needham articles and lectures. *Clerks and Craftsmen in China and the West* (1970) consists of 19 pieces ranging from general reflections to advance versions of chapters of *Science and Civilisation in China* to accounts of incidental researches too detailed or too late for inclusion in the series. Most of the subject matter overlaps with that of Volume 4 and Volume 6, and the collection includes some of the most important writing on Chinese medicine yet to appear in the Occident. Volume 5 is represented only by an article titled "Elixir Poisoning in Mediaeval China," but this is assuredly one of the more curious stories Needham has told. The articles in another collection, *The Grand Titration* (1970), have mostly to do with the relations of science and society as Needham has defined and redefined them over more than two decades. They are with some exceptions based on investigations less systematic than usual, and their arguments (if not their conclusions) tend to be much more tentative. Still, there is nothing to compare with them.

Part III of Volume 4, concerned with civil engineering and nautics, brings the series somewhere near the halfway point in length. There are occasional stretches, perhaps for the first time in the series, that only dedicated students will read thoroughly: detailed descriptions of the road network of nearly 2,000 years ago, of the geographical distribution of bridge types, of hydraulic projects. The maps, charts and tables that these accounts supplement are intelligently designed to lighten mental labor. Two large maps, tipped in next to each other, use colored overlays, one to show Chinese and Portuguese navigations in the 15th century and the other to exhibit principal currents and prevailing winds surrounding Africa and Asia. By reading them against each other (or holding them up to the light together) one can see at a glance how inevitable the early routes of exploration were. Alternating monsoons and a choice of currents running in opposite directions made the Chinese voyages to Africa practically straightforward. The Portuguese, on the other hand, were impeded south of Guinea by the northward Benguela Current and the trade winds blowing from the southeast. By Vasco da Gama's time navigators were circumventing these obstacles by sailing a great arc close to and past the snout of Brazil "until the Roaring Forties shot them through into the Indian Ocean." The implications for a pre-Columbian landfall in America could hardly be made more patent.

Most of the book will enthrall any curious reader, not only because of the inherent interest of its topics but also because individual facts are almost invariably bound in systematic relations. Often this is merely a matter of bringing out the standardizing working systems of the artisans themselves. Here is Li Chieh, author of the greatest treatise on architecture (printed in 1103), summing up the prevalent scheme of modules and modulors: "Thus the height and depth of the roof, the length, curvature and trueness of the members, and the ratios of column and post heights... together with the right use of square and compass, plumbline and ink-box—all proportion and rule depends on the system of standard timber dimensions and the standard divisions of these."

Most of the logical classifications, however, represent order imposed for the first time on large areas of human achievement. Needham's chart of the distribution and possible genetic derivation of types of sails [*see illustration on preceding two pages*], or his taxonomy of bridge structures before the age of steel and concrete, takes into account

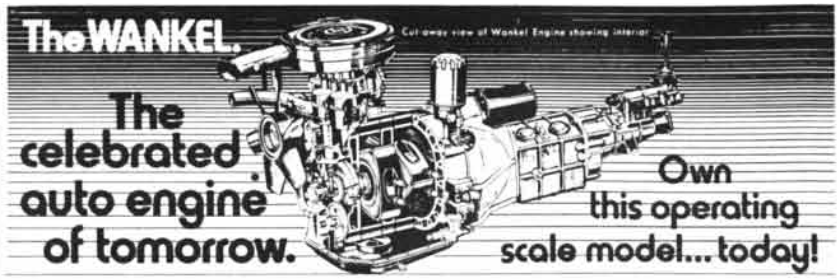
© 1971 SCIENTIFIC AMERICAN, INC

the entire world, not just China. Such catholicity and concern for analogies and comparisons are still rare enough among historians of science and technology that Needham's volumes are among the first one should read for a truly general view of most of the topics they touch on.

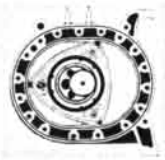
There is also something valuable in them that seems to be disappearing from serious historical writing as historians strive to remake their profession into a pseudo-science, namely a sense of nascent discovery. In Part III of Volume 4 we follow Needham's search for the moment when the axial rudder is first attested. This was a moment hard to detect because the rudder was called by the same name as its predecessor the steering oar, whose blade trailed in the water well behind the stern as the steersman swung its other end. Needham recognizes what no other sinologist had the knowledge to see earlier, that there is no ambiguity in the saying of T'an Ch'iao (about 940) that "the control of a ship carrying 10,000 bushels of freight is assured by means of a piece of wood no longer than one fathom" (about eight feet). A steering oar, even for a small riverboat, would have to be several times as long. Nor is there much doubt about the diary entries of the Japanese monk Ennin visiting China in 839. He writes, for instance, of his ship's being "blown on to rough rocks, and the rudder board broken off." Steering oars do not have anything that could be called a board.

The denouement came several years after Needham's textual argument, which puts the earliest document as far back as the third century, had first been sketched out. Beginning in 1954 several pottery ship models of the first and second centuries were excavated in Canton. The largest was "found to be equipped in very modern style" with a pivoted, slung and balanced rudder. No "proper nautical study" has yet appeared in China, and the only previously published photograph shows the rudder installed backward in the model, but Needham and his collaborator Lu Gwei-Djen were able to study the relics at first hand during two journeys to China. The sketches reproduced from a notebook of the author's 1958 visit are exact and telling, as are many original photographs among the book's 160 plates. There could hardly be a greater contrast with the uninformative snapshots and vague impressions of the great majority of visitors to China over the past couple of decades.

Historians of technology who have never studied the Chinese sources still



The first radical change in automobile power systems in half a century is here. And you can see exactly how it works, in this authentic operating scale model. Someday soon, your car will probably run on a Wankel. This is the engine you've been reading about in *The Wall Street Journal*, *Time* Magazine and *The New York Times*. Wankel-powered cars have been making headlines throughout the world. General Motors recently paid $50 million for the rights to use the revolutionary Wankel Rotary Engine in its cars. And such noted manufacturers as Rolls-Royce, Mercedes-Benz, Alfa Romeo and Citroen are already developing Wankel-powered automobiles.



Works Without Pistons.

What makes the Wankel run? Instead of conventional cylinders and pistons, it uses unique triangular rotors that rotate in combustion chambers shaped like fat figure eights. Why is the pistonless Wankel better? Because it's less than ½ the size and weight of conventional engines of the same horsepower! And has 40% fewer parts. No valves, no rods, no lifters, no cam shafts, no crank shafts. Which means less friction, less wear and tear, less need for repair. And the faster it goes, the quieter it gets.

Fully Authentic Scale Model.

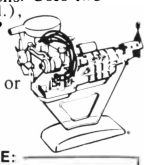
We find the whole thing incredibly fascinating. So will you, with your own motorized see-through model of this automobile engine of the future. We have a kit that duplicates the amazing Wankel in authentic 1/5th scale right down to the very last plug, gear and oil cap. This fully detailed operating model has a transparent, heavy plastic housing. You can actually see the triangular rotors turning on their shaft. You see the continual flash of igniting spark plugs. You see the rotating fan blades. You see the moving fan belt of real rubber. You hear the powerful deep-throated whir of the motor. All controlled by your own touch on the stick-shift on-off switch.

The model that separates the men from the toys.

This is *not* a plaything for preschoolers. It is an intricate, fully operational scale model of a major automotive development. A "must" for hobbyists, auto buffs and guys of all ages who love to build things. An impressive attention-getter for your office or den. A significant visual teaching aid. A history-making forerunner of things to come!

Order No. 43140: Wankel Rotary Engine© Kit.

Comes with all pre-painted parts, motor, display stand and easy-to-follow instructions. Uses two inexpensive AA batteries (not incl.), obtainable everywhere. The price? As amazing as the engine itself:



only $7.50 complete, including postage and handling. Send check or money order. (N.Y. residents add sales tax.)

FREE gift catalog on request.

THE *EVERYTHING* GUARANTEE:

The GALLERY guarantees EVERYTHING: The quality, accuracy of description, availability, prompt delivery. If not delighted, return the Wankel Engine Kit **within 3 weeks after you receive it** (not the usual 10 days) for an instant refund of purchase price.

finer things by mail from The GALLERY
Dept. 2962, Amsterdam, N.Y. 12010

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—and binders—prepacked for sale in pairs only.

*Index for entire year in December issue.**

FILES
Hold 6 issues in each.
Single copies easily accessible.
Price: $4.00 for each pair (U.S.A. only).
Address your order, enclosing check or money order, to: *Department 6F*

BINDERS
Hold 6 issues in each.
Copies open flat.
Price: $4.50 for each pair (U.S.A. only).
Address your order, enclosing check or money order, to: *Department 6B*

New York City residents please add 7% sales tax
Other NYS residents please add 4% state sales tax plus local tax

*Supply of back copies of SCIENTIFIC AMERICAN is limited. To replace missing copies, mail request with your order for files or binders.

SCIENTIFIC AMERICAN, 415 Madison Ave., New York, N.Y. 10017.

© 1971 SCIENTIFIC AMERICAN, INC

persist in offering ponderous explanations for what is supposed to be an inborn Chinese scientific and technological inferiority. We are assured by the distinguished author of a 1967 book on the technological background of European expansion into Asia that "the Chinese could grasp the usefulness of machines pertaining to irrigation, for example, but could not understand the purpose of other Western inventions," and that "it is obvious that the Chinese were not interested in weapons and military matters while the Japanese were." In view of this kind of nonsense it is hardly surprising that Needham should try consistently to make the strongest possible case in the other direction. My own impression is that occasionally he extends the benefit of the doubt where it is either not needed or not justified. The net effect is to lessen the credibility of what even with no indulgence at all would be an overwhelming case for the larger point being argued. I shall give an example with broad implications.

There is a clear contrast between the pacific intentions of the Chinese explorers, who were concerned essentially with establishing diplomatic and trade relations wherever they went, and the intolerance, destructiveness and insatiable greed of the adventurers from Lisbon. Nothing could be less Chinese than these words of João Ribiero in 1685: "From the Cape of Good Hope onwards we were unwilling to leave anything outside of our control; we were anxious to lay hands on everything.... There was not a corner which we did not occupy or desire to have subject to ourselves." This quotation is provided with a page-long footnote as crammed with hostile intentions toward China as the Pentagon papers: an assurance by the man who later became the first Portuguese ambassador to China that Canton could be depopulated by a single warship (1515), an expression of pleasure by a pious French Jesuit in 1698 at the thought of "how easily Lewis the Great would subdue those Provinces, if Nature had made us a little nearer Neighbours to China; he whom the stoutest Places in Europe can at best withstand but during a few days," an allusion to Lord Macartney's ideas about invasion projects in 1794, and so on. The antithesis could hardly have suffered had Needham noted that China in some periods was not averse to kingmaking and similar activities among her immediate neighbors, and showed considerable talent for military conquest during parts of the Han and T'ang eras. Traditional China not only lacked the convenient and automatic means of mass destruction commanded by the superpowers of today; it did not even have the long-range effectiveness of early European naval guns. That does not mean that the Vietnamese remember their prolonged medieval subjugation as having been benign. The significant ground of Needham's contrast lies far from the plane of immutable national character.

Although it may appear odd that a review of a book on Chinese technology has devoted so much space to matters concerning other parts of the globe, this merely reflects the broad conception of Needham's series. Where else can one find a serious critique of two translations of al-Idrīsī on geography, an 18th-century Swahili poem praising Seric porcelain, a discourse on hydraulic metaphors in medicine, an account of the discovery of a Chinese Taoist figurine in Australia, estimates of road mileage per unit of land area in the Roman Empire and in China at the same period (compiled in both cases from primary sources) and a detailed explication (with a map) of the "most complete and subtlest form" of irrigation works in the ancient world (built up in Ceylon between about 300 B.C. and the 12th century)? These are not mere ornaments. Each is woven into the substance of Needham's argument that modern science could not have come into being except out of the shared experience of all mankind. They help to make his book the first true world history of science and technology.

Shorter Reviews

by Philip Morrison

THE ADAPTIVE GEOMETRY OF TREES, by Henry S. Horn. Princeton University Press ($7.95). At about the time of the Battle of Britain a mixed set of English academics attached themselves closely to various military commands, founding a style of problem-solving called operations analysis. Nowadays their transatlantic intellectual descendants, beholden to big computers and think tanks, work with expensive and lengthy simulations and the full apparatus of mathematical statistics. The old style, however, is not dead. Here it appears, as daring, lighthearted, informal and penetrating as ever, in a small tour de force of a book by a young Princeton biologist who has begun "thinking of trees as crafty green strategists."

The old style is clear. You identify some overriding feature of a complex situation. You attach some numbers to its effects, crudely if need be, content that a couple of free parameters and some good luck about built-in correlations will help you out. You deduce a few tendencies that are well known, and you notice with pleasure that unexpected consequences follow from your equations. In a couple of well-chosen cases the data fit the curves. Then you look happily forward to improving it all in a more realistic second-order theory.

Sunlight is the chief food of trees. The fleeting light on the forest floor is complex to compute or even to measure, with the changing direction, the mixture of direct and diffuse light, and the rest of dappled reality. Still, boldly pointing an ordinary light meter with a carefully designed aperture stop straight up at the canopy above, or photographing the canopy directly (many beautiful fish-eye photographs are reproduced here), serves as a good guide in woods that are uniform enough. Measuring sapling growth, or even counting the saplings, species by species under the canopy, shows a clear order of shade tolerance. The quantitative order of tolerance is close to the order assigned long ago by foresters. Red maple should succeed black gum and be replaced by beech; in the woods of the Institute for Advanced Study in Princeton just that succession is found, demonstrated by the size distribution of the several species, seedlings, saplings, poles and grown trees.

Such ideas lead to a simple model. There are two limiting cases of leaf arrangement: a monolayer, with leaves spread in a single horizontal layer, and a multilayer, with leaves set more loosely but at many levels. The multilayer catches the diffuse light, which is fainter than direct sunlight but can still promote photosynthesis; the photochemical reaction rate is saturated at a quarter of the full sunlight intensity. Thus the open field is first invaded by the multilayered species, which are resistant to water loss because of their low heat load and are fond of high light intensity overhead. But their spaced canopy only partly shades the ground. Then the monolayer sapling enters. It does well where the light is scarcer, because it can take it all and leave little but heavy shade below its dense layer of leaves. It overgrows the multilayers to produce in time a thick canopy, resistant to invasion and supporting its wide horizontal spread of crown on slow-growing, strong hardwood branches: the climax forest. Yet the multilayered white pine and the giant sequoia cast unusually deep shade. These exceptions can settle and maintain a resistant monospecies stand.

There is far more to read in this di-

© 1971 SCIENTIFIC AMERICAN, INC

rect, meaty, small tale of the forests and the trees, both seen sharply, than any review can summarize. The mixture of amateur means and penetrating analysis is entirely winning; to a general reader it seems to be an unqualified success. Perhaps professionals will cavil at its simplicity. Most probably they will grumble that it tells nothing new, but if you walk in the woods and you want to understand, try the jargon-free and good-natured text of this ecology without pretensions.

THE WANKEL ENGINE: DESIGN, DEVELOPMENT, APPLICATIONS, by Jan P. Norbye. Chilton Book Company ($15). Rudolf Diesel dreamed of thermodynamic efficiency. By sea and land the remarkable development of his form of internal-combustion engine—fuel injection, high-pressure construction and low-temperature exhaust—has made his name into a lowercase noun. Diesel took out his first patent in 1892. Will the second half of the century burn hydrocarbons in the name of another engineer who spent "a lifetime devoted to the unwavering pursuit of one basic idea"?

Felix Wankel is a professional automotive engineer, self-taught by decades of running his own machine shop, which has grown to the status of a development laboratory and test center for automobile and motorcycle engines. He has been consultant to the biggest German automobile companies in peace and in war. (He knew prison under both the early Third Reich and the French postwar authority.) His dream of a rotary piston engine led him to become a specialist in the problems of gas sealing, which holds the key to the success of an engine whose four cycles occur at different positions of the same many-cusped rotor. Nearly 70, Wankel remains at work, the model consultant, the independent who helps the big companies and their big teams develop the schemes he starts while he seeks new backers and new paths.

Today all the great engine-builders are making Wankels on at least an experimental basis. Rolls-Royce has built a diesel Wankel, supercharged by a Wankel compressor. In 1970 Citroën produced 500 M-35 two-door coupés, with a front-wheel drive, a hydraulic suspension and a one-rotor Wankel engine built by NSU (the original German development company and still the main patent-holder). These cars are not sold outside France, but within the country their owners are in effect testing for Citroën a possible new economy-car formula. The M-35 is a large-scale experiment, and it will never become a production car; it has too many hand-me-down parts. (It is "an Ami-8 with an engine swap.") It is nonetheless a multimillion-dollar wager on Wankel.

General Motors has paid $10 million a year (from 1970 to 1975) for a nonexclusive license to develop and sell the Wankel engine. Daimler-Benz has built a radical mid-engine fiber-glass sports car with a three-rotor Wankel, the C-111.

The only production Wankel-powered cars today are those of NSU and of Toyo Kogyo of Hiroshima (sold under the name Mazda). All these cars are described and pictured here, with an account of how they feel to drive. The author is an expert automotive journalist. He has written an interesting volume, candid, informal and factual. It is at the level of the popular periodical for the mechanically minded reader; it is never deeply technical, but it does have many engineering graphs and charts (not all of which are to the point). It is well informed, and it takes up finance and legislation as well as rotor seals and temperature distribution. The curious antiquated conventions of the patent drawing ornament many of the pages.

The Wankel is a reality. It can be made on many scales, from lawnmower to locomotive, by multiplying the rotors. It will burn a wide range of fuels well. It is inherently smooth, quiet and vibration-free, and it is as efficient as any ordinary four-cycle engine. It turns up very high speeds without cranks, cams and valve gear. It may wear, but it cannot throw a rod when it is run too fast. Its emissions differ little from those of the usual piston engine; it is perhaps less apt to make nitrogen oxides (peak temperatures run a little lower in the Wankel) but rather more inclined to leak unburned fuel. What seems clear is that for its power the Wankel is simpler, smaller, cheaper and easier to build on automatic assembly lines than any of the baroque cam-valve-and-crank engines of today. The overall gain is fully 25 percent, perhaps much more. That is the heart of the matter; it may well be that such production advantages will enable the rotary engine to dominate the industry. In it lies the promise of engineering "tradeoffs" to meet the real problems of the automobile: safety, emission control, size, cost, noise. The Wankel might be the Fabian evasion of the necessity of an automotive revolution.

STATISTICAL ABSTRACT OF THE UNITED STATES: 1971 (92nd edition). U.S. Bureau of the Census ($5.50). The new year brings a crop of almanacs and yearbooks bearing facts and figures for the bookshelves of the land. Several excellent proprietary ones have been reviewed here in other years; let us now praise the fountainhead. The *Abstract* is in some ways not an annual but a decennial; this year it has among its 1,300 tables much from the Constitutional Census of 1970. Eighty years after Herman Hollerith's first punched-card census, the six months up to mid-July, 1971, were not time enough to forestall the admission that a "more comprehensive selection of data from the 1970 census ... will be presented in the 1972 edition."

This is no petty fact book; austere, without the relief of moon phases or of cup winners or of headlined catastrophes, it brings the story of who we are, we Americans, where we cluster, how we are born, die and suffer, where we go and whence we came. It is a ledger of our economy, our elections, our taxes and money, our net of electromagnetic waves and wires at high current and at low, our forests, fisheries, mines and farms, and our houses and factories. Nearly a tenth of the thick volume is devoted to a table listing the 148 "standard metropolitan statistical areas," regions with 200,000 population or more and a central city or two of at least 50,000. The list reports 218 items for each area: the counts of black and white, young and old, city and suburb, marriages and income, work and welfare, housing, trade and crime. Another 40 pages summarize international data chiefly from the United Nations, a most valuable source of comparisons. Above all, this work gives the primary sources of all its tables in some detail in a guide of 50 pages, an invaluable feature for the serious student.

Here is a small sample of the bald facts. The New Jersey area of Paterson-Clifton-Passaic has the highest per capita personal income of all the standard metropolitan statistical areas, enjoying nearly twice the annual income per head of Johnstown, Pa., which is about the lowest on the same basis. It is true that four-fifths of the people around Paterson live outside the central city, but Johnstown is even more suburban. Only 2 percent of Johnstown's people are "Negro and other"; in the New Jersey area the same category is 7 percent. Factory wage rates in the two areas are much the same; the crime rate, however, is far lower in Johnstown. These figures conceal rather than reveal the ebb and flow of industrial fortune.

Not all the data are so deceiving. The

© 1971 SCIENTIFIC AMERICAN, INC

Plant a THINK TANK anywhere and watch the minds grow!

home-office-school-park-club-churches-laboratory

Unique instructional games designed by university professors to make learning fun through brain-to-brain action. Beginning games can be mastered by young children—final games will challenge intelligent adults. These are the famous GAMES FOR THINKERS from WFF 'N PROOF Publishers. Order single kits or save $10.00 by ordering the complete 9-kit THINK TANK.

WFF 'N PROOF (logic)	$ 8.75
WFF (beginner's logic)	1.75
EQUATIONS (mathematics)	5.50
ON-SETS (set theory)	5.50
PROPAGANDA (social studies)	6.50
CONFIGURATIONS (geometry)	5.50
TRI-NIM (problem solving)	4.50
QUERIES 'N THEORIES (science)	8.75
QWIK-SANE (topology puzzle)	1.75
RODIN'S THINKERS BOOKENDS	11.00
Complete 9-kit THINK TANK	
with bookends	49.50
without bookends	39.50

Order from: **WFF 'N PROOF,** 1111-DA Maple Avenue, Turtle Creek, Pa. 15145

*postage included in above prices

Fully guaranteed. Dealer inquiries invited.

Gifts that are a COMPLIMENT to receive!

AMATEUR TELESCOPE MAKING

Edited by Albert G. Ingalls

Book One

497 pages, 300 illustrations
$5.00 postpaid. $5.35 foreign

Book Two

650 pages, 361 illustrations
$6.00 postpaid. $6.35 foreign

Book Three

644 pages, 320 illustrations
$7.00 postpaid. $7.35 foreign

Send postcard
for descriptive circular

SCIENTIFIC AMERICAN

415 Madison Avenue, New York, N.Y. 10017

(Residents of New York City please add 7% sales tax)

(Other NYS residents please add 4% state sales tax plus local tax)

lifetime income of male college graduates is nearly three times that of grade school dropouts; the discrepancy was even a little greater 20 years ago. The assets of the Department of Defense, reckoned per employee in or out of uniform, puts it up there with mining, oil and tobacco; these outfits are much more heavily capitalized per head than any other class of large corporations. For 30 years the calorie intake of the average American has stayed about the same, dropping only 1 or 2 percent; fat and protein have risen as starches drop, keeping the energy flow fixed. We eat more beef now, and fewer potatoes; we consume twice as much meat per mouth as the Swedes, and 30 percent more than the French. If extrapolation does not deceive, the centroid of population will cross the Mississippi some 20 miles below St. Louis before our next national census.

Readers can obtain the metropolitan area statistics as a separate publication for 35 cents. Many will be interested to know that a companion volume of historical tables, spanning our entire past from Colonial times up to 1957 (with a paperback supplement to 1962), is also available.

THE COMPACT EDITION OF THE OXFORD ENGLISH DICTIONARY: COMPLETE TEXT REPRODUCED MICROGRAPHICALLY. VOLUME I, A–O. VOLUME II, P–Z. WITH SUPPLEMENT AND BIBLIOGRAPHY ($75). For 70 years the devoted scholars labored "to present in alphabetical series the words that have formed the English vocabulary." Their work was based on five million excerpts from English literature of every period, "the only possible foundation for the historical treatment of every word and idiom." An army of volunteers sent in citations, one man contributing 165,000 of them! The 10 principal volumes appeared in 1928, with ceremonial presentation of the first copies to George V and the laconic Calvin Coolidge. A smaller supplement, emphasizing the words and usages of the 20th-century burst of science and technology, and of the English language overseas, completed the work in 1932. Since then every substantial library has been proud of the 10 thick volumes and the three thinner ones, stretching down an entire shelf. There rests a treasury of our language, with each word discussed in the large. (The famous entry for the verb "set" fills 18 pages, and a few more pages dispose of the noun.) It comes to 50 million words and 16,000 pages.

"Spectrum" (from *spectare,* "to look") took its modern meaning with Newton's own description of his famous experiment. At least no earlier citation is given, although one suspects that Newton, who does not explicitly define his English use of the familiar Latin word, did not regard it as a novelty to refer to "Image or Spectrum." "Molecule" too is almost straight Latin, a diminutive for "mass"; it seems to have arisen in 17th-century discussions of Cartesianism. "Energy" is good Greek, found in Aristotle, but its definition in physics arose with Thomas Young in 1807. "Gene" surfaces in 1913, when it was already in a medical dictionary; "zoom," somewhat surprisingly, buzzed along in the 1880's.

This edition is notable: it is a remarkable miniaturization of the entire work. Sold at a quarter of the price of the full-sized edition, it presents every word very clearly in just two massive volumes. The compression results from photographic reduction in page size, so that four pages of the original edition appear reproduced on each page of the new edition. The reproduction is praiseworthy for its sharpness and contrast. The smallest typeface is about a fourth as high as the type you are now reading. The edition is printed on a thin, opaque paper, bound into heavy volumes of 2,000 pages, compared with the 1,200 pages of the original volumes. Together these devices gain the factor of nearly six by which the edition is squeezed in bulk.

One must emphasize the readability of the result; very large index headings (the only additions) marking each page make it easy to find the entry, which is large enough to read swiftly with the unaided eye. The quotations, however, cannot be read conveniently by someone with normal vision, let alone by someone with failing visual accommodation. (A magnifier supplied with the volumes is optically well chosen but mechanically barely adequate.) The edition, a happy outcome of modern technology, was suggested by Albert Boni, whose firm has similarly squeezed space out of important library reference works, such as the Catalogue of the British Museum. Skillful printers now display the art of close to one order of magnitude compression, while preserving direct human random access. One hopes this fine example will not be the last one produced for general readers, although clearly the wide appeal of a great dictionary allows the cost to be thinly spread. What bookshelf have you longed to own?

TIME AND THE SPACE-TRAVELLER, by L. Marder. George Allen & Unwin, Ltd (65 shillings). It was Paul Langevin who first pointed out the curious effect

© 1971 SCIENTIFIC AMERICAN, INC

of the relativity of simultaneity on the age of a round-trip traveler. That was all the way back in 1911. Not until now—and one may be overoptimistic—has experiment given a direct verdict: the fast-traveling twin, his speed always judged with respect to the great inertial frame of the galaxies, retains his youth. Only last fall did precise clocks circle the earth, going first class by jet, two flying in opposite directions to allow rough cancellation of the large effects of altitude. Plane travel is hardly what most physicists think of as relativistic, but present atomic clocks keep time to better than 100 nanoseconds per day; they do record an elapsed time less by a couple of hundred nanoseconds on the faster-circling eastbound air voyage, the one that moves with the spin of the earth and not against it. The faster-moving twin does in fact age less than his slower-moving brother.

This unusually clear little monograph, whose mathematics is simple algebra, only here and there an integral sign, treats the twin paradox fairly and broadly, everywhere with good sense and good physics. The British author gives us the pleasure of viewing the problem in a wide context, examining the limits and prospects of space travel, the nature of biological clocks, the experiments that bear on the question, from mesons to Mössbauer, and much more. But his central theme is what happens to twins who part and reunite, explained correctly against an attractive and explicit background of classroom special relativity.

It is puzzling that a dozen successful physicists, experimenters and theorists alike, have over the years steadily and stoutly insisted on the impossibility of what Professor Marder and nearly all other physicists find evident in the equations of Einstein and in a series of relevant, if not direct, experiments. There are more than 300 references in this book; all the credible controversialists receive a fair exposition. Once in a while Marder's patience and good humor allow him a dry little murmur of surprise at a particularly doubtful argument, but the book is admirable for the open and winning way in which it manages to describe and evaluate positions the author cannot accept. Experience suggests that neither this book nor the jet voyages will convince everyone; nothing else has. The account of the puzzle in its setting, its implications and the resolute skeptics nonetheless make the book convincing and enjoyable reading for anyone with a taste for physical argument, whether he already knows about the Lorentz transformation or only wants to learn.

INDEX OF ADVERTISERS

JANUARY 1972

ATLANTIC RICHFIELD COMPANY ..... Inside Front Cover
Agency: Needham, Harper & Steers, Inc.

CALCULATOR COMPANY, THE 8
Agency: John H. Schatteles Associates

CALIFORNIA COMPUTER PRODUCTS, INC. 12
Agency: Dailey & Associates

CELESTRON PACIFIC, INC. 10
Agency: Baytron, Inc.

CHILTON BOOK COMPANY 51
Agency: Sussman & Sugar, Inc.

CHRYSLER CORPORATION, CHRYSLER-PLYMOUTH DIVISION 52
Agency: Young & Rubicam, Inc.

DIGITAL EQUIPMENT CORPORATION 1
Agency: Creamer, Trowbridge, Case & Basford, Inc.

EASTMAN KODAK COMPANY 43
Agency: Rumrill-Hoyt, Inc.

FREEMAN, W. H., AND COMPANY 69, 116

GALLERY, THE 117
Agency: Fairfax, Inc.

HARIAN PUBLICATIONS 123
Agency: Braco/Briskin Assoc. Inc.

HEWLETT-PACKARD 48, 49
Agency: Richardson Seigle Rolfs & McCoy

HISTORY BOOK CLUB, THE 53
Agency: Schwab, Beatty & Porter, Inc.

INDIANA, STATE OF, DEPARTMENT OF COMMERCE 11
Agency: Ruben, Montgomery & Associates, Inc.

KING-SIZE CO., THE 10
Agency: Forest Advertising Corp.

LIBRARY OF SCIENCE, THE 112
Agency: Henderson & Roll, Inc.

LILLY, ELI, AND CO. ..... Back Cover
Agency: Geer, DuBois & Co., Inc.

MANPOWER INC., TECHNICAL SERVICES DIVISION 5
Agency: Nader-Lief, Inc.

NATIONAL CAMERA, INC. 8
Agency: Langley Advertising Agency

PHILIPS RESEARCH LABORATORIES 2
Agency: TAG Intermarco Delamar

PHILLIPS PETROLEUM COMPANY 7
Agency: J. Walter Thompson Company

PLASTICS ENGINEERING COMPANY 9
Agency: Kuttner & Kuttner, Inc.

POLAROID CORPORATION, THE 62, 63
Agency: Doyle Dane Bernbach Inc.

QUESTAR CORPORATION 45

RENAULT, INC. 44
Agency: Gilbert Advertising Agency, Inc.

UNITED AUDIO PRODUCTS, INC. 4
Agency: Ries Cappiello Colwell, Inc.

UNIVERSITY COMPUTING COMPANY ..... 124, Inside Back Cover
Agency: Warner, Bicking & Fenwick, Inc.

WFF'N PROOF 120
Agency: Ad-Com Agency

© 1971 SCIENTIFIC AMERICAN, INC

BIBLIOGRAPHY

Readers interested in further reading on the subjects covered by articles in this issue may find the lists below helpful.

EXTENDING THE NUCLEAR-TEST BAN

DEVELOPMENTS IN TECHNICAL CAPABILITIES FOR DETECTION AND IDENTIFYING NUCLEAR WEAPONS TESTS: HEARINGS BEFORE THE JOINT COMMITTEE ON ATOMIC ENERGY. Congress of the United States. U.S. Government Printing Office, 1963.

NUCLEAR TEST BAN TREATY: HEARINGS BEFORE THE COMMITTEE ON FOREIGN RELATIONS. United States Senate. U.S. Government Printing Office, 1963.

NATIONAL SECURITY AND THE NUCLEAR-TEST BAN. Jerome B. Wiesner and Herbert F. York in *Scientific American,* Vol. 211, No. 4, pages 27–35; October, 1964.

PROSPECTS FOR A COMPREHENSIVE TEST BAN TREATY: HEARINGS BEFORE THE SUB-COMMITTEE ON ARMS CONTROL, INTERNATIONAL ORGANIZATIONS. United States Senate. U.S. Government Printing Office, 1971.

RNA-DIRECTED DNA SYNTHESIS

ONCOGENIC RIBOVIRUSES. Frank Fenner in *The Biology of Animal Viruses: Vol. II.* Academic Press, 1968.

CENTRAL DOGMA OF MOLECULAR BIOLOGY. Francis Crick in *Nature,* Vol. 227, No. 5258, pages 561–563; August 8, 1970.

MECHANISM OF CELL TRANSFORMATION BY RNA TUMOR VIRUSES. Howard M. Temin in *Annual Review of Microbiology: Vol. XXV.* Annual Reviews, Inc., 1971.

THE PROTOVIRUS HYPOTHESIS: SPECULATIONS ON THE SIGNIFICANCE OF RNA-DIRECTED DNA SYNTHESIS FOR NORMAL DEVELOPMENT AND FOR CARCINOGENESIS. Howard M. Temin in *Journal of the National Cancer Institute,* Vol. 46, No. 2, pages III–VII; February, 1971.

HOW IDEOLOGY SHAPES WOMEN'S LIVES

FAMILY, SOCIALIZATION AND INTERACTION PROCESS. Talcott Parsons and Robert F. Bales. The Free Press of Glencoe, 1955.

CHILDREN'S CONCEPTS OF MALE AND FEMALE ROLES. Ruth E. Hartley in *Merrill-Palmer Quarterly,* Vol. 6, No. 2, pages 83–92; January, 1960.

THE DEVELOPMENT OF SEX DIFFERENCES. Edited by Eleanor E. Maccoby. Stanford University Press, 1966.

THE SCANNING ELECTRON MICROSCOPE

FACTORS AFFECTING CONTRAST AND RESOLUTION IN THE SCANNING ELECTRON MICROSCOPE. T. R. Everhart, O. C. Wells and C. W. Oatley in *Journal of Electronics and Control,* Vol. 7, No. 2, pages 97–111; August, 1959.

SCANNING ELECTRON MICROSCOPY. C. W. Oatley, W. C. Nixon and R. F. W. Pease in *Advances in Electronics and Electron Physics: Vol. XXI,* edited by L. Marton. Academic Press, 1965.

THE SCANNING ELECTRON MICROSCOPE: PRINCIPLES AND APPLICATIONS IN BIOLOGY AND MEDICINE. T. L. Hayes and R. F. W. Pease in *Advances in Biological and Medical Physics: Vol. XII,* edited by John H. Lawrence and John W. Gofman. Academic Press, 1968.

GEOTHERMAL POWER

PROCEEDINGS OF THE UNITED NATIONS CONFERENCE ON NEW SOURCES OF ENERGY. VOL. II, GEOTHERMAL ENERGY: I. United Nations, 1964.

PROCEEDINGS OF THE UNITED NATIONS CONFERENCE ON NEW SOURCES OF ENERGY. VOL. III, GEOTHERMAL ENERGY: II. United Nations, 1964.

THE ECONOMIC POTENTIAL OF GEOTHERMAL RESOURCES IN CALIFORNIA. Geothermal Resources Board, State of California, January, 1971.

GEOTHERMAL RESOURCE INVESTIGATIONS, IMPERIAL VALLEY OF CALIFORNIA—STATUS REPORT. United States Department of the Interior, Bureau of Reclamation, April, 1971.

UNITED NATIONS SYMPOSIUM ON THE DEVELOPMENT AND UTILIZATION OF GEOTHERMAL RESOURCES (A SUMMARY REPORT). United Nations ST/TAO/Ser. C/126, 1971.

THE EXPLOITATION OF GEOTHERMAL RESOURCES IN THE PRESENT AND FUTURE. Gunnar Bodvarsson. United Nations, in press.

THE SPECTRUM OF THE AIRGLOW

PHYSICS OF THE AURORA AND AIRGLOW. Joseph W. Chamberlain. Academic Press, 1961.

AURORA AND AIRGLOW. Edited by Billy M. McCormac. Reinhold Publishing Corporation, 1967.

WHY THE STOMACH DOES NOT DIGEST ITSELF

THE STOMACH. Stewart Wolf. Oxford University Press, 1965.

SALICYLATE DAMAGE TO THE GASTRIC MUCOSAL BARRIER. Horace W. Davenport in *The New England Journal of Medicine,* Vol. 276, No. 23, pages 1307–1312; June 8, 1967.

BACK DIFFUSION OF ACID THROUGH THE GASTRIC MUCOSA AND ITS PHYSIOLOGICAL CONSEQUENCE. Horace W. Davenport in *Progress in Gastroenterology: Vol. II,* edited by George B. Jerzy. Grune & Stratton, 1970.

GASTRIC MUCOSAL BARRIER. Kevin J. Ivey in *Gastroenterology,* Vol. 61, No. 2, pages 247–257; August, 1971.

POPULATION GENETICS AND HUMAN ORIGINS

STUDIES ON THE EVOLUTION OF THE PRIMATES. William K. Gregory in *Bulletin of the American Museum of Natural History,* Vol. 35, pages 239–355; June 16, 1916.

PRELIMINARY REVISION OF THE DRYOPITHECINAE (PONGIDAE, ANTHROPOIDEA). E. L. Simons and D. R. Pilbeam in *Folia Primatologica,* Vol. 3, No. 2–3, pages 81–152; 1965.

MATHEMATICAL GAMES

DISJUNCTIVE GAMES WITH THE LAST PLAYER LOSING. P. M. Grundy and C. A. Smith in *Proceedings of the Cambridge Philosophical Society,* Vol. 52, Part 2, pages 527–533; July, 1956.

THE G-VALUES OF VARIOUS GAMES. Richard K. Guy and Cedric A. B. Smith in *Proceedings of the Cambridge Philosophical Society,* Vol. 52, Part 2, pages 514–526; July, 1956.

COMPOUND GAMES WITH COUNTERS. Cedric A. B. Smith in *Journal of Recreational Mathematics,* Vol. 1, No. 2, pages 67–77; April, 1968.

THE AMATEUR SCIENTIST

SOLAR ECLIPSES AND OCCULTATIONS. Ivan I. Mueller in *Spherical and Practical Astronomy as Applied to Geodesy.* Frederick Ungar Publishing Co., 1969.

THE 1972 OCCULTATION SUPPLEMENT. *Sky and Telescope,* Dept. SA, 49–50–51 Bay State Road, Cambridge, Mass. 02138.

© 1971 SCIENTIFIC AMERICAN, INC

STOP SAYING . . .

"I CAN'T AFFORD TO TRAVEL"

8 BOOKS THAT GIVE YOU THE FACTS ON HOW YOU CAN TRAVEL TODAY WITHOUT BEING RICH

AROUND THE WORLD BY FREIGHTER

Where and how to travel by freighter —the lower cost way to travel

For no more than you'd spend at a resort, you can take a never-to-be-forgotten cruise to Rio or Buenos Aires. Or through the Canal or to the West Indies or to England, France, the Mediterranean, etc.

And what accommodations you get—large rooms with beds (not bunks), probably a private bath, lots of good food, and plenty of relaxation as you speed from port to port.

Travel Routes Around the World names the freighter lines (700 of them, with sailings from practically every port in the world), tells where they go, what they charge, briefly describes accommodations plus life on your freighter, clothes to take, etc.

To stop saying that travel is expensive get your copy now. Price $1.50.

AMERICA BY CAR

This big book is your insurance of seeing all the 4-star sights in whatever corner of the U.S., Canada, or Mexico you drive to. Whether you're visiting New England or California, Florida or the National Parks, the Great Lakes, the Mississippi, the East, the South, the Southwest, the Indian country, etc., it tells you day by day and road by road the scenic way to go and it always directs you to the important sights along the way and in the cities. In Niagara or Los Angeles, Washington or New Orleans, the Black Hills or Montreal, it takes the guesswork out of travel.

America is so big you can easily overlook or forget important sights or make many a wrong turn. So get *America by Car,* the book that makes sure you'll see everything of consequence and always travel right. Only $3.50 for this 170,000 word book (as big as 3 ordinary sized novels).

FABULOUS MEXICO—WHERE EVERYTHING COSTS LESS

The land of retirement and vacation bargains where you can build a modern home for $7500 and an American retirement income looks like a fortune, and your vacation money can buy double or more what it might back home. Norman Ford shows you vacation and retirement values where you can live like a prince on what you might just get along on in the U.S.A. He pinpoints areas that look like the South Seas, others where it's like June all year round, towns where many other Americans have retired; shows where to find modern flower-bedecked hotels and inns that charge hardly half of what you'd expect to spend in even such a land of vacation and retirement bargains as Mexico. *Plus a big section where to start your money earning so much more than in the U.S.A.* $2.50.

OFF-THE-BEATEN PATH

—these are America's own Bargain Paradises

Where to retire or vacation at what look like prewar prices and no one ever heard of nerves or worries.

Off-the-Beaten Path names the really low cost Florida retirement and vacationing towns, the topnotch values in Texas, the Southwest, California, the South and East, Canada, and a dozen other areas which the crowds have not yet discovered:

—Fabulous places like that undiscovered region where winters are as warm and sunny as Miami Beach's, yet costs can be 2/3rds less. Or that island that looks like Hawaii yet is 2000 miles nearer. Or France's only remaining outposts in this part of the world . . . or a village more Scottish than Scotland . . . or resort villages without crowds or high prices . . . or island paradises aplenty in the U.S. or Canada . . . or areas with almost a perfect climate. And for good measure you also read about low cost paradises in Hawaii, the Virgin Islands, and Puerto Rico.

A big book, with about 100,000 words. Yet it costs only $2.50.

WHERE WILL YOU GO IN FLORIDA?

Florida needn't be expensive—not if you know just where to go for whatever you seek in Florida. And if there's any man who can give you the facts you want, it's Norman Ford, founder of the world-famous Globe Trotters Club.

His big book, *Norman Ford's Florida,* tells you, first of all, road by road, mile by mile, everything you'll find in Florida, whether you're on vacation or looking over job, business, real estate, or retirement prospects.

Always, he names the hotels, motels, and restaurants where you can stop for the best accommodations and meals at the price you want to pay. For that longer vacation, if you let Norman Ford guide you, you'll find a real "paradise"—Just the spot which has everything you want.

Of course, there's much more to this big book. If you want a home in Florida, he tells you just where to head. If you've ever wanted to run a tourist court or own an orange grove, he tells you today's inside story of these popular investments.

If you want to retire on a small income, Norman Ford tells you exactly where you can retire now on the money you've got, whether it's a little or a lot. Because he always tells you where life in Florida is pleasantest on a small income, he can help you to take life easy now.

Whatever you seek in Florida, *Norman Ford's Florida* gives you the facts you need to find exactly what you want. Well over 100,000 words but it costs only $3.00—only a fraction of the money you'd spend needlessly if you went to Florida blind.

ALL ABOUT ARIZONA —the healthful state

Just as a road map shows you how to reach your destination, this big book leads you to whatever you want in this fast growing state of sun and scenic wonderlands.

What do you want to know about Arizona? Where to retire at low cost? Where are summers cool, winters sunny most of the time? Where are the leading places for a job, a home, etc? What must a newcomer watch out for? Is it true that living costs are less than in the East? What about salaries?

Or do you want to tour this Grand Canyon State? What's the most scenic way to see Arizona by car or otherwise? What is really the most satisfying way to see the Grand Canyon? The Indian reservations? The other 4-star sights? Which are the outstanding places to eat and stay? What are the sure ways to cut travel costs in this big state?

Filled with facts, over 100,000 words long, this book almost brings Arizona to your door answering these and a hundred other questions. To know all you should about Arizona before you go for a home, a job, retirement in the sun, or a really memorable vacation, read this book. Price, $2.95.

SPECIAL OFFER: All 4 books above—*Travel Routes Around the World, America by Car, Fabulous Mexico—Where Everything Costs Less,* and *Off-the-Beaten Path*—($10 value) for only $6.95.

BARGAIN PARADISES OF THE WORLD

West Indies, Mexico, Californias Abroad

This is a book on how to double what your money can buy. For that is what spending a few weeks or months, or even retiring, in the world's Bargain Paradises amounts to.

Throughout this big book you learn where to spend a while in the West Indies, South America, the healthful islands of the South Seas, and the marvelous Balearic Islands where two can live like kings for $50 a week.

You read about cities and towns where it's always spring, about "Californias Abroad," about "Four Modern Shangri-Las," about mountain hideaways, tropical islands as colorful as Tahiti but nearer home, about modern cities where you can live for less, about quiet country lanes and surf-washed coastal resorts.

If you've ever wanted to travel but wondered how you could afford it; if you have a little income but wonder how you'd ever be able to retire on that; if you want a life of luxuries on what you'd get only necessities back home, then you want this book. $2.50.

WHERE TO RETIRE ON A SMALL INCOME

This book selects out of the thousands of communities in the U.S. only those places where the climate is right, living costs are less, the surroundings pleasant, and nature and the community get together to guarantee a good time from fishing, boating, gardening, concerts or the like.

It covers cities, towns, spas, resorts, etc., throughout America—from New England south to Florida, west to California and north to the Pacific Northwest. It includes both Hawaii and the American Virgin Islands.

Some people spend hundreds of dollars trying to get information like this by traveling around the country. Frequently they fail—there is just too much of America to explore. This book saves you from that danger. Yet it costs only $2.50.

A good trip begins with a Harian book

Publishers since 1935

Mail to HARIAN PUBLICATIONS,
74 Walnut Drive
GREENLAWN (Long Island), N.Y. 11740

I have enclosed $ .....(cash, check, or money order). Please send me the books I checked below. YOU WILL REFUND MY MONEY IF I AM NOT SATISFIED.

- ☐ **Travel Routes Around the World** (travel by freighters). **$1.50.**
- ☐ **America by Car. $3.50.**
- ☐ **Fabulous Mexico—Where Everything Costs Less. $2.50.**
- ☐ **Off-the-Beaten Path. $2.50.**
  - ☐ SPECIAL OFFER #1. All 4 books above for $6.95.
- ☐ **Norman Ford's Florida. $3.00.**
- ☐ **All About Arizona—the healthful state. $2.95.**
- ☐ **Where to Retire on a Small Income. $2.50.**
- ☐ **Bargain Paradises of the World. $2.50.**
  - ☐ SPECIAL OFFER #2: Save $8—all 8 books above, $20.95 value, for only $12.95.

Print Name .....

Street Address.....

City .....

State .....Zip Code .....

© 1971 SCIENTIFIC AMERICAN, INC

UCC UNIVERSITY COMPUTING C

Customer Service Headquarter

1500 UCC Tower
214/637-0600

P.
Dallas,

© 1971 SCIENTIFIC AMERICAN, INC

There's a real difference between a computer company and a computing company.

MPANY

). Box 6228
exas 75222

Profit...Your profit!

We're computing people who know how to reduce your computing costs.

We begin by stepping into your shoes. We don't start out with a solution looking for problems to match. And we don't shoehorn your problems into pat, packaged answers. Your objectives—become our objectives.

We use task-force management techniques to assemble teams of performance-oriented experts to optimize the return on your computing investment...you profit from the know-how of our more than 2,500 computing and data communications professionals. Augmented by equipment and services that we developed because our customers needed cost-effective results now—when the computer companies were saying "Not now".

Results like—

- A computer network equipped with IBM, Honeywell, Univac, Control Data and Burroughs systems at more than 30 centers providing customized processing results—a little or a lot—for our customers.
- Our own UCC family of COPE® high-speed remote batch interactive terminals that work with major computer systems.
- A cost/time saving converter that lets IBM users run existing 360 DOS programs on the larger OS systems in either 360 or 370 environments.
- A UCC COPE® Communications Controller that enables all major makes and models of computers and terminals to work together as a total system.

As a customer-oriented computing company, UCC provides many types of computing products and services—communications network design, systems management, remote batch processing, commercial data processing, contract programming, computer graphics, consulting, specialty applications and even computer system leasing.

Satisfaction—

After a few in-depth discussions, we'll give you a detailed proposal that blends your resources and ours to satisfy both your immediate and long-term profit objectives.

And we'll be specific about dollars and timetables and performance. At the very least, we'll give you a highly professional "second opinion". For that reason alone, a get-together certainly will be profitable.

Just write or call, and we'll get down to specifics fast.

Our phone number is (214) 637-0600. Our address is 1520 UCC Tower, P.O. Box 6228, Dallas, Texas 75222.

Performance is our product

UCC
UNIVERSITY COMPUTING COMPANY

© 1971 SCIENTIFIC AMERICAN, INC

For four generations we've been making medicines as if people's lives depended on them.

Lilly

ELI LILLY AND COMPANY, INDIANAPOLIS

© 1971 SCIENTIFIC AMERICAN, INC