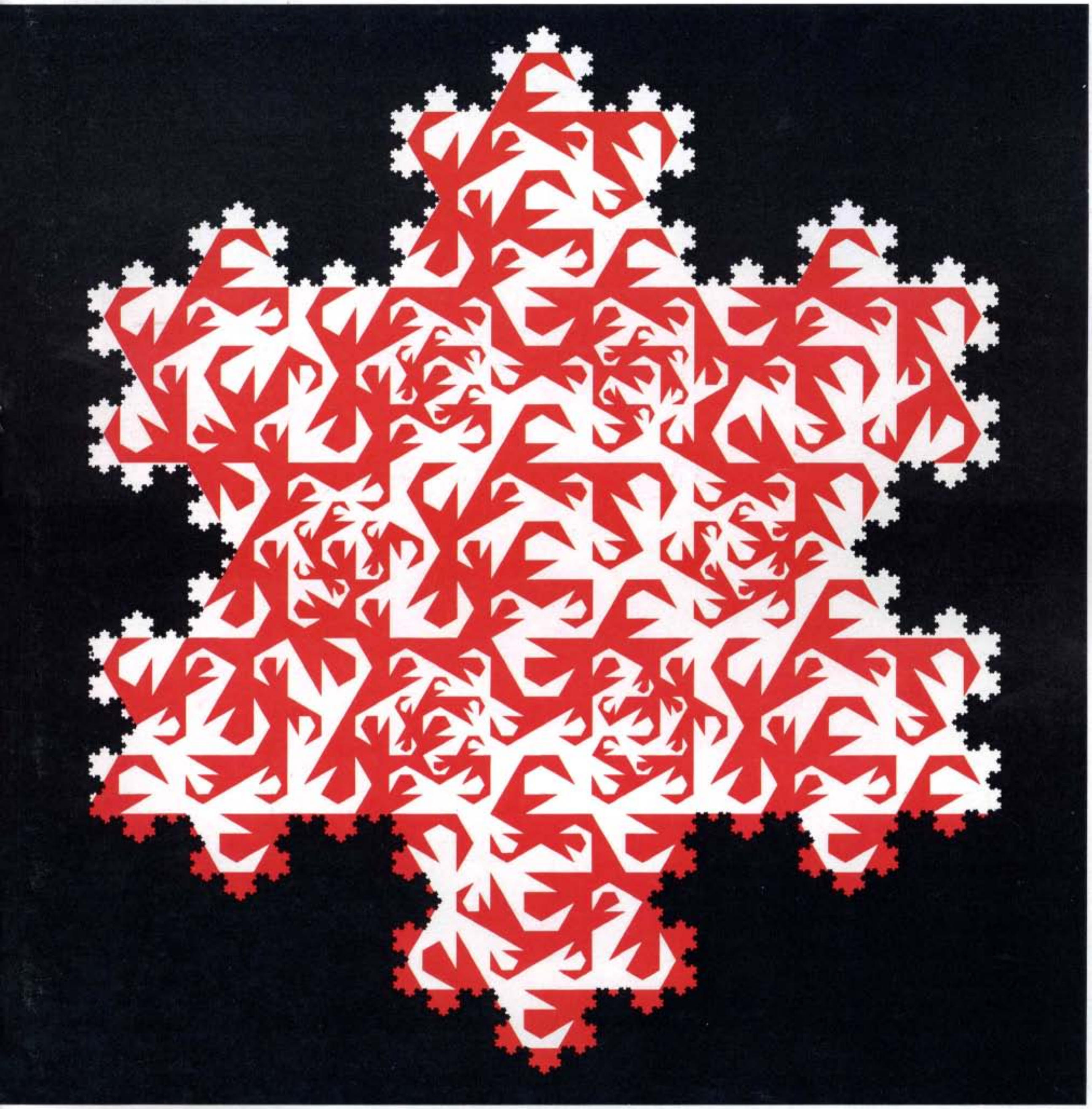
SCIENTIFIC AMERICAN



FRACTAL CURVES

$1.50

April 1978


897-CKK

6 separate brakes, shock absorbers even for the engine, a wax job your eyes may never see: inside the insides of the Mercedes-Benz 450 SEL.

In the eyes of a Mercedes-Benz engineer, the examples of skill and care normally hidden from sight deep within the car are what most strongly stamp a 450SEL as a 450SEL and make it unique.

Herewith, that deep-down 450SEL story – an *engineer's-eye* view.

6,000 invisible welds

Observe that beneath its high-gloss enamel, the 450SEL body is a semi-monocoque or "unitized" structure – a steel hull in effect, studded with no fewer than 6,000 electric welds.

Between the time it first enters and finally leaves the paint shop, this body acquires 77 pounds' worth of protective coatings applied in 9 steps. Anti-corrosion wax is even sprayed into hollow recesses like the insides of the rocker panels before they are permanently sealed.

Almost 5 feet of hiproom

The 450SEL body shell is the largest Mercedes-Benz builds in volume production and the word for its interior space is ample. Yet this so-called "big" Mercedes is contained within an overall length of less than 18 feet–and at 73.6 inches of width is actually *slimmer* than a Nash Rambler compact of 28 years ago.

Linked to its own separate axle, each rear wheel of the 450SEL is freed to act on its own. Result: the left wheel can hit a bump or a dip without affecting the right, and vice versa. Each shock absorber has an upper chamber filled with oil and a lower chamber filled with pressurized *nitrogen* gas as a cushion against vibration.

Excessive vibration is good for neither a car nor its occupants. The engine's rubber-insulated mounts are augmented by two special hydraulic shock absorbers nestled on either side of the 450SEL's 4.5-liter V-8.

Examine the steering system and you will find yet *another* shock absorber, expressly meant to soak up thumps and bumps before they reach your hands on the wheel.

A black box that isn't there

The engineers eliminated pushrod-actuated valves and their adjustment (as well as their clatter) from the 450SEL engine by giving it overhead camshafts instead. Fuel injection is standard–but note: Mercedes-Benz began using fuel injection so long ago (in 1954) that the 450SEL has passed beyond the computerized "little black box" concept to a CIS, or Continuous Injection System, of simplified *mechanical* design.

2 agile tons

If you could peek inside the 450SEL's 3-speed automatic transmission you would find that all shafts and gears in the mechanism are mounted on roller or ball bearings.

This two-ton automobile is maneuvered via a power-assisted steering system based on the recirculating-ball principle. You can ease from lock to lock in a mere 2.7 turns.

Count all the brakes

It is an intriguing fact that the 450SEL is fitted with *six* separate brakes, of two different types:

□ Four disc brakes, each 11 inches in diameter, provide the basic stopping force. The front discs are honeycombed with internal ventilation slots for more rapid heat dissipation.

□ Two drum brakes, one at each rear wheel, handle parking and emergency braking needs.

Velour, veneer, bituminous foil

Soft velour carpeting lies underfoot in the 450SEL. Also underfoot but out of sight: felt, foam-backed rubber matting and even *bituminous foil,* discreet allies in the engineers' quiet war against engine and road noise.

The built-in AM/FM stereo cassette unit has four wide-range speakers. Vacuum power instantly locks or unlocks all four doors, the trunk and the fuel filler port with the twist of a key. The built-in automatic cruise control device is a machine with a memory, able to return the car to a preset speed after you have speeded up or slowed down.

You can specify either leather upholstery or plush velour at no extra cost. You should know that inside the seats are five separate layers of padding and lining, laid overtop a network of steel springs.

Those rich fillets of trim that look like walnut veneer *are* walnut veneer. That dashboard clock is not a clock–it is a quartz-crystal chronometer.

Final revelation: The 60-page Mercedes-Benz brochure covers safety features built into the 450SEL and other models. This makes fascinating reading before you buy the car. and comforting reading afterward.

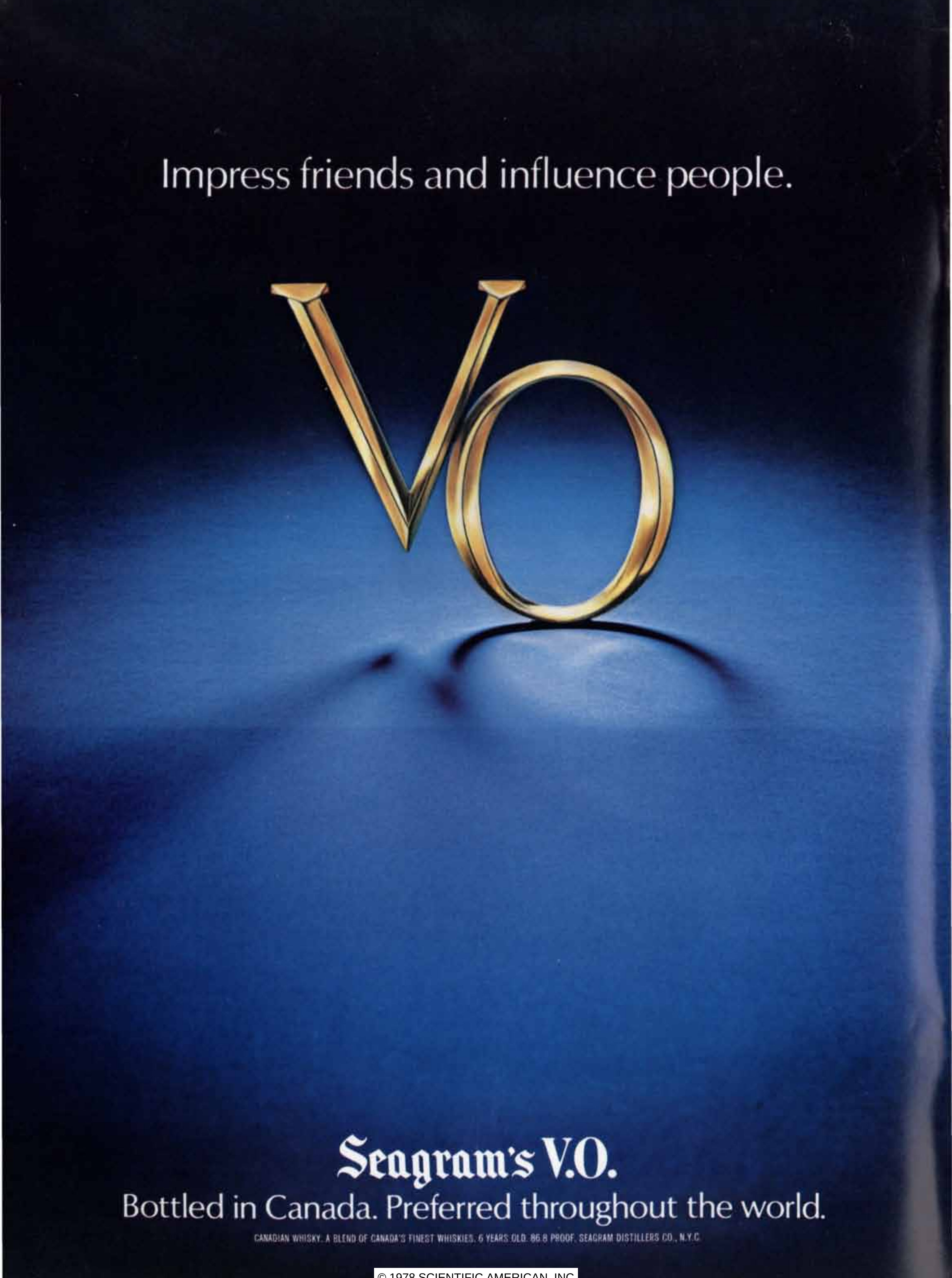
Engineered like no other car in the world

People give many reasons for choosing a Mercedes-Benz. But the company's aim in designing and constructing them is doggedly single-minded. It is to build safe, comfortable, practical cars with as few imperfections as possible.

This philosophy puts engineering ahead of petty economies and precludes the pea-pod mass production of inexpensive cars. It allows little room for compromise, or for shortcuts; just the pursuit of engineering excellence.

A Mercedes-Benz is engineered like no other car in the world.

Electric sunroof (shown) available at extra cost.



Impress friends and influence people.
Seagram's V.O.
Bottled in Canada. Preferred throughout the world.
CANADIAN WHISKY. A BLEND OF CANADA'S FINEST WHISKIES. 6 YEARS OLD. 86.8 PROOF. SEAGRAM DISTILLERS CO., N.Y.C.

Established 1845 SCIENTIFIC AMERICAN April 1978 Volume 238 Number 4

ARTICLES

DEPARTMENTS

PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. 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: $18 PER YEAR, U.S., ITS POSSESSIONS AND CANADA; $22 PER YEAR, ALL OTHER COUNTRIES.

executive health

the report that briefs you on what to watch

Volume XIV, No. 4 • January, 1978 • Pickfair Bldg., Rancho Santa Fe, Calif. 92067 • Area 714:756-2600

Linus Pauling, Ph.D.

VITAMIN C AND HEART DISEASE

Can vitamin C protect you, and how much should you take?

Heart disease and related diseases of the circulatory system are the main cause of death in the United States. Over one million people die of these diseases each year, and probably more than five million people now living are suffering from them in a significant way.

There is no doubt that heart disease is related to the diet. In the 1976 Congressional Hearings on the relation between diet and disease the nation's top health officer, Dr. Theodore Cooper (Assistant Secretary for Health in the Department of Health, Education, and Welfare), stated that

> "While scientists do not yet agree on the specific causal relationships, evidence is mounting and there appears to be general agreement that the kinds and amount of food and beverages we consume and the style of living common in our generally affluent, sedentary society may be the major factors assosociated with the cause of cancer, cardiovascular disease, and other chronic illnesses."

For about 25 years the major culprits in cardiovascular disease have been thought to be saturated fats, cholesterol, and related fat-like substances (lipids). A tremendous campaign has been waged to promote diets with low cholesterol, low saturated fat, and increased polyunsaturated fat. Despite this campaign, the death rate from cardiovascular disease has remained constant during the last 25 years, and it now seems to be almost certain that the assumption that heart disease is caused by a high intake of saturated fats and cholesterol is wrong.

This development does not mean that diet is not important. A high intake of ordinary sugar greatly increases the incidence of cardiovascular disease (see "Sugar: Sweet and Dangerous" in *Executive Health*, Volume 9, Number 1, 1972). Moreover, much evidence has been gathered recently to show that cardiovascular disease can be controlled to a considerable extent by the proper use of vitamin C.

What is cardiovascular disease?

The general term cardiovascular disease comprises various diseases of the heart and blood vessels. Arterio-

THE ONLY THING IT HAS IN COMMON WITH OTHER 7-INCH TAPE DECKS IS THE SIZE OF ITS REELS.

Pioneer's new RT-707 has a lot more in common with today's most sophisticated 10-inch tape decks than it does with most 7-inch tape decks.

Because unlike other 7-inch tape decks, the RT-707 isn't filled with 15 year old ideas.

Take the drive system of the RT-707.

Instead of the old fashioned belt-drive system, the RT-707 is driven by a far more accurate and efficient AC Servo direct-drive motor. This motor generates its own frequency to help correct even the slightest variation in tape speed. Which all but eliminates wow and flutter. And because it doesn't generate heat like the belt-driven "dinosaurs" it doesn't need a fan. So all you'll hear is music with a clarity and crispness not possible on any 7-inch, or many 10-inch tape decks.

Our direct-drive system also makes pitch control possible. To help you regulate the speed of the tape and give you greater control over your recordings.

With technology like this it shouldn't surprise you that our super-sensitive heads will deliver frequencies from 20 to 28,000 Hertz. And our pre-amp section is built to handle 30 decibels more than any other 7-inch tape deck without distorting.

But great sound isn't everything.

As you can see, the RT-707 is smaller and more compact than other tape decks. It's also rack-mountable. And unlike *any* other tape deck, it's stackable. So it'll fit right in with the rest of your components.

But frankly, all the revolutionary thinking that went into the RT-707 wouldn't mean much if it weren't also built to fit comfortably into your budget. It is.

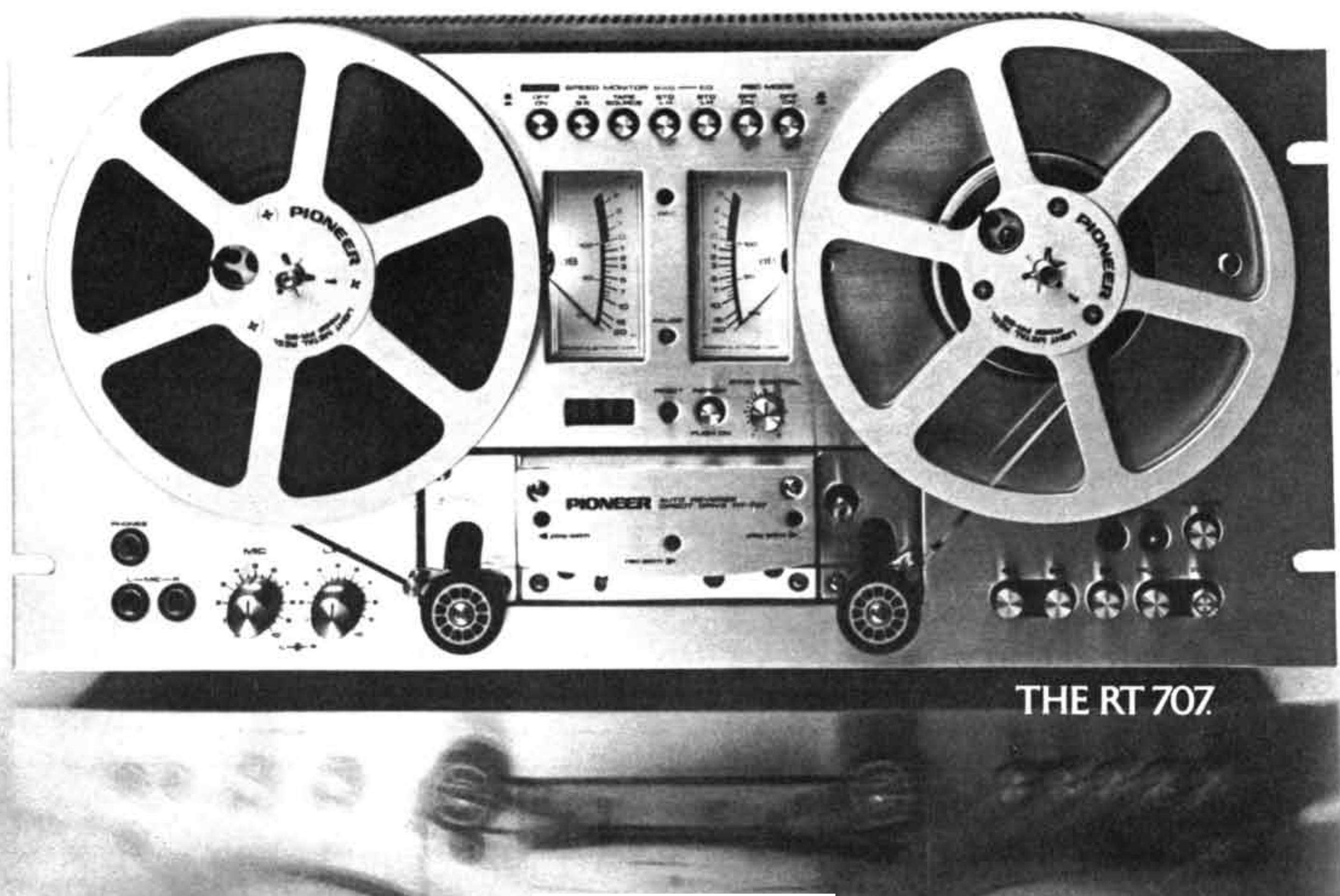
See your Pioneer dealer for a closer look at this extraordinary 7-inch tape deck.

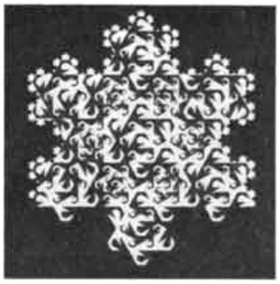
We think you'll find the only things that the RT-707 has in common with other 7-inch tape decks is the size of the reels. And the size of the price.

High Fidelity Components

PIONEER

WE BRING IT BACK ALIVE.





THE COVER

The design on the cover brings together two curves with remarkable mathematical properties (see "Mathematical Games," page 16). The "snowflake" curve that separates the dark outer region from the central motif was discovered in 1904 by the Swedish mathematician Helge von Koch. The most striking property of the snowflake is that it is an infinite curve that bounds a finite area. The boundary between the white and red regions is the third stage in the construction of a curve that was discovered recently by the mathematician Benoît Mandelbrot. It is published here for the first time. At the limit this curve will completely fill the region bounded by the snowflake. Both the snowflake and the Mandelbrot curve are self-similar, that is, if a piece of one of the curves is enlarged, the piece displays the same pattern as the curve. Mandelbrot assigns fractional dimensions to curves of this type and has named them fractal curves.

THE ILLUSTRATIONS

Cover courtesy of Benoît Mandelbrot

Page	Source
18	Ilil Arbel (*top*), W. H. Freeman and Company (*bottom*)
21–28	Richard F. Voss
37	Eleutherian Mills Historical Library
38	*Mimbres Painted Pottery.* © 1977, School of American Research
46–57	Jerome Kuhl
58–59	James Young
60–63	Carol Donner
64	James Young
65	Allen Beechel
66–67	Jack M. Fein
68–72	Guenter Albrecht-Buehler
73	Alan D. Iselin (*top*), Guenter Albrecht-Buehler (*bottom*)
74	*Cell.* © The MIT Press (*top left*); Alan D. Iselin (*top right*); Guenter Albrecht-Buehler (*bottom*)
75	Guenter Albrecht-Buehler (*top*), Alan D. Iselin (*bottom*)
76	Guenter Albrecht-Buehler
91	Lorelle M. Raboni
92	Toshisada Nishida
93–95	Lorelle M. Raboni
96	Lorelle M. Raboni (*top*), Glynn Isaac (*bottom*)
97–102	Lorelle M. Raboni
111	National Geographic Society–Palomar Observatory Sky Survey
112–114	George V. Kelvin
115	National Geographic Society–Palomar Observatory Sky Survey
116	Kitt Peak National Observatory
117	George V. Kelvin
118	Centro Astronomico Hispano–Aleman Almeria
121	De Beers Consolidated Mines Limited
122	Keith G. Cox (*top*), Adolph E. Brotman (*bottom*)
123–128	Adolph E. Brotman
130	Keith G. Cox
132	Adolph E. Brotman
135	Bruce Coleman Incorporated
136–140	Tom Prentiss
145	David K. Lynch (*top*), Robert G. Greenler. © 1977 (*bottom*)
146	David K. Lynch (*top*), Alistair B. Fraser (*bottom*)
147	Austin Hogan
148	A. J. Mallman
149–152	Allen Beechel
154	Michael Goodman
155	Fritz Goro
157–160	Michael Goodman

LETTERS

Sirs:

As George M. Woodwell observes ["The Carbon Dioxide Question," SCIENTIFIC AMERICAN, January], no aspect of policy can remain unaffected by the "greenhouse effect." If our present limited knowledge is any guide, carbon dioxide is a first-order problem, with possible geopolitical consequences on a 25-year time scale. Perhaps some further comments will be useful.

Over the past 117 years the production of carbon dioxide has risen exponentially at 4.25 percent annually, interrupted only by the two world wars and the Great Depression, which slowed our carbon dioxide production by 23 years (perhaps buying us time to reconsider our energy policies). I would emphasize that no smaller-scale disruption of social patterns has had any effect on the growth rate: the advent of nuclear power, the oil embargo and whatever voluntary changes may have occurred have not shown up in the data.

The atmospheric trend can be fitted precisely by an equation with growth parameters taken from the fossil-fuel data and 39 percent of the total carbon dioxide production going into the "ocean." Within the accuracy of the data the contribution from land plants is either zero or proportional to fossil fuel (with a different percentage going into the ocean).

There is no problem in disposing of the "missing" carbon dioxide in the ocean, if one's model is not limited to eddy-diffusion mixing. Biological transport by the diurnal vertical migrations of the "deep scattering layer" organisms could easily transport all of the anthropogenic carbon dioxide across the thermocline. Such a model emphasizes the coupling of carbon dioxide to other nutrients, specifically phosphorus. However, if the coupling is strong, this portion of the carbon dioxide removal from the atmosphere has an upper limit that is constant with time and will not increase as atmospheric carbon dioxide increases. In that case we are headed for trouble faster than any model predicts (because the models assume that removal is proportional to the amount present).

The only possibility I see that offers us a longer respite is for the coupling to be weak, so that carbon (metabolized for energy) and phosphorus (metabolized for growth) are released at different points in an organism's career, and therefore are released at different mean depths in the sea. However, at this point a marine chemist is rapidly getting out of his field of competence. No existing model considers interactions of this subtlety, yet the fact remains that the system is at least this complex.

The potential unpleasantnesses in store, and the fact that the greenhouse is totally man's creation and could be avoided entirely if we chose to avoid it, combine to suggest that the time is ripe for a concerted interdisciplinary attack on carbon dioxide.

FERREN MACINTYRE

Visiting Professor of Oceanography
University of Rhode Island
Kingston

Sirs:

I was interested to read in "The Amateur Scientist" [SCIENTIFIC AMERICAN, December, 1977] Jearl Walker's account of my work on the orientation of Haidinger's brushes, particularly as in many of my own experiments the birefringent wave plates were homemade. Readers who find difficulty in perceiving this visual phenomenon may care to try the effect of first using a blue filter (for example colored acetate) to fatigue the blue receptors of the visual system. When the filter is removed, most subjects see a negative afterimage of Maxwell's spot that is a guide to where the brushes will appear, and when the afterimage fades, the yellow component of the brushes is greatly enhanced. The light blue component is seen as a deeper blue if the green receptors are previously fatigued by exposure to green light rather than to blue light.

The blue component of Haidinger's brushes can be attributed to color contrast developing in a region where there is already a brightness contrast (the yellow component is darker owing to the absorption of blue light) and probably results from inhibitory processes operating in the visual system. The colored shadows referred to by Edwin H. Land in his article on color vision ["The Retinex Theory of Color Vision," SCIENTIFIC AMERICAN, December, 1977] can be regarded as comparable phenomena. When unsaturated colored light casts a shadow onto a white ground, that shadow, if it is sufficiently intense (though still reflecting some light), will assume the complementary color. Thus yellow light will generate violet shadows, orange light blue shadows, red light turquoise shadows and magenta light green shadows. (Conversely, violet, blue, turquoise and green light will respectively generate yellow, orange, red and magenta shadows.) Less intense shadows will appear not as complementary colors but as darker versions of the generating color. Some of these effects were made use of by the Impressionist painters. The complementary shadow colors are reminiscent of the interference colors produced by birefringence with polarized light when a particular wavelength is extinguished. Again one may presume that an inhibitory neural mechanism is at work giving color contrast in regions where there is a brightness contrast, and in this way aiding visual discrimination.

Just as the yellow Haidinger component was seen to be enhanced by previous exposure to the generating color blue, so it is found that adaptation to the various generating colors markedly increases the intensity of shadow colors. The hues of the shadow colors can also be altered by adaptation to some other (nongenerating) colors, in a way that is predictable from the spectral sensitivity of the retinal receptors. Thus the yellow shadow produced by violet-blue light is changed to a green shadow when the subject has previously adapted to red. Both the green shadow produced by magenta light and the turquoise shadow produced by red light are converted to blue shadows after adaptation to green light; this change may be compared with the deepening of the Haidinger-brush blue that results from fatigue of the green receptors.

CHARLES C. D. SHUTE

Professor of Histology
Physiological Laboratory
University of Cambridge
Cambridge, England

Scientific American, April, 1978; Vol. 238, No. 4. 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; Arlene Wright, assistant 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 C. John Kirby, Advertising Director, 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 40 cents.

Subscription correspondence should be addressed to Subscription Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017. For change of address, notify us at least four weeks in advance. Send both old and new addresses and enclose an address imprint from a recent issue. (Date of last issue on your subscription is shown at upper right-hand corner of each month's mailing label.)

Name ____________

New Address ____________

Old Address ____________


1,764 one-gallon gas cans. Enough to reach from the bottom of Royal Gorge, Colorado, to the bridge above. And then some.
Look how much gas you might save with a Chevy Chevette.
Up to 1,764 gallons of gas in five years of normal driving. Imagine saving that much gas. Enough gas to fill 1,764 one-gallon cans. More cans than it takes to reach from the very bottom of the Royal Gorge, Colorado, to the highest suspension bridge in the world.
And here's how to do it. If the car you drive now has an average EPA combined city/highway estimate of 17 miles a gallon, and you switch to a new Chevy Chevette with its EPA combined city/highway estimate of 34 miles a gallon,* in five years of normal driving (12,000 miles per year average) you could save up to 1,764 gallons. But of course that's only an estimate and your gas savings will vary.
Royal Gorge. A demonstration of just *one* advantage of the new Chevy Chevette. It's a lot of car for the money.
***40 mpg hwy/30 mpg city/34 mpg combined.** EPA estimates with 1.6 litre engine and 4-speed manual transmission. California EPA estimates are 36 mpg hwy/27 mpg city/30 mpg combined. Your mileage depends on how you drive, your car's condition, where you drive, and your car's available equipment.
SEE WHAT'S NEW TODAY IN A CHEVROLET.
Chevrolet
GM

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

APRIL, 1928: "The average amount of uranium in the granitic rocks that form the outer layer of the earth's crust can be found with considerable accuracy. It appears to be about seven parts in a million by weight. Its decay must have produced a quantity of lead that must be somewhere in the same rocks. The total content of lead in these rocks, according to the best analyses, averages about 22 parts in a million, but the lead has not the right atomic weight. That it may be a mixture of various 'isotopes' having different atomic weights but the same chemical properties has long been suspected but was first proved by F. W. Aston only a few months ago. With the ingenious 'mass spectrograph,' in which a thin beam of electrically charged atoms is subjected to electric and magnetic fields so proportioned that atoms of each separate atomic weight are directed to a separate point on a photographic plate he has at last, after many vain attempts, attained many definite results with lead which show that there are three kinds of lead atoms with weights 206, 207 and 208. About 30 percent of all the atoms are of the first kind, 20 percent of the second and 50 of the third. Only lead of the first kind can be the product of the decay of uranium. The present amount of this is a little less by weight than that of the uranium itself. A simple calculation shows that it would have been produced by the decay of the uranium in 4,800,000,000 years. If the earth's crust had existed for longer than that, there ought to be more '206-lead' contained in the rocks in proportion to the uranium. We can therefore say, on the basis of this and other evidence, that the earth is more than one and a half billion and less than four billion years old."

"The airship *R-100*, now under construction in England, will be 709 feet long and 133 feet wide at its greatest width. The area of the fabric that will be required to cover the rigid frame is 225,000 square feet for the outer envelope. The framework is to be of novel construction, composed of tubes of duralumin. The passenger accommodations are to be placed within the huge hull, well forward. The plans call for three decks in all, two for the passengers and the third or lower one for the crew. The latter consists of the captain, three commissioned officers, a coxswain, elevator coxswains, riggers, fuel guard, chef and stewards—40 in all. On the passengers' decks are provision for 100 persons. The large dining room in the center deck will accommodate 50 diners at a time. On each side of the passenger compartment, level with the center deck, will be a promenade with specially constructed windows that will afford a perfect view of the countryside or water over which the airship is traveling."

"Concerning the recent demonstration of television given by the General Electric Company, Dr. Lee De Forest, inventor of the three-element vacuum tube, expresses the following opinions: 'I do not think that any marked advance has been made in the Alexanderson television apparatus, except in the synchronization system. I think that television will never be practical in the home, due to the fact that the present methods require large rotating parts operated by a motor. The difficulty is that the operator at the receiving end must constantly regulate a little knob or dial to prevent the picture from becoming distorted. Automatic synchronization, which would overcome this obstacle, is mighty expensive, and I doubt that it will ever be used in the home. We are still a million miles away from the application of television on a large theater screen, because 18 inches today constitutes approximately the largest television screen in use. A new system must be developed, based on another branch of physics, which will get away from heavy and rotating parts before seeing by radio can be made practical for private use.'"

SCIENTIFIC AMERICAN

APRIL, 1878: "What is life? Perhaps the closest definition that science has yet arrived at is that it is the cell's impulse of organization. Whence arises that impulse? Is it a special, extra-material impulse? Or is it only a mode of action of the general force of nature? Are the mysterious properties of the germ the result of molecular combination, as the properties of water arise from the combination of its constituent gases? Here the final battle of biology must be fought. So long as life is surely known to proceed only from antecedent life, just so long will it be impossible to give a decisive answer to the question, What is life? The mystery of life lies in the evolutive power of the germ. If life is a vital spark handed down from organism to organism from the beginning, then it transcends the ken of physical science and must ever remain a mystery. If, on the other hand, life can be proved to begin *de novo* in suitable mixtures of demonstrably dead matter, at that moment life ceases to be the only unique phenomenon in nature, and takes rank among the powers and potencies of ordinary matter. No wonder the controversy assumes at times a bitterness foreign to purely scientific discussions. The issues at stake are of transcendent importance, for upon the supernatural nature and origin of life hang the most revered beliefs and the most momentous theories of the age."

"The study of the solar spectrum and other luminous spectra has been greatly advanced by the intervention of photography, which has been the means of recognizing dark lines or spaces in the ultra violet region, the rays of which produce scarcely any impression on the retina. A large number of such lines have been thus determined by Rutherford, Draper and Mascart. Similarly Vogel has made some new discoveries with regard to the obscure rays in the red region. He has found that it is sufficient to mix with collodion coloring matters that absorb the red rays to render it sensitive to the action of such rays, so that the special designation of 'chemical rays' applied to those of the violet and ultra violet may be considered as obsolete, all the spectral colors being capable of affecting a photographic plate properly prepared."

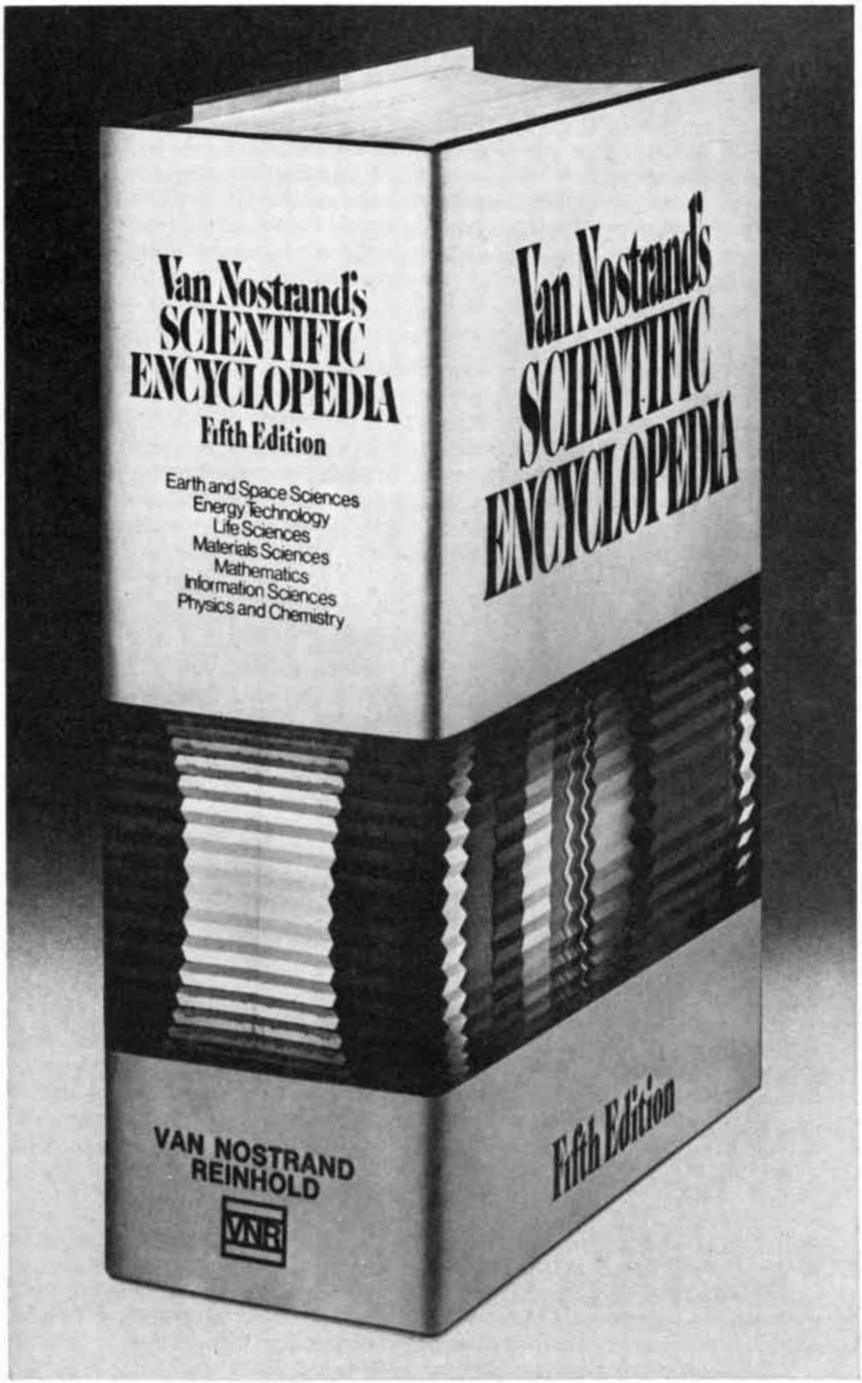
"A reporter of the *World* recently called on Professor Edison at his laboratory in Menlo Park, N.J. After greetings had been exchanged, Mr. Edison said, 'Come downstairs, and I'll show you the new model of the phonograph.' The reporter followed Mr. Edison into the workroom, where about a dozen lathes and machines were in full operation. On one of the tables was the model. The improvement in the phonograph consists in a circular plate being substituted for the cylinder, and clockwork for the crank. 'This clock-movement is a very important improvement,' said Mr. Edison. 'It insures complete regularity and accuracy, and can be thrown out of and into gear instantaneously. We're going to start a publication office in New York when the phonograph is ready.'

'What do you intend to publish?' asked the reporter.

'Music, novels, general literature and many other kinds of matter that are read by persons and reproduced by instruments or their vocal organs for the benefit of themselves and other persons. Take music to begin with. We will phonograph orchestral concerts by brass and string bands, instrumental and vocal solos and part songs. The sheets bearing the sound impressions of this music will be removed from the phonograph and multiplied to any extent by electrotyping, and persons can make selections of any compositions they desire. Then this music may be reproduced by any phonograph, with all the original sweetness and expression, and not only that, but the pitch can be raised or lowered by increasing or diminishing the speed of the phonograph.'"

The long-awaited new edition of a classic.

VAN NOSTRAND'S SCIENTIFIC ENCYCLOPEDIA, completely revised and updated after nearly a decade, is now offered as an introduction to membership in the BOOK-OF-THE-MONTH CLUB®



BOOK-OF-THE-MONTH CLUB, INC., Camp Hill, Pennsylvania 17012

YOURS FOR ONLY

$10.00

PUBLISHER'S PRICE: $67.50
(Save more than 80%)

For four decades *Van Nostrand's Scientific Encyclopedia* has been acknowledged as the most comprehensive one-volume reference of its kind. Now extensively revised—80% of the text and 90% of the illustrations are new—the Fifth Edition is an indispensable source of modern scientific information for professionals and non-professionals alike.

Written with remarkable clarity, the work spans the whole science spectrum—from computer technology and planetary exploration to systems analysis and oceanography—including all the important advances and discoveries of the past decade in every category.

Features of the new Fifth Edition:

- Actual size: 9¼″ by 12″
- 2382 pages
- Over 2500 black-and-white and color photographs, drawings and charts
- More than 2.2 million words
- 7200 articles
- 550 tables
- 190 contributing editors
- Clear, easy-to-read type
- 8000 cross-references in boldface
- Sturdy, reinforced binding

Join the Book-of-the-Month Club now—and get this singular equivalent of a multi-volume science library. As a Club member, you'll continue to enjoy benefits on the best and most important books published today.

Bookstore Quality at Book Club Savings. You conveniently shop at home at considerable savings. *Example:* if you took *Van Nostrand's Scientific Encyclopedia* for $10.00, bought six books and two Book-Dividends, you could save an average of $74—*including* postage and handling. And these are true savings, because every book is exactly like the one sold in good bookstores—the same size, paper, type and binding. You *never* receive cheap book-club editions, altered in size and with inferior paper and bindings.

Book-Dividends. When you remain a Club member after the trial period, every book you buy earns Book-Dividend® credits. These entitle you to choose from a wide variety of significant books at hard-to-believe savings of *at least 70%*.

THE AUTHORS

DAVID J. ROSE and RICHARD K. LESTER ("Nuclear Power, Nuclear Weapons and International Stability") share an interest in the security problems relating to the international supply of nuclear fuel. Rose is professor of nuclear engineering at the Massachusetts Institute of Technology. A native of Canada, he obtained his bachelor's degree in engineering physics from the University of British Columbia and his Ph.D. in physics from M.I.T. in 1950. After seven years at Bell Laboratories he joined the M.I.T. faculty in 1958. Since then his research has encompassed plasma dynamics, controlled nuclear fusion and national energy policy. From 1969 to 1971 Rose served as director of long-range planning at the Oak Ridge National Laboratory, and he is currently a member of the steering committee of the National Academy of Sciences Committee on Nuclear and Alternative Energy Systems. Lester is visiting research fellow in the International Relations Division of the Rockefeller Foundation. Born in England, he obtained his bachelor's degree in chemical engineering from the Imperial College of Science and Technology in 1974, and then came to M.I.T. as a Kennedy Scholar in the department of nuclear engineering. He has been on leave studying nuclear policy issues at the Rockefeller Foundation. Lester recently coauthored a book with Mason Willrich: *Radioactive Waste: Management and Regulation* (Free Press, 1977).

JACK M. FEIN ("Microvascular Surgery for Stroke") is attending neurosurgeon at the Albert Einstein College Hospital in New York and assistant professor of neurosurgery at the Albert Einstein College of Medicine. He attended Yeshiva University and the New York University School of Medicine, where he received his M.D. in 1965. After an internship and residency at Kings County Hospital and Medical Center in Brooklyn he did research on the arterial circulation of the brain at the New York University Medical Center and the National Naval Medical Center in Bethesda, Md., and served as chief of the neurosciences division of the Armed Forces Radiobiology Research Institute. He then practiced and taught neurosurgery for a year at the Georgetown University Medical Center before coming to Albert Einstein in 1974.

GUENTER ALBRECHT-BUEHLER ("The Tracks of Moving Cells") is head of the cell biology section at the Cold Spring Harbor Laboratory on Long Island. Born in Berlin, he studied physics at the Technical University in Munich and worked for three years on the physical properties of excitable membranes in the laboratory of H. Mueller-Mohnssen. He then went on to graduate school in Munich and obtained his Ph.D. in physics in 1971 with a thesis on the application of optics to the study of cell membranes. After three years at the Friedrich Miescher Institute in Basel, where he began to study cell motility, he obtained a fellowship at the University of Florida at Gainesville. Six months later, in 1974, he was invited to the Cold Spring Harbor Laboratory to run the laboratory's new scanning electron microscope. In his free time Albrecht-Buehler often plays the ancient oriental board game Go.

GLYNN ISAAC ("The Food-sharing Behavior of Protohuman Hominids") is professor of anthropology at the University of California at Berkeley, where he teaches both archaeology and human evolution. A native of South Africa, he studied zoology and geology at the University of Cape Town and then attended the University of Cambridge, obtaining his Ph.D. in archaeology in 1968. The final stages of the research reported in his article were supported by a Guggenheim Fellowship and a visiting fellowship at Peterhouse, Cambridge, in 1976. He writes: "Since the age of 12 I have been interested in the study of prehistory as the meeting ground of history and natural history, and have wanted to use archaeology to further our understanding of human evolution. So when in 1961 Louis Leakey offered me and my wife Barbara the chance to join in the work being done in East Africa, we jumped at it. Since 1970 we have been engaged with Richard Leakey in the careful long-term study of one of the world's prime repositories of evidence about early human ways of life."

MICHAEL ZEILIK ("The Birth of Massive Stars") is assistant professor of astronomy at the University of New Mexico. He did his undergraduate work in physics at Princeton University and obtained his master's degree in astronomy from Harvard University in 1969. After teaching for three years at Southern Connecticut State College he continued his study of astronomy at the Center for Astrophysics of the Harvard College Observatory and the Smithsonian Astrophysical Observatory, where he obtained his Ph.D. in 1975. At the University of New Mexico, Zeilik does research on star formation, observing at radio and infrared wavelengths to uncover regions of recent star birth in our galaxy. He is also interested in the teaching of astronomy to nonspecialists by the Personalized System of Instruction (PSI) and has written an introductory textbook in the PSI format titled *Astronomy: The Evolving Universe* (Harper & Row, 1976).

KEITH G. COX ("Kimberlite Pipes") is university lecturer in the department of geology and mineralogy at the University of Oxford. He studied for his bachelor's degree at Oxford, and then in 1956 became a research student in the Research Institute of African Geology at the University of Leeds. An interest in igneous geology led him to do fieldwork in the Karroo volcanic province of South Africa. He writes: "During the course of these studies I visited Lesotho to see the Karroo basalts there and, quite by accident, was directed by a local diamond prospector to the Matsoku kimberlite pipe, at that time unknown to science. I then took up the study of kimberlites as a scientific 'hobby' to accompany my main interest in basaltic rocks, and in 1963 I led an expedition to Matsoku." After leaving Leeds, Cox spent seven years lecturing at the University of Edinburgh before returning to Oxford. He is the coauthor of two textbooks on igneous petrology, one introductory and one advanced, and is managing editor of *Journal of Petrology*.

LORUS J. MILNE and MARGERY MILNE ("Insects of the Water Surface"), husband and wife, are behavioral ecologists at the University of New Hampshire. They received their Ph.D. in biology respectively from Harvard University and Radcliffe College. They have recently become interested in the surface film of lakes and streams, which provides a habitat for a variety of superbly adapted insect forms. They write: "For the insects that live at the transparent boundary between the wet world and the dry, the water molecules provide a flexible support, a banquet table, a communication network and a resilient bed upon which to make love." This is the ninth article by the Milnes to appear in *Scientific American*.

DAVID K. LYNCH ("Atmospheric Halos") is visiting associate professor of physics at the California Institute of Technology and assistant research astronomer at the University of California at Berkeley. He received his bachelor's degree in astrophysics from Indiana University and his Ph.D. in astronomy from the University of Texas at Austin in 1975. In the course of his career as a graduate student he worked at the McDonald Observatory, the Sacramento Peak Observatory and the Aerospace Solar Observatory. Lynch's research interests include the properties of the interstellar medium, solar astronomy and meteorological optics, which provides the theme of his article.

You're looking at what our competitors don't want you to see.

Stereo Receivers	Sugg. Ret. Price*	Min. RMS Power Per Channel into 8 Ohms	Total Harmonic Distortion at Rated Power (Max.)	FM Sensitivity IHF '58	FM Sensitivity Stereo—50dB†
SA-5770	$799.95	165 watts from 20Hz-20kHz	0.08%	1.8μV	35.7dBf
SA-5570	499.95	85 watts from 20Hz-20kHz	0.1	1.8μV	36.2dBf
SA-5470	399.95	65 watts from 20Hz-20kHz	0.1	1.8μV	36.2dBf
SA-5370	329.95	48 watts from 20Hz-20kHz	0.1	1.9μV	37.2dBf
SA-5270	279.95	35 watts from 20Hz-20kHz	0.3	1.9μV	37.2dBf
SA-5170	229.95	25 watts from 20Hz-20kHz	0.5	1.9μV	37.2dBf
SA-5070	179.95	15 watts from 40Hz-20kHz	0.8	2.0μV	37.3dBf

*Technics recommended price, but actual retail price will be set by dealers.

†New IHF '75 standard

PPR. The price performance relationship of Technics new receiver line led by the SA-5770 (shown below). PPR is our way of telling you how much performance, technology and power you're getting for your money. And it may be the reason why so many people are buying and recommending Technics receivers.

When you listen to our new receivers, you'll hear what PPR means to your ears: The big, wide-band power output of our direct-coupled amplifiers. Made possible by the current handling capacity of our large transformers and conservatively rated power supply capacitors. And the hidden dynamic range you'll discover in your records, up to 78 dB S/N. Thanks to load-resistant 3-stage IC's in the phono equalizer sections.

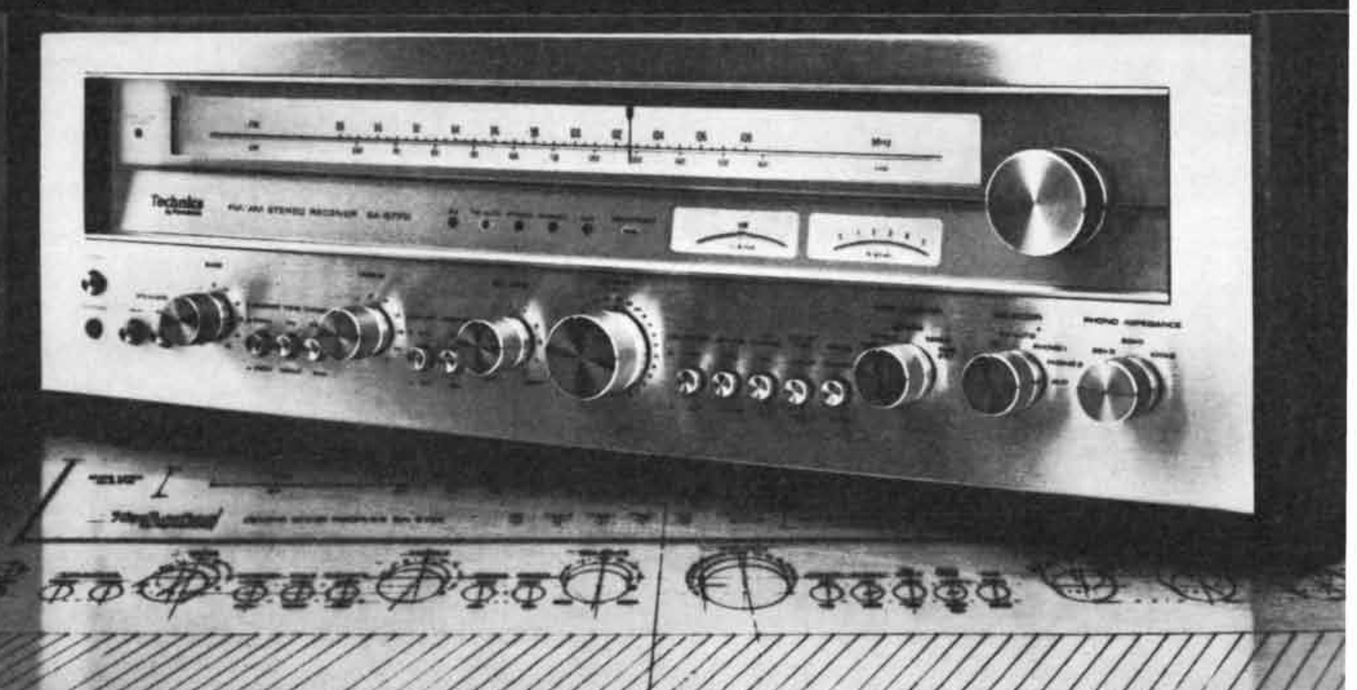
If you want to hear clear FM. From an over-crowded band or a marginal signal. You can. And with outstanding separation. Because every Technics receiver boasts Phase Locked Loop IC's, flat group delay filters and an FM linear dial scale.

What you won't hear is annoying distortion. Because it's so low, it's virtually inaudible, even in our economy model.

Technics new receivers. Judge them on performance. But buy them on PPR.

All cabinetry is simulated wood

Technics
by Panasonic



MATHEMATICAL GAMES

White and brown music, fractal curves and one-over-f fluctuations

by Martin Gardner

"For when there are no words [accompanying music] it is very difficult to recognize the meaning of the harmony and rhythm, or to see that any worthy object is imitated by them."

—Plato, *Laws*, Book II

Plato and Aristotle agreed that in some fashion all the fine arts, including music, "imitate" nature, and from their day until the late 18th century "imitation" was a central concept in western aesthetics. It is obvious how representational painting and sculpture "represent," and how fiction and the stage copy life, but in what sense does music imitate?

By the mid-18th century philosophers and critics were still arguing over exactly how the arts imitate and whether the term is relevant to music. The rhythms of music may be said to imitate such natural rhythms as heartbeats, walking, running, flapping wings, waving fins, water waves, the periodic motions of heavenly bodies and so on, but this does not explain why we enjoy music more than, say, the sound of cicadas or the ticking of clocks. Musical pleasure derives mainly from tone patterns, and nature, though noisy, is singularly devoid of tones. Occasionally wind blows over some object to produce a tone, cats howl, birds warble, bowstrings twang. A Greek legend tells how Hermes invented the lyre: he found a turtle shell with tendons attached to it that produced musical tones when they were plucked.

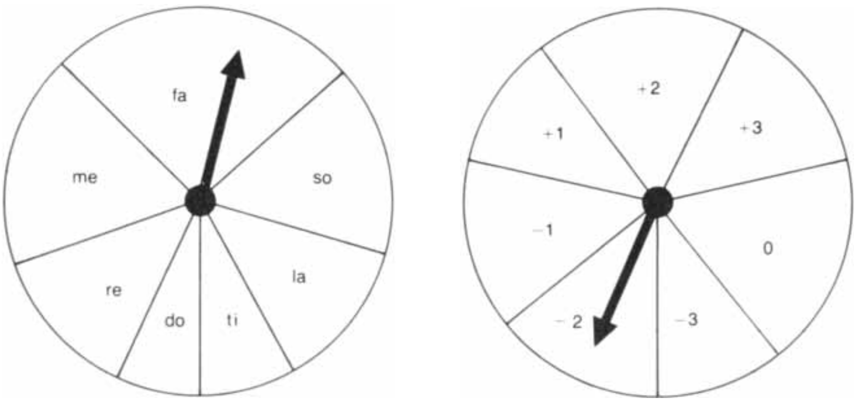
Above all, human beings sing. Musical instruments may be said to imitate song, but what does singing imitate? A sad, happy, angry or serene song somehow resembles sadness, joy, anger or serenity, but if a melody has no words and invokes no special mood, what does it copy? It is easy to understand Plato's mystification.

There is one exception: the kind of imitation that plays a role in "program music." A lyre is severely limited in the natural sounds it can copy, but such limitations do not apply to symphonic or electronic music. Program music has no difficulty featuring the sounds of thunder, wind, rain, fire, ocean waves and brook murmurings; bird calls (cuckoos and crowing cocks have been particularly popular), frog croaks, the gaits of animals (the thundering hoofbeats in Wagner's *Ride of the Valkyries*), the flights of bumblebees; the rolling of trains, the clang of hammers; the battle sounds of marching soldiers, clashing armies, roaring cannons and exploding bombs. *Slaughter on Tenth Avenue* includes a pistol shot and the wail of a police-car siren. In Bach's *Saint Matthew Passion* we hear the earthquake and the ripping of the temple veil. In the *Alpine Symphony* by Richard Strauss cowbells are imitated by the shaking of cowbells. Strauss insisted he could tell that a certain female character in Felix Mottl's *Don Juan* had red hair, and he once said that someday music would be able to distinguish the clattering of spoons from that of forks.

Such imitative noises are surely a trivial aspect of music even when it accompanies opera, ballet or the cinema; besides, such sounds play no role whatever in "absolute music," music not intended to "mean" anything. A Platonist might argue that abstract music imitates emotions, or beauty, or the divine harmony of God or the gods, but on more mundane levels music is the least imitative of the arts. Even nonobjective paintings resemble certain patterns of nature, but nonobjective music resembles nothing except itself.

Since the turn of the century most critics have agreed that "imitation" has been given so many meanings (almost all are found in Plato) that it has become a useless synonym for "resemblance." When it is made precise with reference to literature or the visual arts, its meaning is obvious and trivial. When it is applied to music, its meaning is too fuzzy to be helpful. This month we take a look at a surprising discovery by Richard F. Voss, a young physicist from Minnesota who joined the Thomas J. Watson Research Center of the International Business Machines Corporation after obtaining his Ph.D. at the University of California at Berkeley under the guidance of John Clarke. This work is not likely to restore "imitation" to the lexicon of musical criticism, but it does suggest a curious way in which good music may mirror a subtle statistical property of the world.

The key concepts behind Voss's discovery are what mathematicians and physicists call the spectral density (or power spectrum) of a fluctuating quantity, and its "autocorrelation." These deep notions are technical and hard to understand. Benoît Mandelbrot, who is also at the Watson Research Center, and whose recent work makes extensive use of spectral densities and autocorrelation functions, has suggested a way of avoiding them here. Let the tape of a sound be played faster or slower than normal. One expects the character of the sound to change considerably. A violin, for example, no longer sounds like a violin. There is a special class of sounds, however, that behave quite differently. If you play a recording of such a sound at a different speed, you have only to adjust the volume to make it sound exactly as before. Mandelbrot calls such sounds "scaling noises."



Spinners for white music (left) *and brown music* (right)

Jameson Irish "~~Scotch~~ on the rocks."

If you like Scotch, you'll love light Jameson Irish.

Try a glass of Jameson Irish the way you would your favorite Scotch.

You'll notice how much it tastes like fine Scotch—only lighter and more delicate. Not smoky tasting like Scotch.

The dedicated Scotch drinker will instantly appreciate this flavor difference.

Though it may take a little time getting used to saying, "Jameson Irish on the rocks, please."

Jameson. World's largest-selling Irish Whiskey.

80 & 86 PROOF • CALVERT DIST. CO., N.Y.C.

By far the simplest example of a scaling noise is what in electronics and information theory is called white noise (or "Johnson noise"). To be white is to be colorless. White noise is a colorless hiss that is just as dull whether you play it faster or slower. Its autocorrelation function, which measures how its fluctuations at any moment are related to previous fluctuations, is zero. The most commonly encountered white noise is the thermal noise produced by the random motions of electrons through an electrical resistance. It causes most of the static in a radio or amplifier and the "snow" on radar and television screens when there is no input.

With randomizers such as dice or spinners it is easy to generate white noise that can then be used for composing a random "white tune," one with no correlation between any two notes. Our scale will be one octave of seven white keys on a piano: do, re, me, fa, so, la, ti. Fa is our middle frequency. Now construct a spinner such as the one shown at the left in the illustration on page 16. Divide the circle into seven sectors and label them with the notes. It matters not at all what arc lengths are assigned to these sectors; they can be completely arbitrary. On the spinner shown some order has been imposed by giving fa the longest arc (the highest probability of being chosen) and assigning decreasing probabilities to pairs of notes that are equal distances above and below fa. This has the effect of clustering the tones around fa.

To produce a "white melody" simply spin the spinner as often as you like, recording each chosen note. Since no tone is related in any way to the sequence of notes that precedes it, the result is a totally uncorrelated sequence. If you like, you can divide the circle into more parts and let the spinner select notes that range over the entire piano keyboard, black keys as well as white.

To make your white melody more sophisticated, use another spinner, its circle divided into four parts (with any proportions you like) and labeled 1, 1/2, 1/4 and 1/8 so that you can assign a full, a half, a quarter or an eighth of a beat to each tone. After the composition is completed, tap it out on the piano. The music will sound just like what it is: random music of the dull kind that a two-year-old or a monkey might produce by hitting keys with one finger.

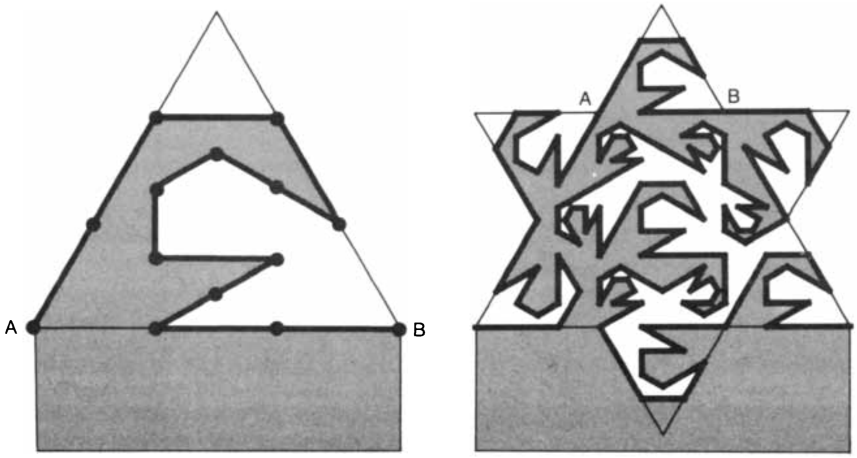
A more complicated kind of scaling noise is one that is sometimes called Brownian noise because it is characteristic of Brownian motion, the random movements of small particles suspended in a liquid and buffeted by the thermal agitation of molecules. Each particle executes a three-dimensional "random walk," the positions in which form a highly correlated sequence. The particle, so to speak, always "remembers" where it has been.

When tones fluctuate in this fashion, let us follow Voss and call it Brownian music or brown music. We can produce it easily with a spinner and a circle divided into seven parts as before, but now we label the regions, as is shown at the right in the illustration on page 16, to represent intervals between successive tones. These step sizes and their probabilities can be whatever we like. On the spinner shown plus means a step up the scale of one, two or three notes and minus means a step down of the same intervals.

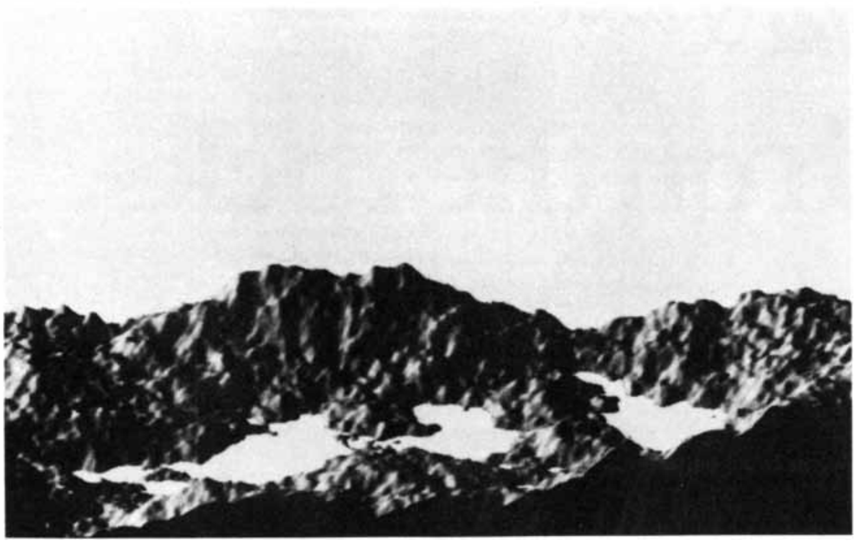
Start the melody on the piano's middle C, then use the spinner to generate a linear random walk up and down the keyboard. The tune will wander here and there, and will eventually wander off the keyboard. If we treat the ends of the keyboard as "absorbing barriers," the tune ends when we encounter one of them. We need not go into the ways in which we can treat the barriers as reflecting barriers, allowing the tune to bounce back, or as elastic barriers. To make the barriers elastic we must add rules so that the farther the tone gets from middle C, the greater is the likelihood it will step back toward C, like a marble wobbling from side to side as it rolls down a curved trough.

As before, we can make our brown music more sophisticated by varying the tone durations. If we like, we can do this in a brown way by using another spinner to give not the duration but the increase or decrease of the duration—another random walk but one along a different street. The result is a tune that sounds quite different from a white tune because it is strongly correlated, but a tune that still has little aesthetic appeal. It simply wanders up and down like a drunk weaving through an alley, never producing anything that resembles good music.

If we want to mediate between the ex-



The first two steps in constructing Benoît Mandelbrot's Peano-snowflake curve



Brownian landscape generated by a computer program written by Richard F. Voss

Contax:
A Step Beyond Nikon.

Your local Contax dealer is waiting to show you stunning audio/visual evidence of a precision instrument that steps beyond Nikon: Contax.

Go to your local Contax dealer. See his audio/visual presentation entitled *Contax: A Step Beyond*. Ask to hold a Contax RTS and a Nikon F2AS. Then compare.

Contax doesn't look like a Nikon.

Nikon is one of the most recognized cameras in America. Truly a fine camera.

Contax is one of the most recognized cameras in Europe and Asia. That's because the Contax RTS is the result of the advanced technology of German optics by Carl Zeiss, and high state-of-the-art Japanese electronics.

Both Contax and Nikon are recognized by their professional looks. But, *only a Contax looks like a Contax*.

Contax doesn't sound like a Nikon.

Nikon's shutter works mechanically. You hear a mechanical sound. Contax uses a sophisticated electromagnetic release system.

The unique sound is the electronic contact of a Contax micro-switch activated by a feather touch.

The Contax electromagnetic shutter you hear is so precise you can eliminate "movement worry" at shutter release. You can breathe a little easier with Contax. *Only a Contax sounds like a Contax*.

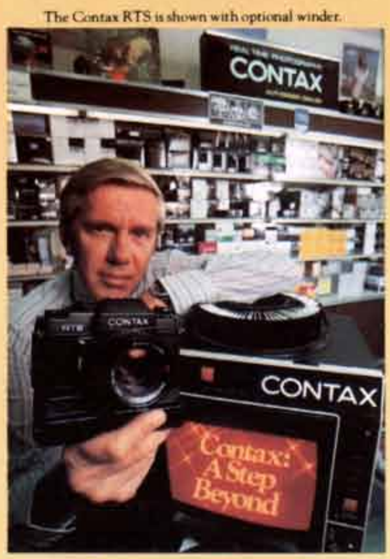
Contax doesn't feel like a Nikon.

Hold a Contax RTS in one hand. The Nikon F2AS in the other. Contax is smaller. Lighter. A comfortable compromise between compact and standard 35mm size.

Nikon is a traditional design. The Contax body is modern and sleek. Controls are molded closer to the body. Fingers have natural access to them.

The design is faultless. It should be. The Porsche Design Group labored over a year in its creation. *Only a Contax feels like a Contax.*

The Contax RTS is shown with optional winder.



Contax doesn't operate like a Nikon.

The Contax RTS is fully automatic. The Nikon F2AS is not.

Shooting the Contax RTS is as simple as *focus* and *touch*. Two steps. The Nikon F2AS (without the addition of another accessory) requires a manual matching step before you focus and push. Three steps.

Contax has 16-dot LED shutter speed display that automatically shows shutter speed selection. The F2AS does not.

Contax is one of the fastest and simplest operating cameras ever developed. *Only a Contax operates like a Contax.*

See *Contax: A Step Beyond* at your local Contax dealer.

Contax must be seen . . . heard . . . felt . . . and operated. Your dealer can show you a stunning audio/visual presentation on this remarkable precision instrument. Ask to compare the Contax RTS with Nikon, or any other sophisticated camera system.

Then, step beyond.

Call the Contax toll-free hotline for your nearest Contax dealer, and audio/visual show.

(800) 327-5351

In Florida call (800) 432-8407 toll free.

Contax/411 Sette Drive
Paramus, New Jersey 07652

CONTAX RTS
Real Time System

tremes of white and brown, we can do it in two essentially different ways. The way chosen by previous composers of "stochastic music" is to adopt transition rules. These are rules that select each note on the basis of the last three or four. For example, one can analyze Bach's music and determine how often a certain note follows, say, a certain triplet of preceding notes. The random selection of each note is then weighted with probabilities derived from a statistical analysis of all Bach quadruplets. If there are certain transitions that never appear in Bach's music, we add rejection rules to prevent the undesirable transitions. The result is stochastic music that resembles Bach but only superficially. It sounds Bachlike in the short run but random in the long run. Consider the melody over periods of four or five notes and the tones are strongly correlated. Compare a run of five notes with another five-note run later on and you are back to white noise. One run has no correlation with the other. Almost all stochastic music produced so far has been of this sort. It sounds musical if you listen to any small part but random and uninteresting when you try to grasp the pattern as a whole.

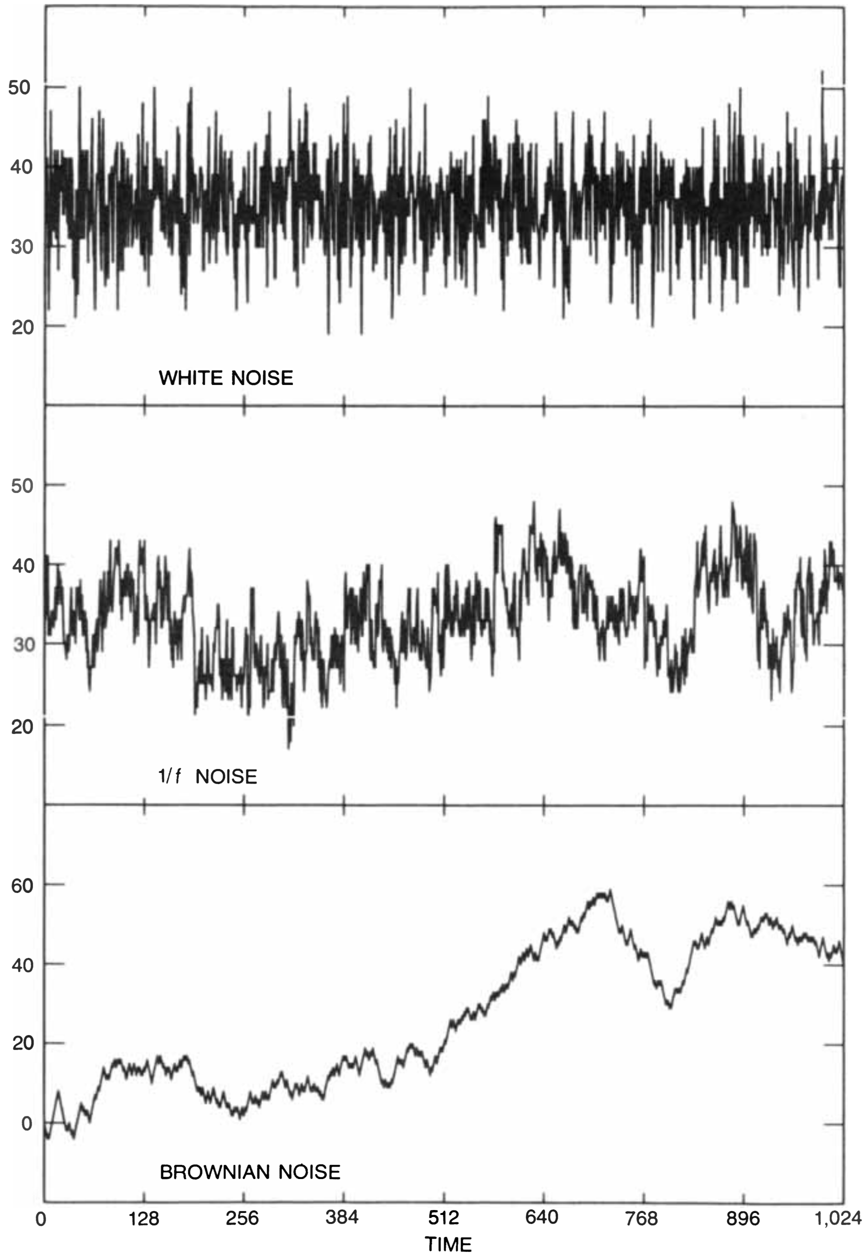
Voss's insight was to compromise between white and brown input by selecting a scaling noise exactly halfway between. In spectral terminology it is called $1/f$ noise. (White noise has a spectral density of $1/f^0$, brownian noise a spectral density of $1/f^2$. In "one-over-f" noise the exponent of f is 1 or very close to 1.) Tunes based on $1/f$ noise are moderately correlated, not just over short runs but throughout runs of any size. It turns out that almost every listener agrees that such music is much more pleasing than white or brown music.

In electronics $1/f$ noise is well known but poorly understood. It is sometimes called flicker noise. Mandelbrot, whose book *Fractals: Form, Chance and Dimension* (W. H. Freeman and Company, 1977) has already become a modern classic, was the first to recognize how widespread $1/f$ noise is in nature, outside of physics, and how often one encounters other scaling fluctuations. For example, he discovered that the record of the annual flood levels of the Nile is a $1/f$ fluctuation. He also investigated how the curves that graph such fluctuations are related to "fractals," a term that he invented. A scaling fractal can be defined roughly as any geometrical pattern (other than Euclidean lines, planes and surfaces) with the remarkable property that no matter how closely you inspect it it always looks the same, just as a slowed or speeded scaling noise always sounds the same. Mandelbrot coined the term fractal because he assigns to each of the curves a fractional dimension greater than its topological dimension.

Among the fractals that exhibit strong regularity the best-known are the Peano curves that completely fill a finite region and the beautiful snowflake curve discovered by the Swedish mathematician Helge von Koch in 1904. The Koch snowflake appears on the cover of this issue as the boundary of the dark "sea" that surrounds the central motif. (For details on the snowflake's construction, see this department for December, 1976.)

The most interesting part of the cover is the fractal curve that forms the central design. It was discovered quite recently by Mandelbrot, who has allowed *Scientific American* to publish it here for the first time. If you trace the boundary between the red and white regions from the tip of the point of the star at the lower left to the tip of the point of the star at the lower right, you will find this boundary to be a single curve. It is the third stage in the construction of a new Peano curve. At the limit this lovely curve will completely fill a region bounded by the traditional snowflake! Thus Mandelbrot's curve brings together two pathbreaking fractals: the oldest of them all, Giuseppe Peano's 1890 curve, and Koch's later snowflake.

The secret of the curve's construction is the use of line segments of two unequal lengths and oriented in 12 different directions. The curve is much less regular than previous Peano curves and therefore closer to the modeling of natu-



Typical patterns of white, 1/f and Brownian noise ral phenomena, the central theme of Mandelbrot's book. One can see in the pattern such natural forms as the gnarled branches of a tree or the shapes of flickering flames.

At the left in the upper illustration on page 18 is the first step of the construction. A crooked line of nine segments is drawn on and within an equilateral triangle. Four of the segments are then divided into two equal parts, creating a line from A to B that consists of 13 long and short segments. The second step replaces each of these 13 segments with a smaller replica of the crooked line. These replicas (necessarily of unequal size) are oriented as is shown inside the star at the right in the illustration. A third repetition of the procedure generates the curve on the cover. (It belongs to a family of curves arising from William Gosper's discovery of the "flowsnake," a fractal pictured in the column cited above and in Mandelbrot's book.) When the construction is repeated to infinity, the limit is a Peano curve that totally fills a region bordered by the Koch snowflake. The Peano curve has the usual dimension of 2, but its border, a scaling fractal of infinite length, has (as is explained in Mandelbrot's book) a fractal dimension of log 4/log 3, or 1.2618. . . .

Unlike these striking artificial curves the fractals that occur in nature—coastlines, rivers, trees, star clustering, clouds and so on—are so irregular that their self-similarity (scaling) must be treated statistically. Consider the profile of the mountain range in the lower illustration on page 18, reproduced from Mandelbrot's book. This is not a photograph. It is a computer-generated mountain scene based on a modified Brownian noise. Any vertical cross section of the topography has a profile that models a random walk. The white patches, representing water or snow in the hollows below a certain altitude, were added to enhance the relief.

	↓ BLUE	↓ GREEN	↓ RED
0	0	0	0
1	0	0	1
2	0	1	0
3	0	1	1
4	1	0	0
5	1	0	1
6	1	1	0
7	1	1	1

Binary chart for Voss's 1/f dice algorithm

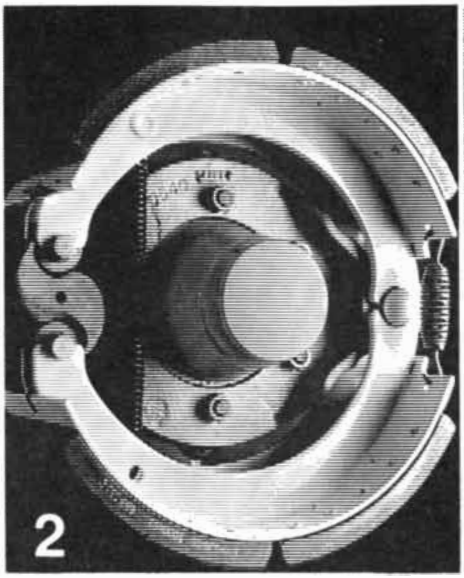
The profile at the top of the mountain range is a scaling fractal. This means that if you enlarge any small portion of it, it will have the same statistical character as the line you now see. If it were a true fractal, this property would continue forever as smaller and smaller segments are enlarged, but of course such a curve can neither be drawn nor appear in nature. A coastline, for example, may be self-similar when viewed from a height of several miles down to several feet, but below that the fractal property is lost. Even the Brownian motion of a particle is limited by the size of its microsteps.

Since mountain ranges approximate random walks, one can create "mountain music" by photographing a mountain range and translating its fluctuating heights to tones that fluctuate in time. If we view nature statically, frozen in time, we can find thousands of natural curves that can be used in this way to produce stochastic music. Such music is usually too brown, too correlated, however, to be interesting. Like natural white noise, natural brown noise may do well enough, perhaps, for the patterns of abstract art but not so well as a basis for music.

When we analyze the dynamic world, made up of quantities constantly changing in time, we find a wealth of fractal-like fluctuations that have $1/f$ spectral densities. In his book Mandelbrot cites a few: variations in sunspots, the wobbling of the earth's axis, undersea currents, membrane currents in the nervous system of animals, the fluctuating levels of rivers and so on. Uncertainties in time measured by an atomic clock are $1/f$: the error is 10^{-12} regardless of whether one is measuring an error on a second, minute or hour. Scientists tend to overlook $1/f$ noises because there are no good theories to account for them, but there is scarcely an aspect of nature in which they cannot be found.

T. Musha, a physicist at the Tokyo Institute of Technology, recently discovered that traffic flow past a certain spot on a Japanese expressway exhibited $1/f$ fluctuation. In a more startling experiment, not yet published, Musha rotated a radar beam emanating from a coastal location to get a maximum variety of landscape on the radar screen. When he rotated the beam once, variations in the distances of all objects scanned by the beam produced a Brownian spectrum. But when he rotated it twice and then subtracted one curve from the other the resulting curve—representing all the changes of the scene—was close to $1/f$.

We are now approaching an understanding of Voss's daring conjecture. The changing landscape of the world (or



Eaton Update:

1 Getting the power around the corner

Eaton hydrostatic transmissions offer designers of agricultural equipment great flexibility by replacing awkward mechanical linkages with hydraulic lines that can go around or through other components. They allow the equipment to work more efficiently because operators have an infinite selection of operating modes.

Eaton's commitment to hydrostatics started in the mid-sixties. A new plant in Spencer, Iowa, is operating at maximum capacity, and we are planning another expansion to serve this growing market.

2 The advantages of being single

With Eaton's Single Anchor Pin brake system you can reline a truck brake in just two minutes—and the only tool you need is a screwdriver. With conventional double anchor pin systems the job can take hours, even with special tools.

This Eaton exclusive is saving down-time for truckers all over the country. But it's only one of our many contributions to trucking efficiency. We're a leader, worldwide, in axles and heavy-duty transmissions, and we're winning an ever-stronger position in components like brakes and anti-lock systems, fan drives, and air conditioning.

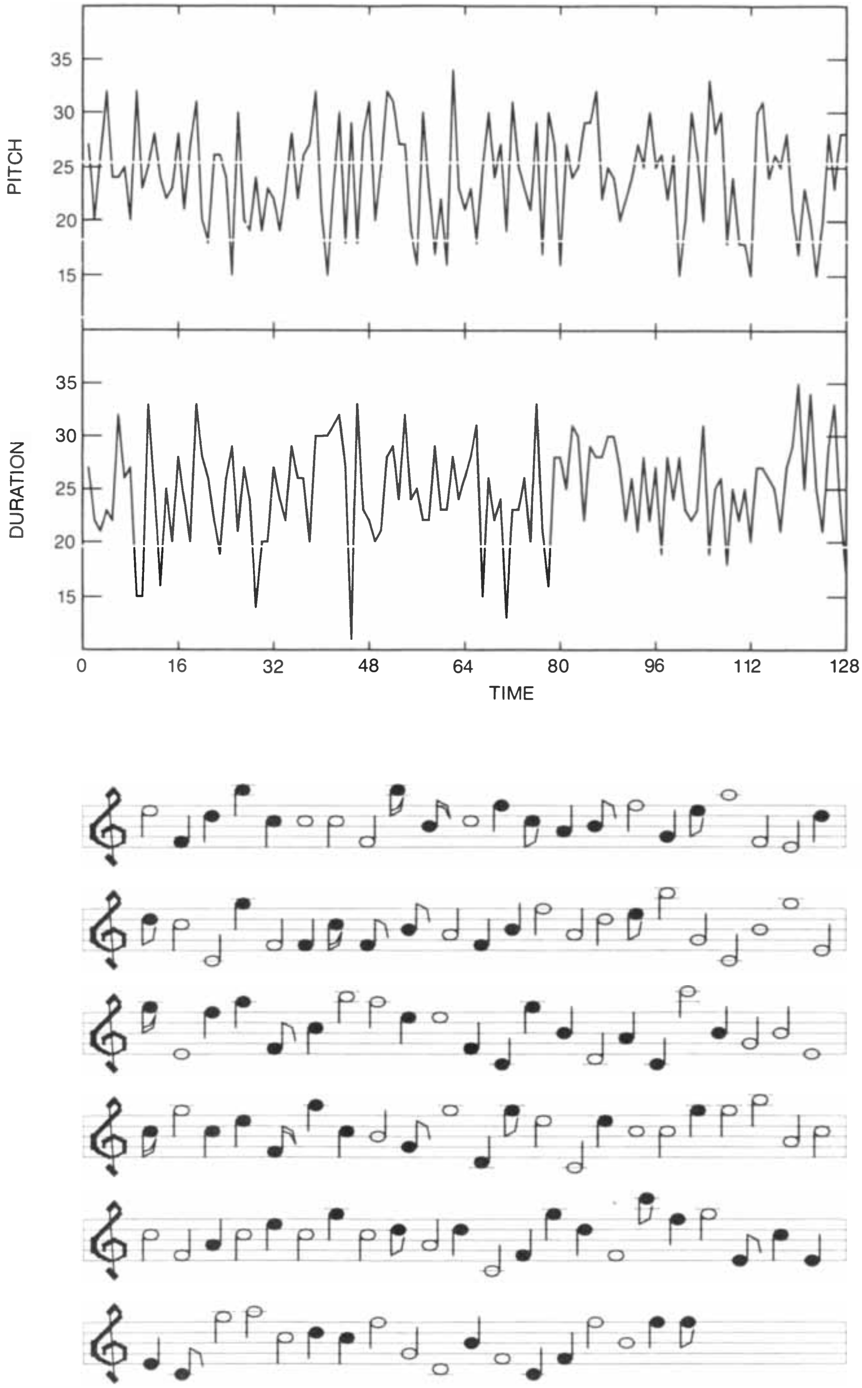
3 636-acre survival course

The Eaton Proving Ground at Marshall, Michigan, subjects truck and automobile components to horrendous treatment. The facilities include a 1.6 mile oval track for sustained endurance testing; road surfaces that are just plain hostile; a salt bath for corrosive brake lining tests; and a 1200-foot skid pad that can simulate any skid condition.

It's all part of Eaton's total commitment to quality. We market new engineering ideas only after having proved that they excel at surviving in the real world.

Eaton is a family of technologically related businesses with a balanced combination of manufacturing and engineering skills. We're always looking for new ways to use these skills in markets where needs are growing. This approach to the management of change has been achieving record sales and earnings. For the complete story, write to: Eaton Corporation, 100 Erieview Plaza, Cleveland, Ohio 44114.

EATON



White music

to put it another way the changing content of our total experience) seems to cluster around $1/f$ noise. It is certainly not entirely uncorrelated, like white noise, nor is it as strongly correlated as brown noise. From the cradle to the grave our brain is processing the fluctuating data that come to it from its sensors. If we measure this noise at the peripheries of the nervous system (under the skin of the fingers), it tends, Mandelbrot says, to be white. The closer one gets to the brain, however, the closer the electrical fluctuations approach $1/f$. The nervous system seems to act like a complex filtering device, screening out irrelevant elements and processing only the patterns of change that are useful for intelligent behavior.

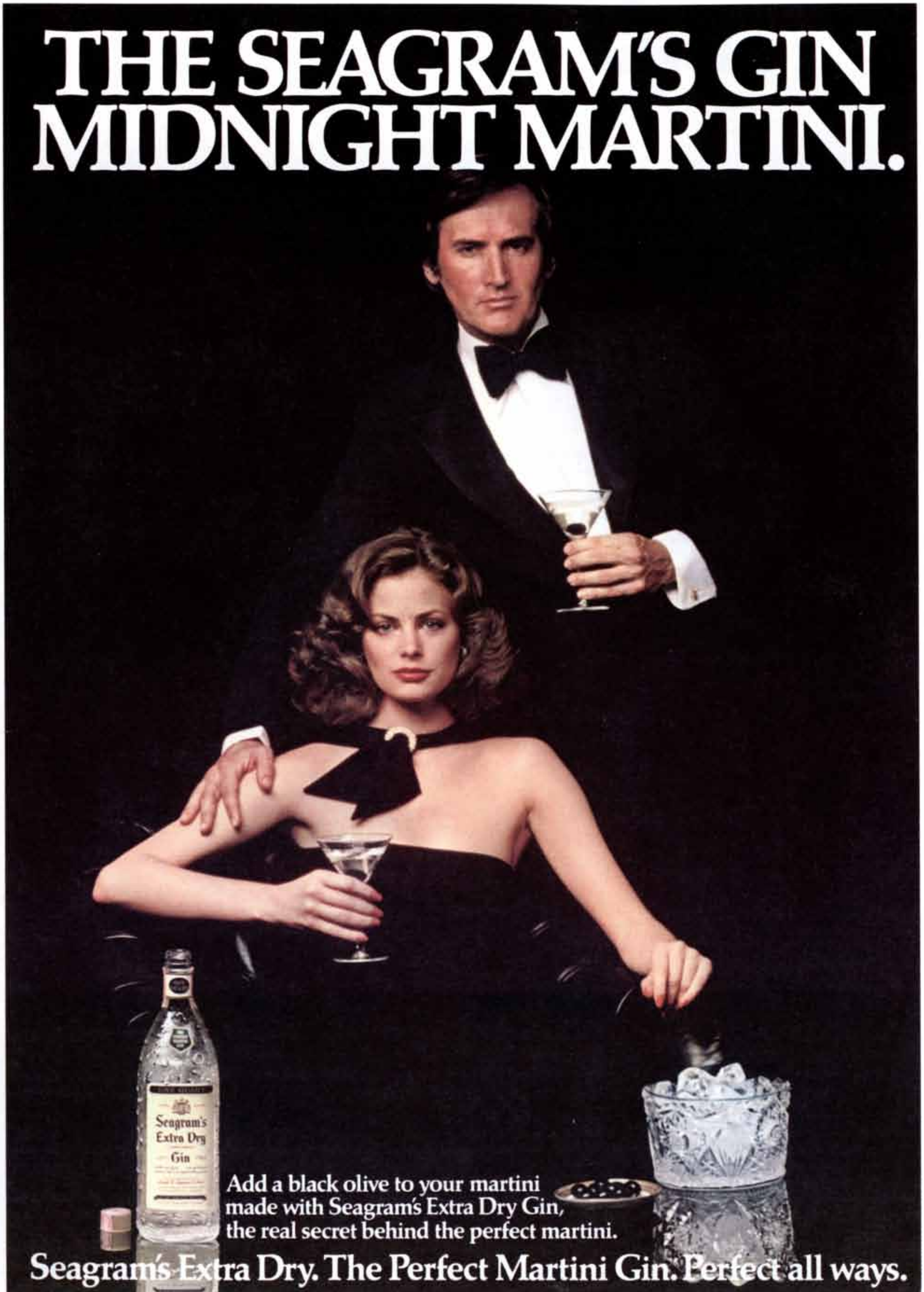
On the canvas of a painting colors and shapes are static, reflecting the world's static patterns. Is it possible, Mandelbrot asked himself many years ago, that even completely nonobjective art, when it is pleasing, reflects fractal patterns of nature? Mandelbrot has some unpublished speculations along these lines. He is fond of abstract art, and maintains that there is a sharp distinction between such art that has a fractal base and such art that does not, and that the former type is widely considered the more beautiful. Perhaps this is why photographers with a keen sense of aesthetics find it easy to take pictures, particularly photomicrographs, of natural patterns that are almost indistinguishable from abstract expressionist art.

Motion can be added to visual art, of course, in the form of the motion picture, the stage, kinetic art and the dance, but in music we have meaningless, nonrepresentational tones that fluctuate to create a pattern that can be appreciated only over a period of time. Is it possible, Voss asked himself, that the pleasures of music are partly related to scaling noise of $1/f$ spectral density? That is, is this music "imitating" the $1/f$ quality of our flickering experience?

That may or may not be true, but there is no doubt that music of almost every variety does exhibit $1/f$ fluctuations in its changes of pitch as well as in the changing loudness of its tones. Voss found this to be true of classical music, jazz and rock. He suspects it is true of all music. He was therefore not surprised that when he used a $1/f$ flicker noise from a transistor to generate a random tune, it turned out to be more pleasing than tunes based on white and brown noise sources.

The illustration on page 21, supplied by Voss, shows typical patterns of white, $1/f$ and brown when noise values (vertical) are plotted against time (horizontal). These patterns were obtained by a computer program that simulates the generation of the three kinds of sequences by tossing dice. The white noise is based on the sum obtained by repeated tosses of 10 dice. These sums range from 10 to 60, but the probabilities naturally force a clustering around the median. The Brownian noise was generated by tossing a single die and going up one step on the scale if the number was even and down a step if the number was odd.

The $1/f$ noise was also generated by simulating the tossing of 10 dice. Although $1/f$ noise is extremely common in nature, it was assumed until a few months ago that it is unusually cumbersome to simulate $1/f$ noise by randomizers or computers. Previous composers of stochastic music probably did not even know about $1/f$ noise, but if they did, they would have had considerable difficulty generating it. As this article was being prepared Voss was asked if he could devise a simple procedure by which readers could produce their own $1/f$ tunes. He gave some thought to the problem and to his surprise hit on a clever way of simplifying existing $1/f$

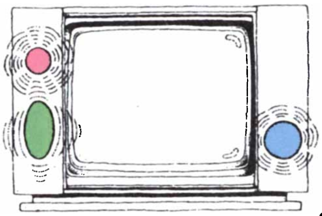

THE SEAGRAM'S GIN MIDNIGHT MARTINI.
Seagram's Extra Dry Gin
Add a black olive to your martini made with Seagram's Extra Dry Gin, the real secret behind the perfect martini.
Seagram's Extra Dry. The Perfect Martini Gin. Perfect all ways.

How to buy a television set with your eyes closed.

Ever since the early days of television, manufacturers have stressed picture, picture, picture.

As a result, television sound has changed little since the introduction of the first sets over 50 years ago.

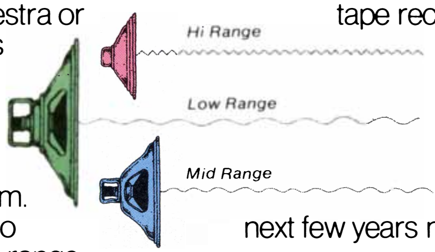
Hear as well as you see. But now to go along with Quasar's sharpest, clearest picture yet, we add Audio Spectrum Sound.™



A Quasar exclusive, it finally makes television sound as good as it looks.

Not one speaker but three. You see, whether it's a 100 piece orchestra or a simple conversation, most televisions cram everything through one speaker.

But our new Audio Spectrum Sound sets have three speakers. Each technically tailored to reproduce a precise portion of the sound spectrum.



Higher highs, lower lows. So there's a conventional speaker for mid-range sounds. A bass speaker specifically designed to deliver lower lows, plus another speaker for higher highs.

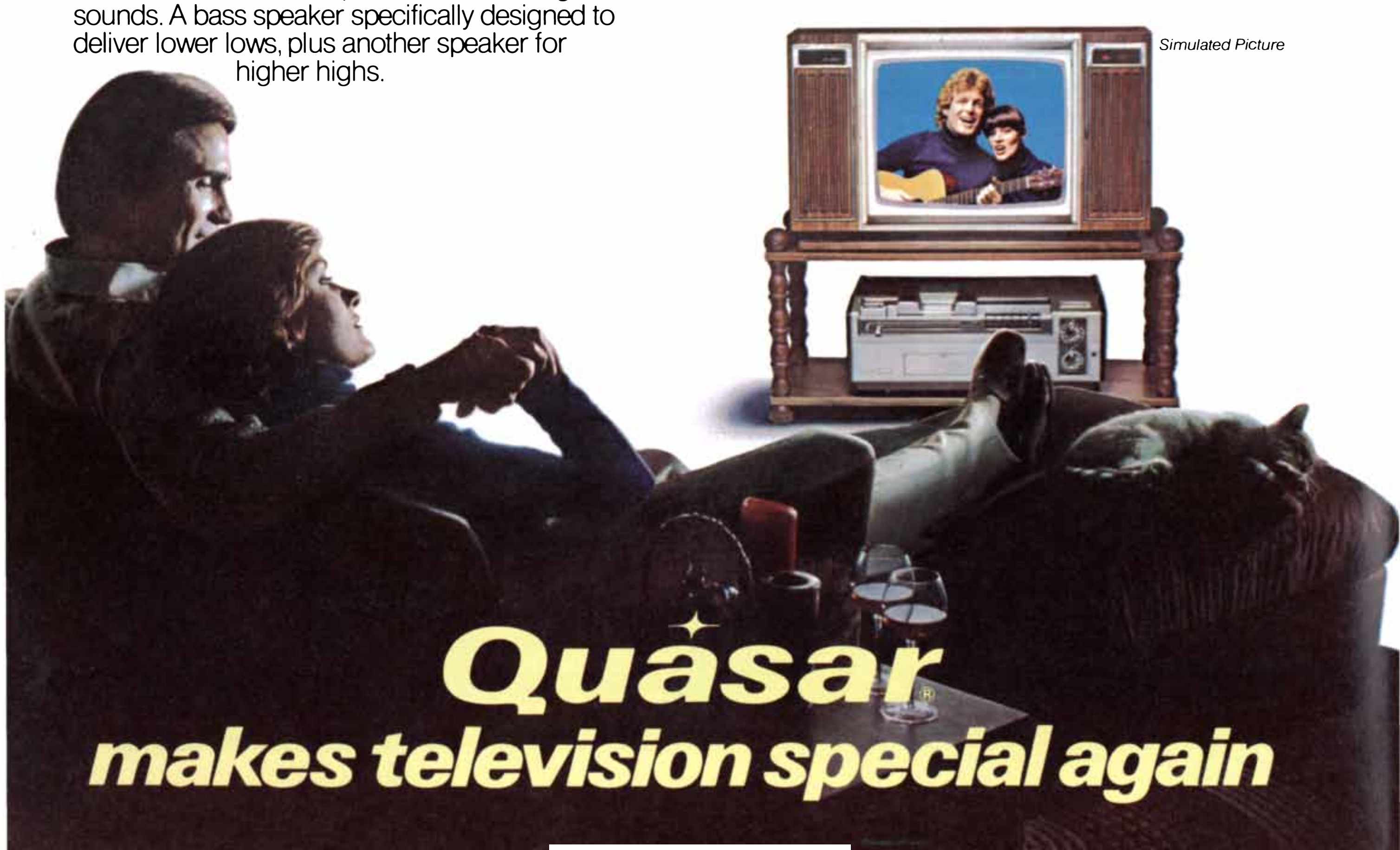
Together they add a whole new exciting dimension of reality to every show you watch.

Why it's *so* real, that when you watch a concert you'll think you're in the front row instead of your living room.

Some features we'd never change. Besides a great new sound system, Audio Spectrum Sound sets also feature traditional Quasar quality and reliability. Because the show must always go on. And on. And on.

So before buying your next TV, see your Quasar dealer and hear the difference Audio Spectrum Sound makes. And while you're there, see Quasar's Great Time Machine. It's a home video tape recorder that lets you record your favorite programs to watch at your favorite times. (That's it under the TV set below.)

But remember, if you don't hear Audio Spectrum Sound, you might spend the next few years missing out on a lot.



Simulated Picture

computer algorithms that does the trick beautifully.

The method is best explained by considering a sequence of eight notes chosen from a scale of 16 tones. We use three dice of three colors: red, green and blue. Their possible sums range from 3 to 18. Select 16 adjacent notes on a piano, black keys as well as white if you like, and number them 3 through 18.

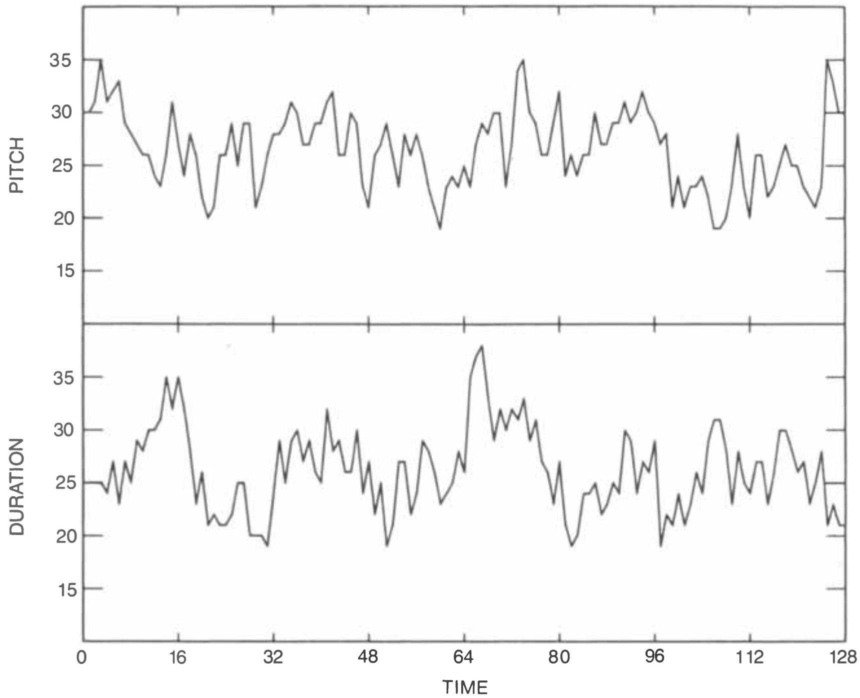
Write down the first eight numbers, 0 through 7, in binary notation, and assign a die color to each column as is shown in the illustration on page 22. The first note of our tune is obtained by tossing all three dice and picking the tone that corresponds to the sum. Note that in going from 000 to 001 only the red digit changes. Leave the green and blue dice undisturbed, still showing the numbers of the previous toss. Pick up only the red die and toss it. The new sum of all three dice gives the second note of your tune. In the next transition, from 001 to 010, both the red and green digits change. Pick up the red and green dice, leaving the blue one undisturbed, and toss the pair. The sum of all three dice gives the third tone. The fourth note is found by shaking only the red die, the fifth by shaking all three. The procedure, in short, is to shake only those dice that correspond to digit changes.

It is not hard to see how this algorithm produces a sequence halfway between white and brown. The least significant digits, those to the right, change often. The more significant digits, those to the left, are stabler. As a result dice corresponding to them make a constant contribution to the sum over long periods of time. The resulting sequence is not precisely $1/f$ but is so close to it that it is impossible to distinguish melodies formed in this way from tunes generated by natural $1/f$ noise. Four dice can be used the same way for a $1/f$ sequence of 16 notes chosen from a scale of 21 tones. With 10 dice you can generate a melody of 2^{10}, or 1,024, notes from a scale of 55 tones. Similar algorithms can of course be implemented with generalized dice (octahedrons, dodecahedrons and so on), spinners or even tossed coins.

With the same dice simulation program Voss has supplied three typical melodies based on white, $1/f$ and brown noise. The computer printouts of the melodies are shown in the illustrations on page 24, this page and page 28. In each case Voss varied both the melody and the tone duration with the same kind of noise. Above each tune are shown the noise patterns that were used.

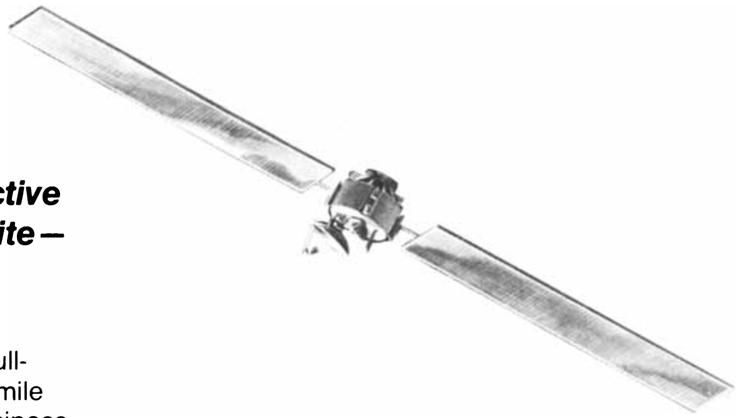
Over a period of two years tunes of the three kinds were played at various universities and research laboratories, for many hundreds of people. Most listeners found the white music too random, the brown too correlated and the $1/f$ "just about right." Indeed, it takes only a glance at the music itself to see how the $1/f$ property mediates between the two extremes. Voss's earlier $1/f$ music was based on natural $1/f$ noise, usually electronic, even though one of his best compositions derives from the record of the annual flood levels of the Nile. He has made no attempt to impose constant rhythms. When he applied $1/f$ noise to a pentatonic (five-tone) scale and also varied the rhythm with $1/f$ noise, the music strongly resembled Oriental music. He has not tried to improve his $1/f$ music by adding transition or rejection rules. It is his belief that stochastic music with such rules will be greatly improved if the underlying choices are based on $1/f$ noise rather than the white noise so far used.

Note that $1/f$ music is halfway between white and brown in a fractal sense, not in the manner of music that has transition rules added to white music. As we have seen, such music reverts to white when we compare widely separated parts. But $1/f$ music has the frac-



1/f music

HP measurement and computer advances

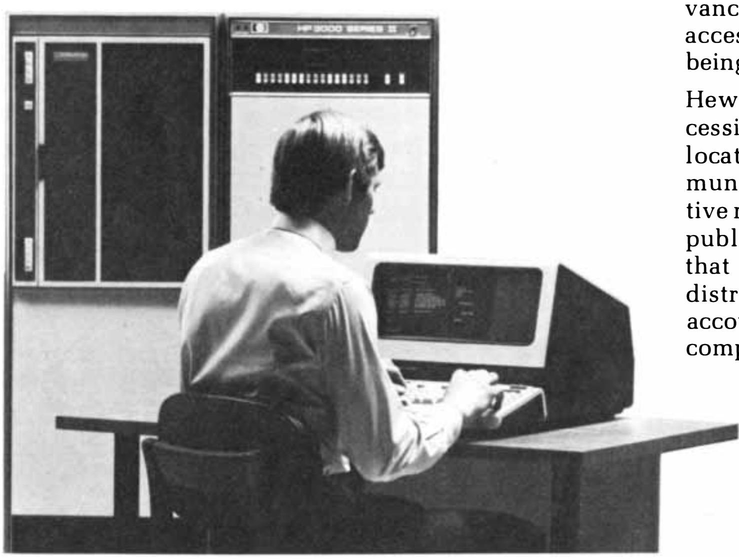


SBS extends frontiers in interactive data communications via satellite—with the help of HP computers.

Using a satellite to transmit data, voice, full-motion and freeze-frame video, and facsimile documents—all interactively—Satellite Business Systems (SBS) has undertaken a pace-setting experiment in advanced communications for geographically dispersed organizations. HP 3000 Series II computers were chosen for the data processing and operations management portions of the experiment.

Project Prelude, an experiment in advanced communications, has been successfully concluded by Satellite Business Systems in cooperation with host companies that included Rockwell International Corp., Texaco Inc., and Montgomery Ward & Co., Inc. The experiment demonstrated the feasibility of high-speed, low-cost intracompany communications via satellite among widely dispersed facilities.

Providing the link was the Communications Technology Satellite (CTS), an experimental communications satellite used jointly by NASA and the Canadian Department of Communications.



"This is the first CTS experiment to include small earth stations on customer premises and integrated audio/data/image communications in a digital format," says Tom Rush, SBS Project Coordinator. "This system was designed to explore large-scale economies and improved control in organizations with heavy information traffic among geographically dispersed facilities.

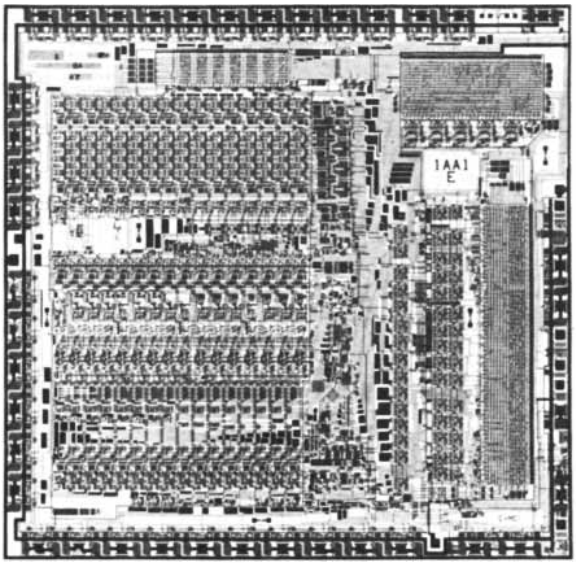
"We selected HP 3000 Series II computer systems because their standard software could handle the fast data rates, and they certainly had proven distributed data processing capability."

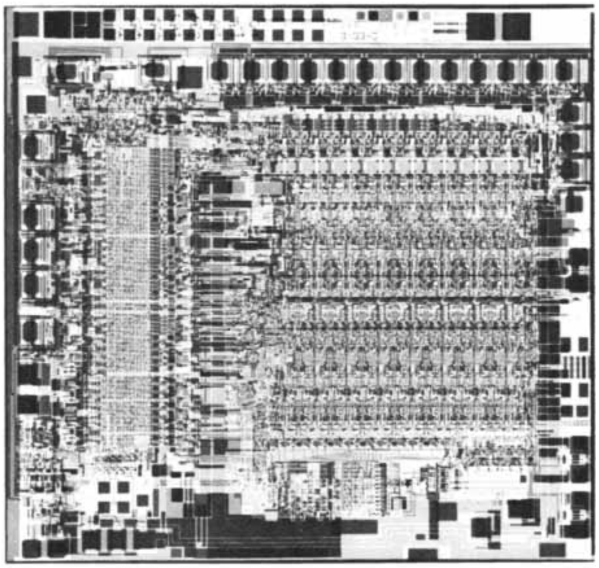
HP distributed processing could extend your possibilities.

HP's distributed processing capabilities stem from a generalized network architecture that can fit computer systems to your needs without reshaping your organizational structure. The availability of low-cost, high-performance computers and advanced systems network software can put easily accessible processing power wherever the work is being done.

Hewlett-Packard is committed to distributed processing as the sensible way to enable functional locations within an organization to process, communicate, and share information for a more effective response to changing conditions. We have just published an issue of the *Hewlett-Packard Journal* that deals with some of the technical niceties of distributed systems networks, including a fuller account of Project Prelude and satellite-linked computer networks. Mail the coupon for your copy.

extend your possibilities.





These photographic enlargements of two HP microprocessor chips—one with NMOS II large-scale integrated circuits at right and one with CMOS/SOS at left—illustrate a major reason for the rapid growth of digital circuits: increasing functional density at a decreasing cost. These chips contain as many as 10,000 transistors and are about this big ■

How to operate more confidently in an increasingly digital world.

While analog circuits generally convey data by a continuum of voltages, the digital world of microprocessors and LSI circuits relies on two voltage levels, conveying information by switching between them. Because of this fundamental difference, digital circuits pose a special set of problems for both designers and troubleshooters. And HP has some excellent solutions.

As the solid state revolution increases their functional density at a decreasing cost, digital circuits are infiltrating areas that previously relied on analog technology, and are cropping up in an endless procession of products that include microwave ovens, automobiles, home computers, process control (analog from its inception), telecommunications, and a range of industrial controls.

In short, the technical, industrial, and consumer worlds are turning digital. While this greatly extends possibilities, it also introduces some complications—most notably in the design, debugging, software development, and diagnosis of these functional leviathans in minnow-sized packages. While it may not be immediately obvious, this digital data domain requires special measurement tools. And here Hewlett-Packard is making significant strides, continually introducing new measurement approaches and instruments designed to make the digital revolution manageable, and thus enable technical people to operate more confidently.

If you are wrestling with any aspect of digital circuitry, you will be interested in reading the February issue of the *Hewlett-Packard Journal*, which explores HP's growing family of digital logic analysis test instruments at some length, and offers a few viewpoints of possible interest. Just mail the coupon, and we will be pleased to send you a copy.

HEWLETT hp PACKARD

1503 Page Mill Road, Palo Alto, California 94304

For assistance call: Washington (301) 948-6370, Chicago (312) 255-9800, Atlanta (404) 955-1500, Los Angeles (213) 970-7500

Mail to: Hewlett-Packard, 1503 Page Mill Road, Palo Alto, CA 94304.
Please send me the Hewlett-Packard Journal dealing with:

( ) HP distributed processing
( ) HP digital logic analysis

Name ____________________
Company ____________________
Address ____________________
City ____________ State ________ Zip ________

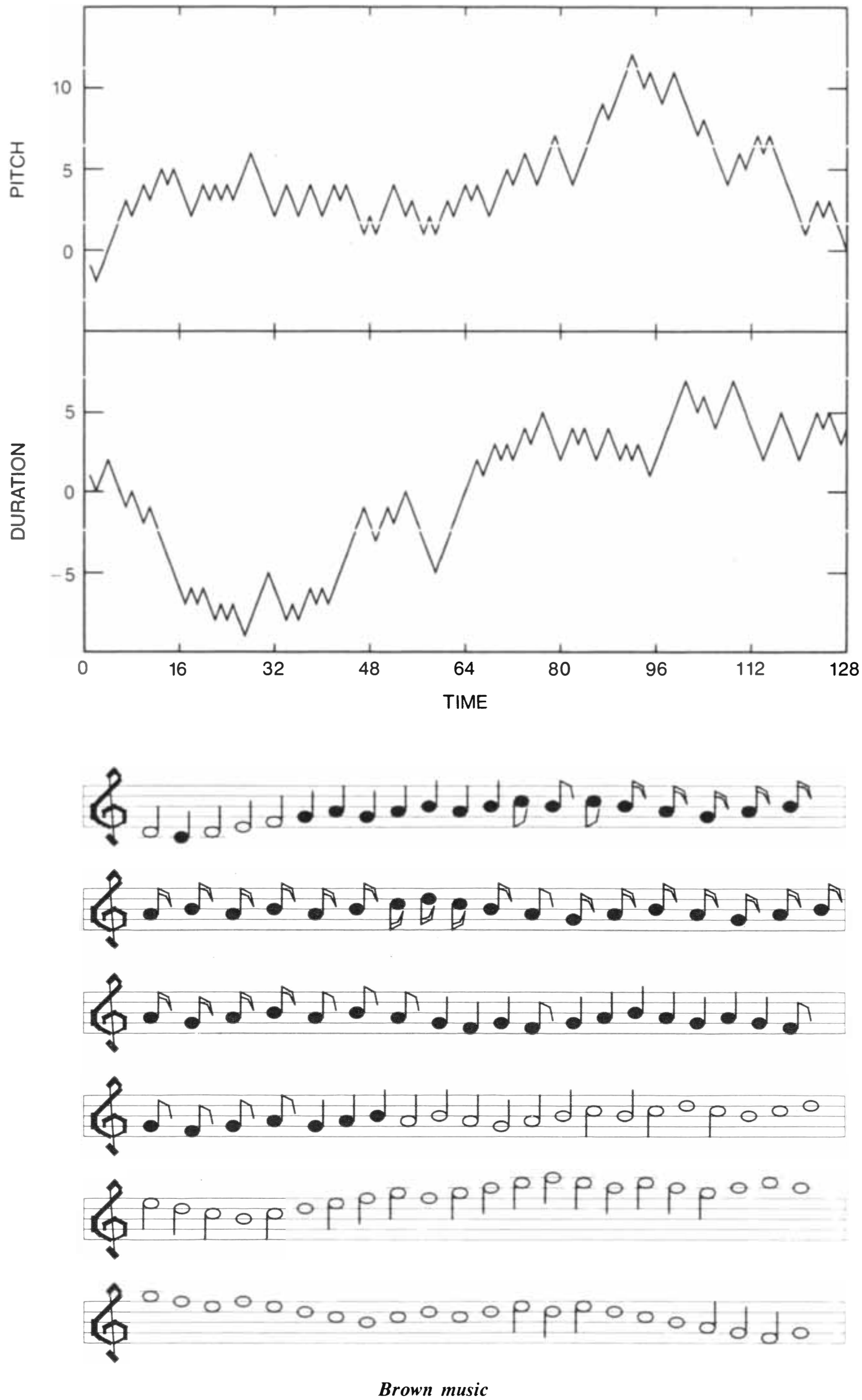
*Domestic U.S. prices only.

00844

tal self-similarity of a coastline or a mountain range. Analyze the fluctuations on a small scale, from note to note, and it is $1/f$. The same is true if you break a long tune into 10-note sections and compare them. The tune never forgets where it has been. There is always some correlation with its entire past.

It is commonplace in musical criticism to say that we enjoy good music because it offers a mixture of order and surprise. How could it be otherwise? Surprise would not be surprise if there were not sufficient order for us to anticipate what is likely to come next. If we guess too accurately, say in listening to a tune that is no more than walking up and down the keyboard in one-step intervals, there is no surprise at all. Good music, like a person's life or the pageant of history, is a wondrous mixture of expectation and unanticipated turns. There is nothing new about this insight, but what Voss has done is to suggest a mathematical measure for the mixture.

I cannot resist mentioning three curious ways of transforming a melody to a different one with the same $1/f$ spectral density for both tone patterns and durations. One is to write the melody backward, another is to turn it upside down and the third is to do both. These trans-



Brown music

formations are easily accomplished on a player piano by reversing and/or inverting the paper roll. If a record or tape is played backward, unpleasant effects result from a reversal of the dying-away quality of tones. (Piano music sounds like organ music.) Reversal or inversion naturally destroys the composer's transition patterns, and that is probably what makes the music sound so much worse than it does when it is played normally. Since Voss composed his tunes without regard for short-range transition rules, however, the tunes all sound the same when they are played in either direction.

Canons for two voices were sometimes deliberately written, particularly in the 15th century, so that one melody is the other backward, and composers often reversed short sequences for contrapuntal effects in longer works. The illustration on page 31 shows a famous canon that Mozart wrote as a joke. In this instance the second melody is the one you see taken backward and upside down. Thus if the sheet is placed flat on a table, with one singer on one side and the other singer on the other, the singers can read from the same sheet as they harmonize!

No one pretends, of course, that stochastic $1/f$ music, even with added transition and rejection rules, can compete with the music of good composers. We know that certain frequency ratios, such as the three-to-two ratio of a perfect fifth, are more pleasing than others, either when the two tones are played simultaneously or in sequence. But just what composers do when they weave their beautiful patterns of meaningless sounds remains a mystery that even they do not understand.

It is here that Plato and Aristotle seem to disagree. Plato viewed all the fine arts with suspicion. They are, he said (or at least his Socrates said), imitations of imitations. Each time something is copied something is lost. A picture of a bed is not as good as a real bed, and a real bed is not as good as the universal, perfect idea of bedness. Plato was less concerned with the sheer delight of art than with its effects on character, and for that reason his *Republic* and *Laws* recommend strong state censorship of all the fine arts.

Aristotle, on the other hand, recognized that the fine arts are of value to a state primarily because they give pleasure, and that this pleasure springs from the fact that artists do much more than make poor copies.

They said, "You have a blue guitar,
You do not play things as they are."
The man replied, "Things as they are
Are changed upon the blue guitar."

Wallace Stevens intended his blue guitar to stand for all the arts, but music, more than any other art and regardless

OMNI'S ARMY.

The new Dodge Omni. People just seem to like it. Because Omni tries so hard to please. With a 99.2-inch wheelbase. A 1.7 litre overhead cam engine. And a likeable base sticker price of only $3706,* including standard features like an AM radio, vinyl body side moldings, and white sidewall tires.

The Omni shown below with a number of stylish additions has a sticker price of just $3981.*

$3981.* (PRICE OF CAR AS SHOWN)
$3706.* (BASE STICKER PRICE)

MOTOR TREND CAR OF THE YEAR

Even big people like Omni. For example, the four weekend hackers you see here. All four of their golf bags fit in a covered compartment behind the rear seat. If this were a twosome, that rear seat could be folded down giving 35.8 cubic feet of space. Enough to handle their bags and a couple of carts. Omni's built here in America with American-sized people very much in mind.

Lots of room. That's one advantage of Omni's front-wheel drive. The size of the tunnel you find in the middle of most conventional rear-wheel-drive cars is reduced. So there's more room for legs and stuff.

Then there's ride. It's stable. Comfortable. Because the right kind of attention was paid to things like spring rates and jounce travel. Front-wheel drive with rack-and-pinion steering does its part, too.

Another thing, with four good-sized suburbanites like you see here, Omni's gritty four-cylinder engine can make this car move with plenty of authority. No sweat at expressway speeds.

Remarkable, when you consider the EPA estimated mileage ratings achieved by that engine and Omni's standard four-speed manual transmission and a 3.3 transaxle ratio:

EPA ESTIMATES:
38 MPG HIGHWAY/25 MPG CITY.

Now, your mileage may vary according to your car's condition, equipment, and your driving habits. And California mileage is lower. But these are noteworthy ratings for a car that can do all the Omni can do.

Here's a suggestion. Visit your Dodge Dealer. Have him wheel out an Omni. Take it out on the freeway. Over a bumpy road. We have a feeling when you're through, you'll be ready to buy, or lease. Omni's like that. It attracts fans.

A PRODUCT OF CHRYSLER CORPORATION

DODGE OMNI.
IT DOES IT ALL.

*Excluding taxes and destination charges.

of what imitative aspects it may have, involves the making of something utterly new. You may occasionally encounter natural scenes that remind you of a painting, or episodes in life that make you think of a novel or a play. You will never come on anything in nature that sounds like a symphony. As to whether mathematicians will someday write computer programs that will create good music—even a simple, memorable tune—time alone will tell.

Last month's logic problems from *What Is the Name of This Book?* by Raymond M. Smullyan are answered:

1. The Lion can say "I lied yesterday" only on two days: Monday and Thursday. The Unicorn can make the same statement only on Thursday and Sunday. Therefore the only day on which both the Unicorn and the Lion can make the statement is Thursday.

2. The inscriptions on the gold and

Mozart's palindromic and invertible canon lead caskets say the opposite, so that one of them must be true. Since at most only one statement is true, the statement on the silver casket is false. The portrait is therefore in the silver casket.

3. If *B* is innocent, then we know (by fact 1) that either *A* or *C* is guilty. If *B* is guilty, he must have had an accomplice because he cannot drive; therefore again *A* or *C* must be guilty. Consequently *A* or *C* or both are guilty. If *C* is innocent, *A* must be guilty. If *C* is guilty, then (by fact 2) *A* is also guilty. Therefore *A* is guilty.

4. You say "I am not a poor knight." The girl reasons that if you were a knave, you would indeed not be a poor knight; therefore your statement would be true. Since a knave never makes a true statement, the contradiction eliminates the assumption. Hence you are a knight. Knights speak truly, and so you are not a poor knight.

5. An inhabitant on the Island of Zombies has replied "Bal" to the question, "Does bal mean yes?" If bal means yes, then bal is the truthful answer; therefore the speaker is human. If bal means no, then that too is truthful; therefore the speaker is human. It is not possible to determine what "Bal" means, but the answer does prove that the islander is human.

6. To determine in one yes-no question what "Bal" means ask the islander if he is human. Since both human beings and zombies answer yes to such a question, if the islander answers "Bal," the word means yes. If the islander answers "Da," then "Da" means yes and "Bal" means no.

7. To tell whether a Transylvanian is a vampire by asking one yes-no question, ask him if he is sane. A vampire will say no and a human being will say yes. (I leave the proof to the reader.) To tell whether the Transylvanian is sane ask him if he is a vampire.

Admirers of M. C. Escher, and anyone interested in the beautiful symmetries of regular polyhedrons, will be fascinated by *M. C. Escher Kaleidocycles,* a package of mathematical toys published in November by Ballantine Books. The package contains 17 sheets (many brilliantly colored) of partially die-cut boards and a book of instructions for assembling the "Escher sculpture."

The models range from Platonic solids decorated with Escher patterns to rings of tetrahedra that "flex" in strange ways. They were designed by Doris Schattschneider, a mathematician at Moravian College, and Wallace Walker, a New York graphic designer. Mrs. Schattschneider contributes to the booklet an informative essay on the mathematical structure of the models and reproduces some Escher graphics never published before.

BOOKS

Muscle power, the unique painted pottery of the Mimbres culture and other matters

by Philip Morrison

MECHANICS AND ENERGETICS OF ANIMAL LOCOMOTION, edited by R. McN. Alexander and G. Goldspink. Halsted Press, John Wiley & Sons ($37.50). PEDAL POWER: IN WORK, LEISURE, AND TRANSPORTATION, edited by James C. McCullagh. Rodale Press, Emmaus, Pa. ($4.95). Every animal that moves has the analogue of an engine, an electronic control system and a specific dynamical mechanism for converting internal energy into the appropriate motion. Two decades ago the work of the late Professor Sir James Gray and his school outlined the problems and teased out just how the forces that move each creature arise from its interaction with the environment. *Mechanics and Energetics of Animal Locomotion* is a compendium, tightly packed with results and brief analyses and rich in its references, of the work of a group of British and American investigators, a comprehensive summary of what we know. Our view is based on a much deeper understanding of muscle itself, on big gains in neurophysiology (particularly in insects), on physiological measurements made possible by wind or water tunnels and on performance analysis and field tests carried out at the level of detail of quantitative engineering design.

The engine is of course muscle, fueled by ATP (except for the fast contractions of certain tiny creatures such as *Stentor,* where a calcium potential gives rise to tension directly). One chapter surveys the intricate design of muscle and its force and work relations seen as a function of rates and lengths. A second chapter details the molecular mechanisms of muscle motion and fuel supply; a third reviews the prime mover's energy economy. One point is basic and is emphasized anew: muscles stretched while still developing tension (doing negative work, like the muscles of a man walking downstairs) consume much less fuel. A. V. Hill once demonstrated this to the Royal Society. A slight woman pedaling to slow down easily resisted the best efforts of a "large healthy male athlete" doing the positive work on their coupled ergometers.

This perception is a key to much natural design. The fast-beating wings of a mosquito would consume much negative work even at this big discount rate, but elastic hinges (made of a special protein super-rubber) mean that no such work is done. The wings resonate, more or less; the muscles do not work against the inertial forces but only against the air (working a trifle more to make up for the lack of total perfection in the elastic material). How the neural controls manage it is now pretty clear. The wing muscles contract in response mainly to their own stretch; infrequent neuron impulses merely keep them in an activated state. A hopping kangaroo shows similar elastic savings. The jump of the flea is a spring-catapulted feat, wound up by slow muscle contraction and suddenly released. Muscle fiber could not react swiftly enough on a flea scale, even though there are fast muscle fibers and slow ones. Many muscles contain more than one type of fiber, each reaching its optimum efficiency at a different rate of contraction, "as it were, a two- or three-geared system."

The gait of the cockroach has been studied in detail and is described here in the context of many patterns of loping quadrupeds: duty factors of the several feet, stability and favored speeds of gait-pattern shift. There is even a hopeful block diagram of the neural circuitry that might underlie the gait, with an input block marked "Command to walk" (a roach thought?). Archy would have liked it.

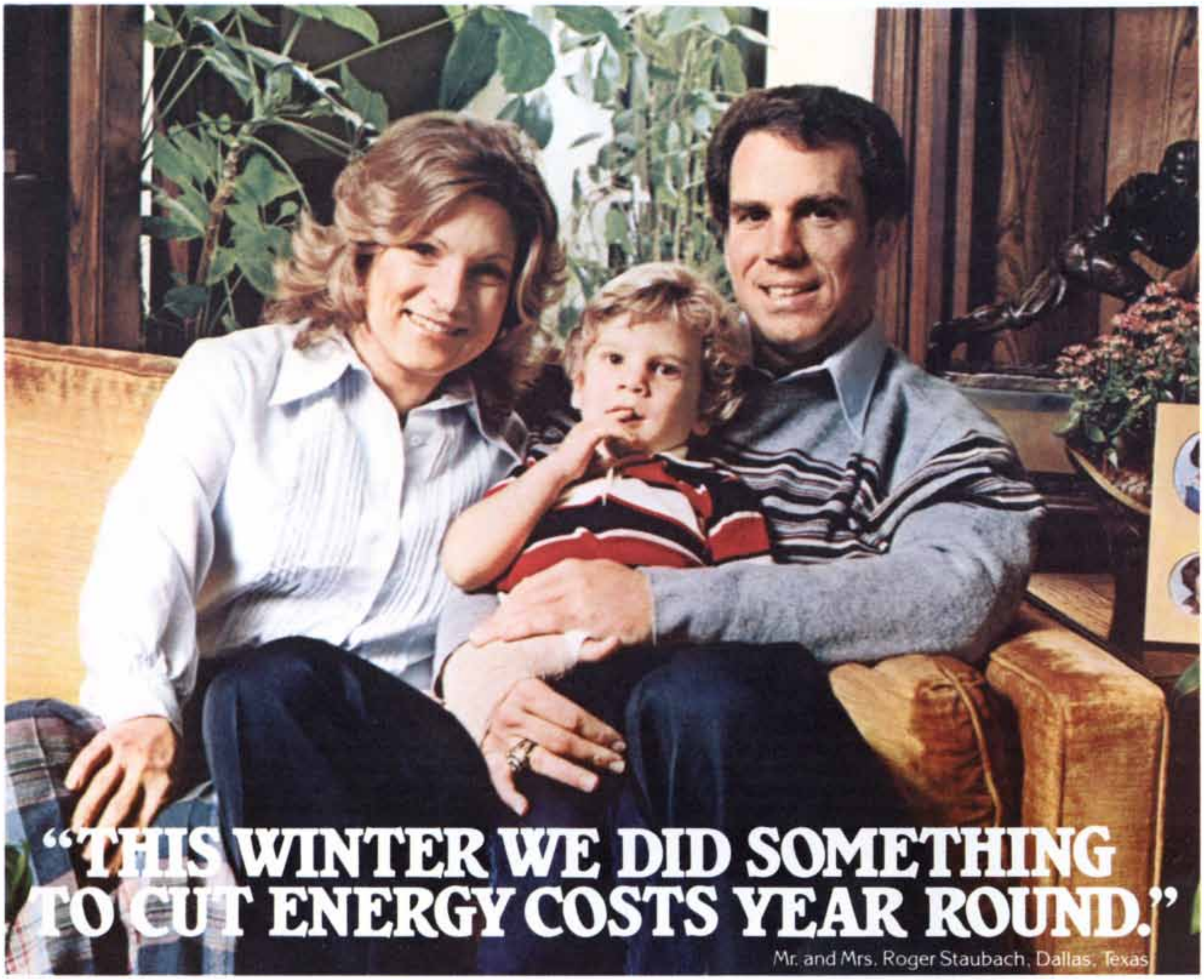
If it walks, runs, trots, crawls or burrows, there is some account of its dynamics, neatly done. A careful mechanical treatment in dimensionless terms by R. McN. Alexander of the University of Leeds helps one understand the effects of scale. Swimming (along with rowing and jet propulsion in water) is discussed in depth; so is flight, from hovering through gliding and soaring to fast flapping. There is a report of C. J. Pennycuick with his data on East African vultures and storks, collected as the investigator and the birds flew around together under the same cumulus cloud a few years ago. (He occupied a small motorized glider.) A bumblebee lies a little off the mean hovering curves, with a fast beat and small wings for its size, but there is no doubt of the basis of its flight design; we understand it mainly from the brilliant work of the late Torkel Weis-Fogh. (Professor Alexander gives us the review chapters on flight and swimming that Weis-Fogh did not live to prepare.)

Fishes have three ways to avoid sinking to the bottom. Hydrofoil-like fins work, given continuous swimming; low-density compounds such as squalene can balance off the denser tissue; a gas-filled swim bladder is energetically a better idea—maybe. Slow swimmers in fact do better with the light low-density compounds, but fast swimming speeds favor the active-lift scheme (all that extra volume makes for extra drag, which is serious at higher speeds). The gas bladder is best of all, but it takes chemical work to keep it at the ambient pressure; only big fish can afford to maintain a costly gas bubble at depth. The final chapter deals with single cells and their motion through a sticky world. Amoeba cytoplasm sets tough problems in rheology, but it is there that the motion of these "lobose carnivores" is to be sought.

This work is a particularly neat example of British bookmaking, although it is expensive for its modest size. There are many drawings; nearly all the other pictures are electron micrographs. It is a rich lode of physics and engineering, and an up-to-date road map to a big literature of wonderful freshness and ingenuity, in the high tradition of the physiologists of golden Bloomsbury times.

Only *Homo sapiens* has the muscle-driven wheel, which is celebrated in *Pedal Power,* a big paperback. David G. Wilson, a spry M.I.T. mechanical engineer, can be seen almost any day skimming the Cambridge streets in his own redesigned recumbent bicycle (lie back and enjoy the ride); he says it is safer, more efficient and more comfortable. Here we read of historical and modern pedicabs, bicycles with outriggers to ride the rails, pedal boats and lathes and a neat design for a homemade anchored bike to power kitchen, shop and garden tasks (or even television) while you exercise for 20 minutes. The theme is in the title: this is the best means of coupling the muscles of the human thigh and back to yield usable rotary power, at the rate of an easy tenth of a horsepower—or more for spurts and from athletes.

The work is at once practical and philosophical, not to say here and there rather hortatory; it offers information at many levels yet by no means offends simple economic man. We can do a lot more with pedals, even where food is dearer than petroleum. Stuart S. Wilson of Oxford outlines useful ways to harness pedal power in the context of the Third World, where power lines and refineries are distant. "Perhaps an interface between East and West is the bicy-



"In recent years, we've done a lot of little things to try to save energy. Always with the same result. Our fuel bills have gone up anyway. So this winter we did a big thing. We put in a Year-round One heat pump air conditioner from Carrier.

"Already the Year-round One has cut our heating bills. This summer it should save us even more energy on cooling.

"Currently, we're building a new, larger home for our family. We'll heat it and cool it entirely with Carrier heat pumps. Given the energy situation, we think it's the only way to go.

"And given Carrier's expertise in both air conditioners and heat pumps, we think Carrier is the only way to go.

"Cut your fuel bills, too. See your Carrier Certified Heat Pump dealer."

WE CAN'T CONTROL THE WEATHER.
BUT WE CAN HELP YOU CONTROL ITS COST.

Number One Air Conditioning Maker
Division of Carrier Corporation
Carrier

cle, the machine which makes us all brothers and sisters."

MIMBRES PAINTED POTTERY, by J. J. Brody. School of American Research, University of New Mexico Press ($22.50). The Mimbres is not much of a river. It rises among well-watered, intricate volcanic peaks but dries up as it wanders south into the desert plain near Deming in southwestern New Mexico, 30 miles north of the Mexican border. Along the Mimbres, from the intermittent lower reaches ("once a pleasant if hot country") up to the cooler and greener mountain terraces, 100 or so village ruins have been found. They were villages of pit houses and adobe complexes, typically sheltering a few dozen families of self-sufficient people whose staple was maize. They were hunters and gatherers as well, expert at making "full use of the animal and vegetable life that thrived around them," from bison to yucca, as the cycle of the year offered.

Their epoch is not easy to fix, but the Mimbres Phase of the villages is established by a few carbon dates and tree rings and by association with well-dated pottery trade wares. Their style of painted pottery began around the 10th century of this era and continued at a peak for some 200 years, fading with the abandonment of the valley settlements. We can partly trace their origins out of the complex influences of the 1,000 years before their central time, but we do not know why they left and we can see only dimly their later traces south and east of the Mimbres. Neither drought nor erosion nor military necessity drove them away. Their sites were open and far from places of refuge, and there is no "evidence of violence at any Mimbres site thus far excavated." Theirs seems to have been a quiet, peaceful life, self-contained but not isolated.

A dozen generations of artists painted the Mimbres black-on-white pictures, usually on utilitarian near-hemispherical bowls 10 inches across. The pots were used for food and its preparation; they were made in self-consuming kilns of 15 or 20 vessels stacked with dried wood and set alight for an hour or two out in the open, erratically windblown from a reductive state to an oxidized state. The potters were almost surely the women (by analogy with widespread tradition), but the painters were men and women alike. The ware is hard and waterproof and was easy to make, although it is fragile. The pots were slipped with a fine white kaolin taken from a very few deposits. With an artfully chewed brush of limp, wet yucca and a black paste ground out of the iron oxide pebbles commonly found in the area, the Mimbres painter worked with sure strokes; erasing was impractical.

The first few painted pots were found by people at U.S. Army posts in that

Apache country around the turn of this century. Private collectors and finally professional archaeologists brought the Mimbres culture to light. The 1920's saw the chief excavations; the 1930's did little more. With some 1,500 painted pots safe in museums, the interests of the scholars shifted. "Visually or emotionally powerful artifacts such as Mimbres paintings were potential distractors" to students, who now sought sites for their universal implications.

Then in the 1960's the marketplace seized Mimbres. New York dealers sent their exigent agents, local pothunters cratered the sites like the DMZ and bulldozers pushed aside the dirt by day and night. An estimate of 10,000 painted pots now in private and public hands is not implausible; most of them are of unknown provenance. It is a familiar history of "exploitation of an irreplaceable resource by irresponsible commercial interests, and of the inability of the scientific community" to act in protection. Professionals merely cried "thou shalt not dig" and lost the opportunity to enlist fascinated amateurs in the protection of sites. The amateurs, snubbed, themselves became looters.

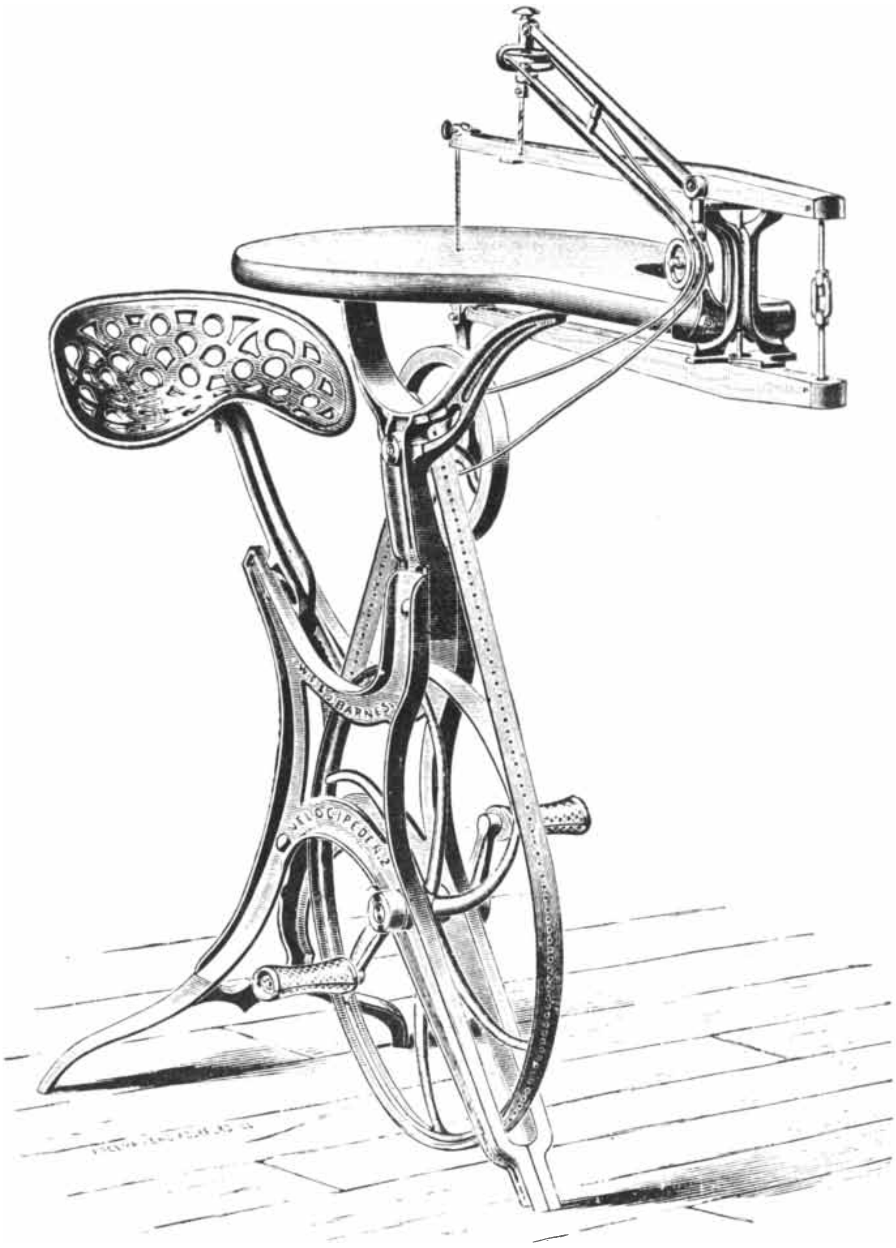
The Mimbres painters produced both nonfigurative and figurative paintings. The geometric patterns, mostly in line, are tensely organized on the pot surface and often "optical" in the up-to-date sense. The lines are so dense in contrast as to load the vision, with a balance of binary tensions: "dark-light, positive-negative, curve-straight, in-out." (The pot painters of Acoma Pueblo today make conscious use of some Mimbres patterns.) The figurative paintings at their best adapt themselves to the azimuthal symmetry of a round pot, to make a picture space out of multiple figures or out of an illusion of interaction that unifies the scene however it is viewed, "pictures with a balance between content and decorative values that are among the most important ever made by Native American artists during pre-Columbian times."

This handsome volume offers us about 150 pots in black-and-white photographs (and a dozen, plus two landscapes, in color), nearly all sensitively made by Fred Stimson, a tireless and fortunate traveler to collections. Professor Brody enters deep waters in his account of the Mimbres paintings and what they might mean. We simply do not know, he argues, and the efforts to impute to them the folksiness and charm of a few poses or clan symbols or animal conversations are simply projections of our own attitudes. There is little we can say, although we can admire the composition and the catholic representation of moons, ant-lions and antelopes, and also the clear sign of mind—not only hand and eye—at work. The viewer is reminded a little of the great cave paintings of Europe when he sees the armadillo with a deerhead mask or the pair of dancers in feline costumes, human hands and feet showing beneath the head and skin.

The iconography of Mesoamerica hints that many motifs refer to a mythology of death. The paradox is deeper than that. The part-time artists of the Mimbres decorated lots of pots. The geometric ones are often clearly marred by long use as food vessels. Not so the figurative pots, which are mainly unused. Almost all the pots we have were grave goods, themselves "killed" by piercing the bottom with a pointed tool before burial. These people lived close to their departed; among 1,009 burials recovered at the large Swarts Ruin more than 900 were found under occupied rooms in unlined pits, with offerings of pottery, tools, turquoise or shell, rarely of food.

Here much of the archaeology comes into question, if we follow the philosophical inquiry of this book. How can we sympathetically unbury what artists long ago plainly meant to set below ground forever? The issues are profound; let us leave them between Brody and his reader. A reviewer can only add that they who wrought were our kindred; in a world grown much larger perhaps they would share the creator's joy in an admiring and informed audience, however unexpected, however ignorant.

NON-INVASIVE CLINICAL MEASUREMENT, edited by David Taylor and Joan Whamond. University Park Press ($39.50). The bats and the whales were



A pedal-power scroll saw of the 19th century, from **Pedal Power**

An abstract pattern on a Mimbres bowl, from **Mimbres Painted Pottery**

A representational pattern on a Mimbres bowl, from **Mimbres Painted Pottery**

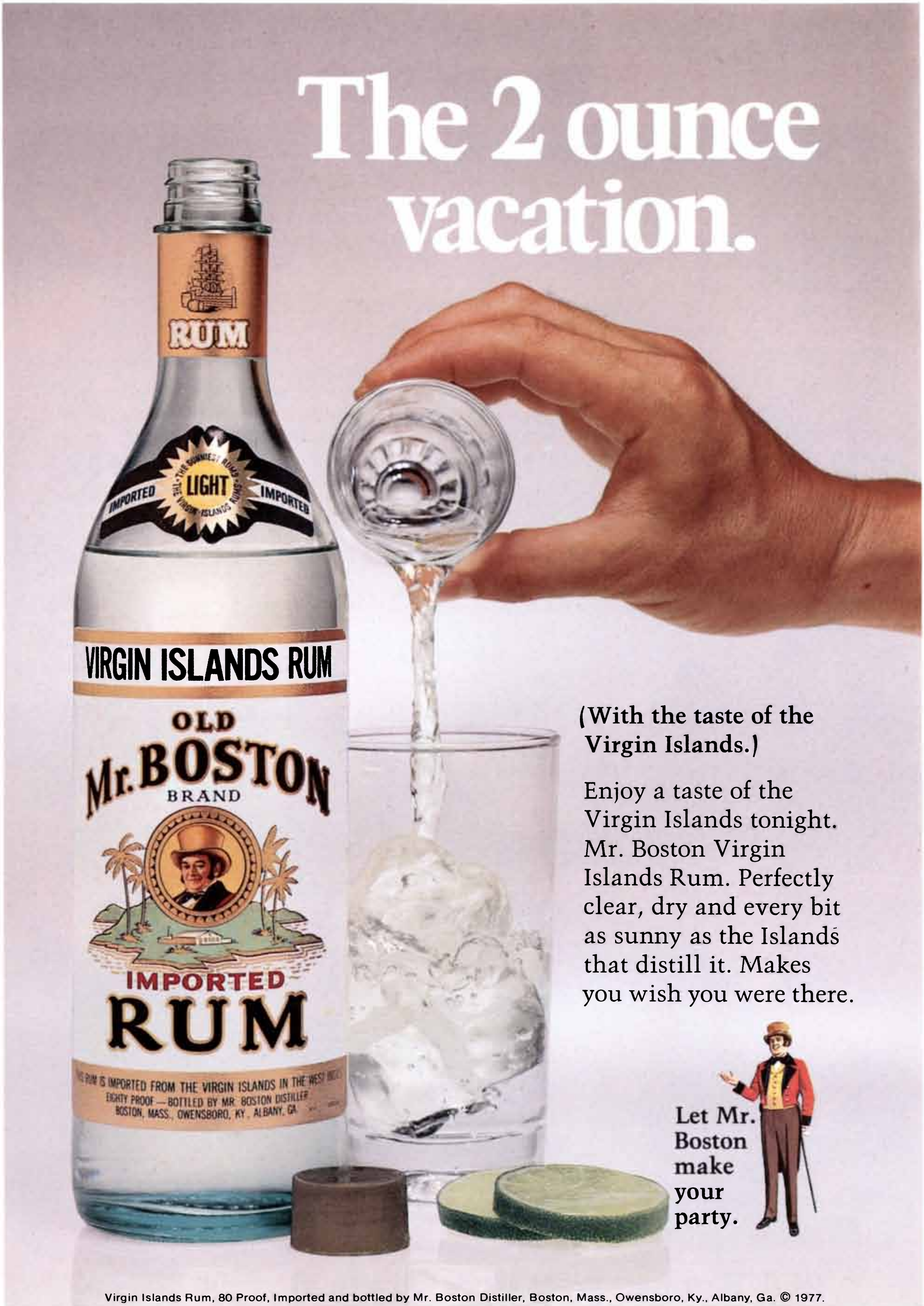
before us, but now we humans make routine use of pulses of ultrasound (or of microwave) to map the night or the depths. About half the papers in this experts' symposium let the reader in on the edge of recent progress in the sonar exploration of that most intimate and inner sea, the human body itself. Any physician who takes seriously his sworn undertaking at least to do no harm must question the damage secondary to many kinds of probing that are valuable for the rational care of the ill. The needlepoint, intrusive little tubes and the ionizing X-ray beam can carry risk and pain to the very patient their information is intended to help. Hence the "noninvasive" instruments described in this up-to-date collection of 15 detailed papers.

Ultrasound pulses at a few megahertz have for some years found application for mapping the living body from without. The *A*-mode scan of wartime radar parlance, in which the oscilloscope trace plots the delay time for the echo return against azimuth, was perhaps first. Today the *B* mode, with a scanning ultrasound beam whose echo intensity is signaled point by point to build a glowing two-dimensional image on the screen, is a decade old and much used. The skull of the fetus can be measured in utero, for example, and the date of delivery predicted with a standard deviation of some 12 days.

The authors are from bioengineering groups from Toronto to Bucharest. (There are no authors from the U.S.) Imaging is only part of the story. The penetrating sound wavelengths that are employed are only a few tens of times longer than the diameter of the red cells of the blood; the faint scatter from these particles moving with the flow is not hard to detect, and the Doppler shift of a few kilocycles is easily measured in the echo. The geometry of flow is complex: it is not easy to sight straight down the axis of the hidden aorta. Clever probe geometry with twin probes at right angles can give an angle-independent frequency shift. The superposition of a *B*-scan image with a *B*-scan Doppler looks still more promising. It requires much digital data handling by those remarkable microelectronic chips.

The flow profile within half-clogged blood vessels and everyday absolute measurement of flow speed and effective area are not far off. The ambiguity of range in ordinary pulsed Doppler systems, set by the conflict between the pulse-repetition rate and the Doppler shift, limits the measurable product of range and blood velocity. To get around this, batlike frequency-modulation schemes and even pulse-code modulation are being worked out, with a sophistication more familiar in the context of antisubmarine warfare than in the hospital ward.

The amount of oxygen carried in the


The 2 ounce vacation.
RUM
LIGHT
IMPORTED
IMPORTED
VIRGIN ISLANDS RUM
OLD
Mr. BOSTON
BRAND
IMPORTED
RUM
(With the taste of the Virgin Islands.)
Enjoy a taste of the Virgin Islands tonight. Mr. Boston Virgin Islands Rum. Perfectly clear, dry and every bit as sunny as the Islands that distill it. Makes you wish you were there.
Let Mr. Boston make your party.
Virgin Islands Rum, 80 Proof, Imported and bottled by Mr. Boston Distiller, Boston, Mass., Owensboro, Ky., Albany, Ga. © 1977.

You're gonna love our Great Little Car. Mazda GLC.

Great little piston engine. Great little goodies. Great little price.

REMOTE HATCH RELEASE

The 1978 Mazda GLC is a phenomenal automobile. It's not often a car of special merit comes along that doesn't cost a small fortune. It's not often a car comes along with this kind of equipment, this kind of styling, for this kind of money: $3595** for the GLC Deluxe 3-door Hatchback.

An electric remote control button under the instrument panel allows you to unlatch the hatch from the driver's seat. Or, you may open the hatch from the outside, when that's more convenient. Every Deluxe GLC has a rear wiper, washer, and defroster as standard equipment. A choice of 4-speed, 5-speed, or automatic transmission. And a rear seat that splits in half, so you can carry people and stuff, or just stuff. There's a lot more to a GLC than this of course. Yet it doesn't cost a lot of money — whether it's the GLC Deluxe 3-door Hatchback, or the GLC 5-door Hatchback, or the GLC Sport, or the GLC Standard model.

Mazda's Great Little Car is a great little car. You're gonna love it.

REAR WIPER/WASHER/DEFROSTER

SPLIT REAR SEAT

CHOICE OF TRANSMISSIONS

46 MPG H'WAY

35 MPG CITY*

*EPA estimates based on optional 5-speed transmission. Your mileage may vary depending on how and where you drive, your car's condition, and optional equipment. 42/33 mpg Calif.

**POE Los Angeles prices (slightly higher in California, and other areas). Taxes, license, freight, and optional equipment are extra. GLC prices for models not shown range from $3245 to $3920

mazda

GLC. Now, four great little cars in all, from $3245.**

arterial bloodstream becomes marginal during such crises as anesthesia and premature birth. Samples taken from time to time are hardly able to map the changes expected; coughing in a normal subject can induce changes of tens of a percent for a minute or two. Continuous noninvasive recording of the oxygen concentration in the blood is offered by one system described here. It is a recognizable outgrowth of the microscopic observation most of us who were once students can recall: red cells pulsing fitfully along the capillaries of the webbed foot of a frog. The blood gases of deep arterial flow are pretty accurately represented in blood brought to flow in the capillaries close beneath the skin by heat applied locally. Some oxygen diffuses through the intact skin. An elegant skin-contact probe, with its own little heater carefully controlled by thermistors, samples the leaking oxygen by means of a thin membrane of Teflon and cellophane. Within the probe a tiny silver/silver chloride and platinum cell polarographically measures the oxgyen tension, which is linearly related to the reduction current of a few nanoamps. The probe responds in some 10 seconds; it is easily calibrated, and a long series of clinical correlations is cited. The apparatus developed by the University of Marburg group writing here is already being made commercially.

There is a long way to go before this expertise yields reliable and inexpensive devices suitable at the bedside for the very ill. The senior editor, from the Royal College of Surgeons, offers a brief overview of the past and the future; the individual papers are technical, and each is of limited scope. We need a better method to image the fetus during the first trimester, "when abnormality should be diagnosed." Ultrasound is too coarse. Possibly microwave radar of very short wavelengths could be used, but the guarantee against teratogenic effects must be well-founded to justify microwave exposure at this most sensitive time. (No comment is offered here about the safety of ultrasound.) Preventive medicine demands more attention, particularly for the ills of the population under 40. Progress there needs long-term noninvasive mobile monitors, with tape storage of data or perhaps radio links. The pressure and flow of blood and the blood gases are the main data we now think we want; electrical signals from the heart and brain are already being monitored in ambulatory subjects. There is a challenge here that ought to go a long way toward offering the systems and the data-processing communities the moral equivalent of war.

NATIONAL DIRECTORY OF ADDRESSES AND TELEPHONE NUMBERS, edited and produced by Stanley R. Greenfield. Nicholas Publishing Co., Inc. Bantam Books ($9.95). The late Constance H. Winchell of Columbia University, to whom this paperbound volume about the size and shape of the Boston telephone book is dedicated, was a distinguished reference librarian. Experience led her and her colleagues to dream of new reference books; this concept was the first in their list of desires, a substitute for most of the inquiries brought to the shelf in any big library where the limp telephone directories slump.

These green pages offer more than 50,000 entries, each listed at least twice: once classified into some category and once more in a merged alphabetical listing more than 450 columns long. It is a software wonder in Bell System format, an American guide to organizations of every kind. Few individuals make it (some brokers); mostly it lists those artificial individuals we all interrogate, answer, hire, attend, buy from, serve or offer allegiance: the 10,000 corporations with annual sales of $10 million or more, Federal offices galore, the agencies of the 50 states and every county courthouse, the world of diplomats in the U.S. and those sent by the U.S. outward, colleges and graduate schools, those foundations with enough assets, hospitals with more than 300 beds, hotels and airports, bus and rail terminals, county medical societies and business, educational, cultural and fraternal societies of every stamp. The media are here in detail: 800 radio and television stations, two dozen hot-line data bases, nearly 600 magazines, publishers of books and daily newspapers with circulation above 50,000, plus two dozen groups of "ethnic newspapers" by national group. Museums, symphony orchestras, public libraries, commercial and public parks and one brave column of dance companies are to be found.

Not everything one wants is here, of course. Administrative headquarters usually stand for the whole: the National Aeronautics and Space Administration is in Washington but not at Goddard, Huntsville or Pasadena. Field stations and laboratories are given little space; Los Alamos is only a county and Kitt Peak National Observatory is missed, although its boss AURA is listed. The National Bureau of Standards is here, but not its Colorado center with Station WWV. The major outlying military installations in the U.S. are an exception; they are all here, some listed in every state save Iowa and West Virginia. The media are very thoroughly done, but such big industries as ordnance are omitted; you can find *Hustler* but not Hughes Aircraft. A dozen newspapers with the name *Times* are published daily, and they all carry stories about the hundreds of professional teams listed here by sport, from the Baltimore Orioles to the San Diego Friars.

The volume is strong on useful lists

drawn from national associations and from a variety of other reference works, which are cited. It is not quite free of error. The American Physical Society, for example, makes it under "Health Regimens," along with the National Nudist Council of Whitehouse, Ohio. Why is no U.S. embassy listed in Sri Lanka? If someone you know is busted in Cleveland, you might find it urgent to seek the professional services of Benny the Bondsman. The volume is surely useful, and it is probably irresistible to the curious.

OPTIMAL STRATEGIES IN SPORTS, edited by Shaul P. Ladany and Robert E. Machol. North-Holland Publishing Company ($20). The managers of General Motors, say, are familiar by now (at least vicariously) with linear programming, Monte Carlo simulation, Markov processes, optimization schemes and the rest of the arsenal of operations research, dear to planners in war and peace. But what about the managers of the Boston Red Sox and the Denver Broncos, not to say the Little League coaches?

A pioneer of sports operations research is Earnshaw Cook, whose Monte Carlo study on baseball was reviewed in these columns 10 years back. Here is Cook again, with a five-page guide to his earlier work set among a dozen other pieces on baseball strategy, scheduling and player evaluation. The book is an anthology of some 30 pieces, not very mathematical but somewhat so, on baseball, football, basketball, hockey, cricket, tennis, golf, field sports and more, mostly reprinted from a variety of journals, a few original here. All of them employ either quantitative operations methods or the more familiar methods of mechanical analysis to draw inferences about how to improve the game. Want a table for the most important points in a tennis match? Like to know where to take off for a long jump? (The marked foul line is not the best place to aim at.) Interested in a simple strategy for running a foot race, its parameters fitted well to the records, distance by distance? One author even offers another and more mathematical way to fail to find a strategy to beat the horses. (That work was supported in part by Victory Knox, which paid $91.50 for $2.) Anyone with a little pleasure in simple equations will be dazzled by a study cited from Thomas McMahon of Harvard. He extracted from a few lines of analysis the remarkable result that the speed of an n-man rowing shell increases as the ninth root of n: an eight-rower shell is only 25 percent faster than a single scull. It even fits the data pretty well.

There is a fine annotated bibliography, up to date to 1977. Play ball—by the numbers!


trimline

Ultimately It's Marantz.
Go For It.

Marantz goes beyond THD to lower TIM (transient intermodulation distortion).

Because TIM doesn't show up on conventional amplifier testing equipment, most manufacturers and their engineers aren't even aware that it exists in their amplifiers. Even if they were, they probably wouldn't know what to do about it.

But because Marantz builds for the music and not just the specs we know how destructive TIM can be to pure sound reproduction. And we've developed a revolutionary new circuit design to eliminate it.

The reduction of TIM can be the single most important element in making an amplifier sound better. For instance, two amplifiers with identical total harmonic distortion (THD) specifications should sound the same when compared...but the one with low TIM will sound audibly better! That's because TIM adds an unnatural harshness to the music. It's not only detrimental to pure sound reproduction, but it can have an emotional effect that you experience as "listening fatigue."

TIM is caused by an improper design of "negative feedback circuitry," by other manufacturers. Every modern amplifier uses it to lower THD. But *excessive* negative feedback coupled with an insufficient slew rate* can lead to gross internal overloads under the constantly changing transient and sound levels of music. That distortion is TIM.

The gentle slopes of continuous sine wave test signals normally used to test an amplifier simply cannot detect TIM distortion. It requires the type of extremely sophisticated spectrum analysis equipment developed by the space industry to analyze radio frequencies.

Our answer to TIM is a circuit design that ensures the widest bandwidth and the lowest obtainable THD *before* negative feedback is applied. The Marantz 170DC Stereo Power Amplifier (the 1152DC and 1180DC also use this circuitry), for instance, needs only 1/100th (− 40 dB) the amount of negative feedback commonly required by other amplifiers to yield the same low THD figures.

Incredibly, Marantz amplifiers with low TIM design can deliver flat frequency response from 0 Hz to 20 kHz *without* the use of negative feedback. And this same circuit design provides the optimum slew rate for minimum TIM and maximum reliability.

Result: Marantz reduces TIM to an inaudible level, which means you get clear, clean sound from all your records and tapes. Think of Marantz with low TIM as a window to the original performance.

If you truly want the best reproduction of musical sound available anywhere—and are willing to spend a little more to get it—then go for it.

Go for Marantz.

25th Anniversary marantz®

*The maximum change of voltage per unit time. © 1978 Marantz Co., Inc., a subsidiary of Superscope, Inc., 20525 Nordhoff St. Chatsworth, CA 91311. Prices and models subject to change without notice. Your Marantz dealer has the full line of Marantz amplifiers. Look for him in the Yellow Pages.

Established 1845 SCIENTIFIC AMERICAN April 1978 Volume 238 Number 4

Nuclear Power, Nuclear Weapons and International Stability

Irresolution over domestic energy policy and the role of nuclear power may act to undermine current U.S. efforts to control the proliferation of nuclear weapons

by David J. Rose and Richard K. Lester

A year ago this month the Carter Administration put before Congress a comprehensive national energy plan that included as one of its key components a revision of this country's long-standing policy on the development of civilian nuclear power. The proposed change, which would have the effect of curtailing certain aspects of the U.S. nuclear-power program and of placing new restrictions on the export of nuclear materials, equipment and services, was based explicitly on the assumption that there is a positive correlation between the worldwide spread of nuclear-power plants and their associated technology on the one hand, and the proliferation of nuclear weapons and the risk of nuclear war on the other. This point of view has become the topic of a lively debate; at the periphery of opinion some see nuclear war lurking behind every reactor on foreign soil, whereas others argue that the connection between civilian nuclear power and nuclear-weapons proliferation is vanishingly small.

We shall advance here the heretical proposition that the supposed correlation may go the other way, and that the recent actions and statements of the U.S. Government have taken little account of this possibility. In brief, it seems to us that if the U.S. were to forgo the option of expanding its nuclear-energy supply, the global scarcity of usable energy resources would force other countries to opt even more vigorously for nuclear power, and moreover to do so in ways that would tend to be internationally destabilizing. Thus actions taken with the earnest intent of strengthening world security would ultimately weaken it. We believe further that any policy that seeks to divide the world into nuclear "have" and "have not" nations by attempting to lock up the assets of nuclear technology will lead to neither a just nor a sustainable world society but to the inverse. In any event the technology itself probably cannot be effectively contained. We believe that the dangers of nuclear proliferation can be eliminated only by building a society that sees no advantage in having nuclear weapons in the first place. Accordingly we view the problem of the proliferation of nuclear weapons as an important issue not just in the context of nuclear power but in a larger context.

Fundamental tensions exist between the energy objectives and the nonproliferation objectives of U.S. policy, and on a different plane between the respective consequences of measures designed to achieve their primary effect either domestically or internationally. In what follows we shall analyze the complex set of interrelated issues that bear on the entire question of nuclear power and world security.

The most important of the new nuclear measures announced by the Administration last April were that the U.S. would defer indefinitely the reprocessing of spent nuclear fuel from domestic nuclear-power plants to recover and recycle plutonium and unused fissionable uranium, and that it would try to persuade other nations to follow its lead. Legislation submitted to Congress demonstrated the Administration's intention to restrict exports related to the nuclear-fuel cycle and to prevent the retransfer of exported U.S. nuclear technology to third parties. (A modified version of the Adminstration's nonproliferation bill has since then been approved by Congress.) Along with these restrictions the U.S. capacity to enrich uranium to standards suitable for use in conventional light-water (as opposed to heavy-water) reactors was to be increased to help meet the growing world demand for this service.

On the domestic front the regulatory requirements for installing light-water reactors were to be streamlined. The Administration also proposed a substantial slowdown of the U.S. program to develop a liquid-metal-cooled fast breeder reactor, a type of nuclear-power plant designed to create and consume fissionable plutonium out of the vast store of ordinary, nonfissionable uranium in the earth's crust. Specifically, the Clinch River breeder reactor, which was to have been built in Tennessee at a cost of some $2 billion, was scheduled for cancellation. The operation of breeder reactors requires reprocessing the nuclear fuel periodically, so that the retreat from fuel reprocessing and the deemphasis of breeder-reactor development complement each other. (Unfortunately the Clinch River reactor has become a focal point of the debate for both the critics and the proponents of nuclear power. The situation is doubly unfortunate because on the one hand that particular program was technologically and institutionally vulnerable and on the other the stopping of it has not helped resolve the deeper issues we discuss.)

The U.S. was also to redirect its nuclear research and development programs to place more emphasis on alternative fuel cycles and reactor designs that might offer reduced access to material suitable for use in nuclear weapons. This initiative has been carried into the world arena with the establishment of an international program to evaluate alternative technical and institutional strategies for the nuclear-fuel cycle.

In the months since the Administration's program was announced it has provoked much discussion within the U.S. and throughout the world. It has dissatisfied both critics and supporters of the U.S. nuclear-power program, and (partly because of the way it was presented) it has generated concern in many foreign capitals. Some of this country's partners in the development of nuclear power feel that they were not consulted adequately during the genesis of the new policy, and that policy communications, at least initially, have been clumsy and insensitive. Deeper-rooted anxieties underlie this irritation, however, since fuel enriched in the U.S. and reactors manufactured by U.S. companies still play significant roles in many national nuclear-power programs, and the effects of U.S. nuclear policies are widely felt.

What was the motivation behind the Administration's new nuclear policy and the related Congressional actions? Several possibilities come to mind. The first is simply that the Administration means what it says, namely that its goal is to increase international stability by taking actions thought to inhibit the proliferation of nuclear weapons. It would do so by reducing the availability of nuclear materials and technology helpful to a weapons program, even though the same materials and technology had hitherto been commonly assumed to be a part of civilian power programs.

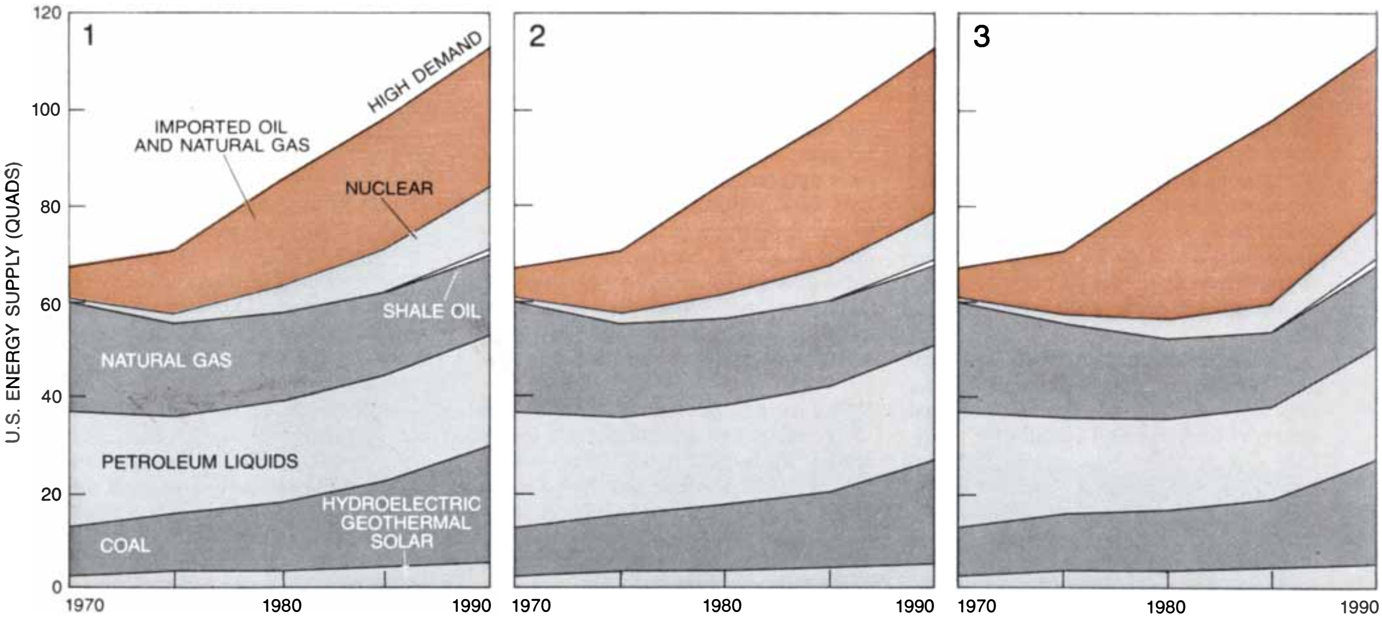
Another possibility is that both the Administration and Congress are undecided as to whether the collapse of the U.S. nuclear industry is desirable. This indecision contributes to the Government's apparent inability to formulate a coherent nuclear-energy strategy.

It is also possible that the Administration announced its policy in sympathetic response to the critics of nuclear power but expects that the policy will not work, and that after this demonstration of good faith a new, more pragmatic program will be unveiled at an appropriate time. One danger in such a tactic is that the Administration might delay its denouement too long. The reaction to an Administration generally perceived to be resigned to the demise of nuclear power in the U.S., or even actively to desire it, might develop its own irreversible momentum.

It may also be that the Administration, frustrated by the diplomatic rigidity of world discussions about international security and arms limitation, is casting about for some new approaches. Such approaches are to be found; we hope that this article is an example. These four general motivations are not mutually inconsistent, and the Government could shift its priorities among them as the consequences of its actions unfold.

Various circumstances created the context in which the Administration formulated its national energy program and in which its set of nuclear-energy and nuclear-nonproliferation goals were to be pursued. The first circumstance was the urgent need to reduce the growing outflow of U.S. funds, currently running at more than $40 billion per year, to pay for imported petroleum. Such an expenditure, although not liable to bankrupt the country, was inconceivable as recently as 1973. Furthermore, the situation suggested excessive dependence on the policies of the principal oil-supplying nations. The increasing dependence on oil imports was (and is) seen as probably the major contributor to the U.S. energy problem.

The second circumstance was the presence of vast coal deposits in the U.S., amounting to perhaps 10 times the entire domestic oil and gas resource. Only 19 percent of the U.S. energy supply comes from the 650 million tons of coal currently mined each year. Estimates of the amount of coal recoverable at current prices with current technology range from 200 to 600 billion tons; thus even if coal production were to be tripled, the minimum estimate amounts to a 100-year supply. Moreover, the total amount of coal in the U.S. might be as much as 3,000 billion tons recoverable eventually at increased cost—a quarter of the earth's known reserves. Besides coal, an energy resource equivalent to perhaps 1,000 billion tons of coal resides in the oil-shale deposits in the vicinity of northwestern Colorado. Technological advances would surely make many of these resources available later, it was thought. In no absolute sense is the U.S. "running out of energy." Thus arose the goal of increasing coal production rapidly, to several times the present rate by the year 2000. In this



ALTERNATIVE U.S. ENERGY PROJECTIONS through 1990 are outlined in the set of graphs beginning on these two pages. The graphs represent 12 possible demand-and-supply scenarios constructed by the staff of the Congressional Research Service on the basis of different assumptions about economic growth rates, energy prices, the elasticity of energy demand and the constraints on various energy supplies. The tables from which the graphs were drawn were compiled in the course of a two-year study conducted at the request of several committees of Congress responsible for dealing with energy-related issues; the data appeared originally in *Project Interdependence: U.S.*

way the U.S. could reduce its dependence on imported petroleum, and perhaps also afford a more leisurely nuclear program.

The third circumstance was that energy conservation began to be taken seriously. Many studies under way between 1975 and 1977 showed not only that substantial increases in energy efficiency were possible but also that much energy was being wasted. Energy conservation had received significant recognition but little actual support during the previous Administrations, and Congress had not been overactive compared with what could have been done.

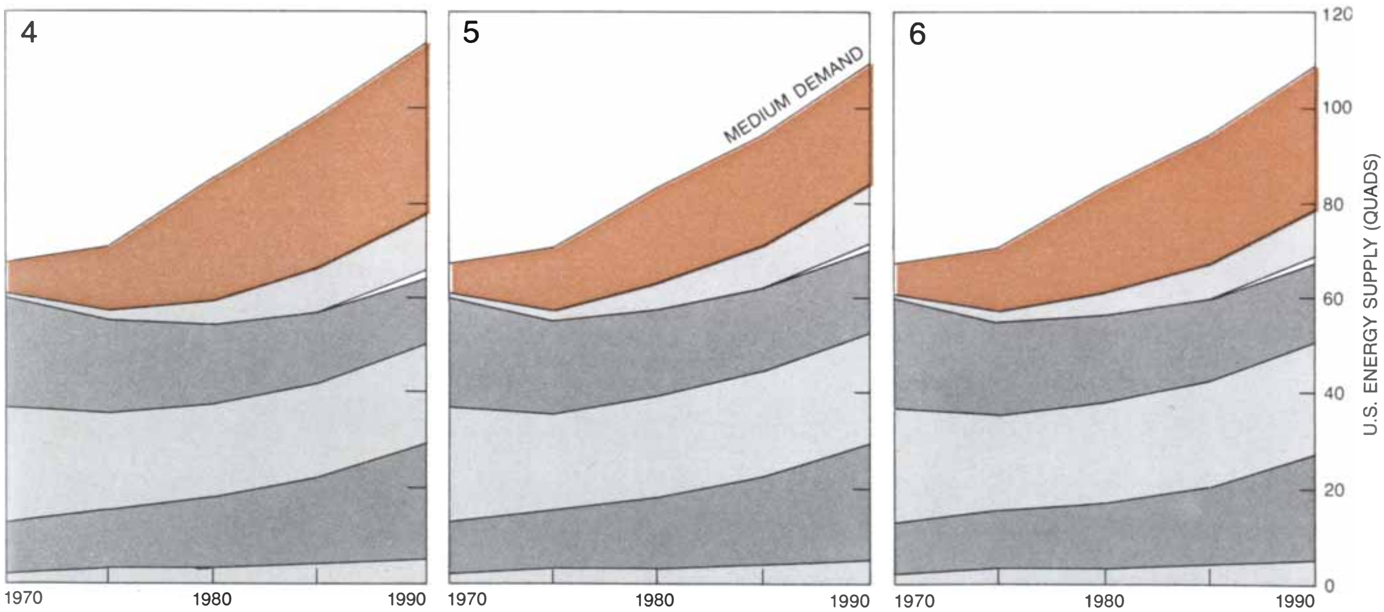
Several arguments had been marshaled against conservation, the main one being that economic activity and energy use were closely bound; hence restricting energy use would probably exacerbate a recession or cause one. In the short term energy and economic activity are indeed closely bound, because machines use energy, and they cannot be replaced overnight. By replacing more energy-intensive machines at the end of their life span with more energy-frugal ones, however, the energy demand could be cut in a matter of decades by 1 or 2 percent per year from what would have been otherwise forecast. With an economic growth rate of 3 or 4 percent per year, energy use might then grow at only half that rate; by the year 2000 the gross national product would have almost tripled, and the energy used per unit of economic output would decline to about 60 percent of its present value. Even so, domestic energy use would have increased by a factor of approximately 1.6, through the diligent exploitation of coal, solar power, light-water reactors and perhaps other technologies. (These numbers are meant only to indicate what many energy planners thought would be possible.)

The fourth major circumstance relates to several aspects of nuclear power itself. First, the U.S. industrial capacity to make light-water reactors is large—perhaps too large. A substantial part of this capacity would be needed to produce some of the base-load electric-generating plants, leaving coal for other electric plants and many other uses. Second, the nuclear-power industry, beleaguered by critics of many persuasions and by a Nuclear Regulatory Commission that it had come to regard as increasingly demanding, also needed some organizational relief. Thus arose the goal of simplifying procedures for fulfilling siting and other licensing requirements. The light-water-reactor industry was to be encouraged by these activities, and electric utilities would be encouraged to "go nuclear" by building light-water reactors wherever such plants were economically attractive.

The fifth circumstance relates to the uranium resources, particularly in the U.S., with which to fuel all those light-water reactors. Each reactor that produces 1,000 megawatts of electric power requires about 5,500 tons of uranium (in the form of uranium oxide) to operate during an expected 30-year life span. The Administration knew that the equivalent of about 680,000 tons of uranium oxide had been located in deposits in the U.S. with characteristics that would make economic recovery possible with current technology, together with an additional 140,000 tons that will be available between now and the end of the century as a by-product of other mineral-extraction operations. It also estimated that roughly another three million tons would be found when it was necessary and that this amount could be produced at a cost of $50 per pound or less. All this uranium would fuel about 700 reactors for their full life span or an even larger number if a full lifetime commitment of fuel were not made for each plant when it began operation.

Considering also that nuclear-power stations take 10 years or more to build and that orders would increase gradually, the Administration judged that adequate nuclear fuel would be available to last several decades into the next century. All this could be done without reprocessing spent fuel from the reactors. Besides, a number of studies had shown that the recycling of uranium and plutonium in the current generation of light-water reactors would at best be only marginally attractive economically and could in fact result in higher fuel-cycle costs than a "once through" fuel cycle.

This brings us to the sixth circumstance: the Administration's concern about the connection between the technology of nuclear-fuel reprocessing and the development of a nuclear-weapons capability. The past few years have brought increasing doubt about the ability of international safeguards to function satisfactorily in a "plutonium economy," that is, one in which large amounts of plutonium would be present at various stages of the fuel cycle in comparatively extractable form. The objective of international safeguards is



and World Energy Outlook Through 1990, **a 939-page report published in November by the U.S. Government Printing Office. The first six scenarios in the set, depicted in the graphs on these two pages, are characterized as follows: high demand, high coal and nuclear supply (*1*); high demand, medium supply (*2*); high demand, low supply (*3*); high demand, low oil and gas supply, high coal and nuclear supply (*4*); medium demand, high coal and nuclear supply (*5*); medium demand, medium supply, also referred to in the study as the "base case" (*6*). All figures are given in "quads": quadrillions (10^{15}) of British thermal units. The graphs are continued on the next two pages.**

to detect the theft or diversion of nuclear material by nations early enough for diplomatic or other international countermeasures to achieve their objective before the material can be made into an explosive. It was argued, however, that plutonium that had been recovered by spent-fuel reprocessing and then recycled could be turned into an explosive so rapidly after national diversion of the recycled fuel that the ability of the safeguards system to work adequately would be fatally undermined, even if the loss of material were detected. International safeguards do not prevent diversion; they deter it through the threat of timely detection. If plutonium were adopted as a commercial fuel, the deterrence effect of safeguards would be lost. In addition, the presence of plutonium would increase the risks of nuclear terrorism, and there were unresolved questions about the effectiveness of the predominantly national safeguards that would be introduced to deal with this threat.

If, on the other hand, there were no reprocessing, there would be no plutonium either to fuel breeder reactors or to make plutonium-based nuclear weapons; a nation that did not reprocess the fuel from its nuclear-power reactors could not then imperceptibly slip into the position of having a nuclear-weapons capability, and it could not in some temporary passion easily pervert a civilian nuclear-power program. Either it would have to extract almost pure fissionable uranium 235 from natural uranium, an activity that is associated with nuclear weapons and not at all with conventional light-water or heavy-water reactors, or it would have to flagrantly set about reprocessing used reactor fuel to extract plutonium. Furthermore, both of these activities are widely thought to be beyond the present capability of any subnational group acting clandestinely.

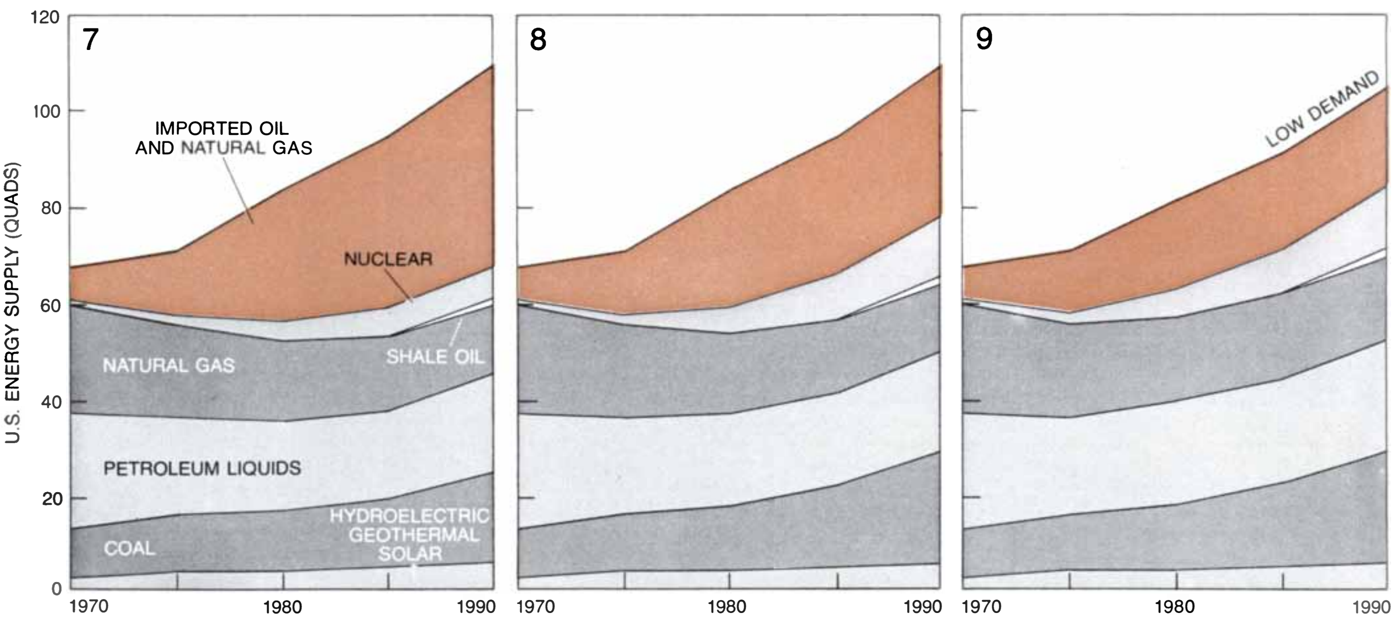
If the U.S. was mainly worried about the international proliferation of nuclear weapons, why then did it stop the domestic reprocessing of nuclear fuel? The best answer seems to be that in order to argue its case persuasively, it would have, so to speak, to come to court with clean hands.

To be sure, breeder reactors would be delayed, perhaps indefinitely if some better prospect such as economic central-station solar power or controlled fusion came along. Clearly the latter options were to be encouraged. Meanwhile even the U.S. breeder program was to benefit because there would be time to explore a more varied set of technological options, both conceptually and through experiments on a modest scale. Far from being canceled, the breeder program would take on a needed diversity; perhaps more nearly proliferation-proof fuel cycles could be found.

Many doubts about the public acceptability of nuclear power had built up, and the new goals would surely be seen as being responsive to those doubts: no plutonium, no reprocessing, no breeder reactor in this century and so forth. Meanwhile little public concern had yet arisen over the social, environmental or health hazards of coal, the energy-supply option the Administration planned to promote most vigorously.

The Administration's perception of this complex issue can be analyzed with the aid of a sequential logic diagram, which is useful in clarifying some of the main proliferation-related trains of thought and their impact on international security [*see illustration on pages 54 and 55*]. Two main paths appear in the diagram. First, there is a horizontal decision path that could be followed by a nation (let us call it *Y*) that does not now possess the nuclear technologies in question. The central questions affecting international stability are whether or not nation *Y* decides to develop a general capability with respect to nuclear weapons and, once it has decided to do so, how long it would take to acquire that capability. The second path consists of several vertically arranged inputs to *Y*'s decisions. The U.S. sees the capability to reprocess nuclear fuel as a stimulant to weapons proliferation because *Y* would then have a source of weapons-grade fissionable material; such accessibility might be instrumental in a decision by *Y* to acquire nuclear weapons, and in any case once the decision had been made the reprocessing capability would help *Y* to move toward the right in the diagram: toward the weapons themselves.

The sequential arrangement of events depicted here is too simple. A decision to acquire nuclear weapons might be a prolonged process, and it could be made in parallel with (or could even follow) the acquisition of some of the technological components of a civilian nuclear-power program, which would also be useful for the development of a weapons system. For example, that would be the case with a commercial reprocessing



U.S. ENERGY ALTERNATIVES are continued on these two pages. The characterizations of the six remaining scenarios in the Congressional Research Service study are as follows: medium demand, low supply (*7*); medium demand, low oil and gas supply, high coal and nuclear supply (*8*); low demand, high coal and nuclear supply (*9*); low demand, medium supply (*10*); low demand, low supply (*11*); low demand, low oil and gas supply, high coal and nuclear supply (*12*). The authors of the *Project Interdependence* report point out that even with a low energy-demand projection coupled with high coal use, an expanded nuclear-power capacity, an increase of about 50

plant, assuming that in its normal operating procedure it produces plutonium separated partially or completely from the other constituents of irradiated fuel. Such a reprocessing plant would therefore be ambiguously perceived by observers, even if that were not the intention of a peaceful nation *Y*. The Administration saw that by blocking the connection at the top of the figure a substantial barrier would be erected against *Y*'s either sidling consciously or sliding unconsciously into a technological competence applicable to the manufacture of weapons.

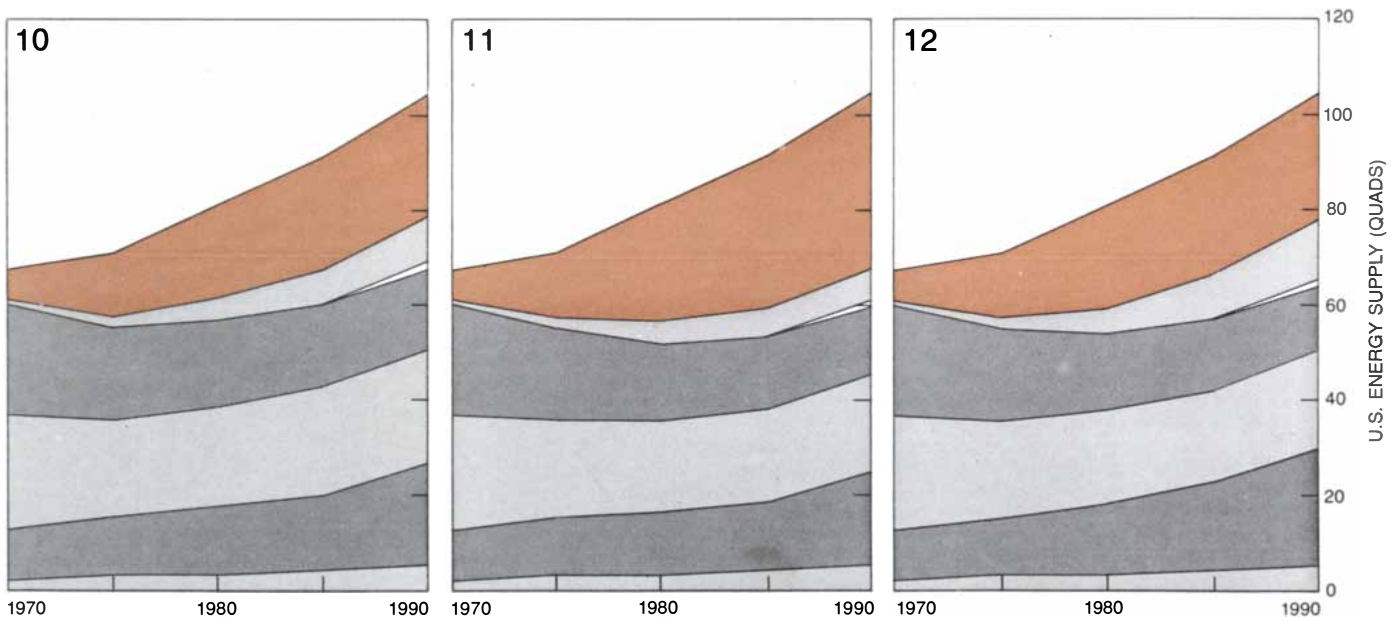
Of course, the Administration realized that other routes exist whereby *Y* could obtain fissionable material suitable for weapons. First, it might import the necessary technology from elsewhere, but for some time the U.S. has been actively attempting to close those routes by seeking to persuade the other major suppliers of nuclear materials and equipment (through bilateral channels and in the multilateral forum of the London Suppliers Group of nuclear exporting nations) to exercise restraint in the transfer of "sensitive" items that might offer increased access to weapons-grade material.

All the other routes involve a conscious decision by nation *Y* and a substantial effort on its part. It could develop its own civilian fuel-reprocessing technology, fully intending peaceful uses only, then have it subverted later after a change of attitude on the part of its government. Alternatively, it could attempt the production of weapons-grade plutonium in research reactors (as India did for its 1974 nuclear explosion) or in a small clandestine reactor, in either case recovering the plutonium from the irradiated fuel in a small reprocessing plant built expressly for the purpose, a task much easier (and cheaper) than the development of the technology and the construction of the plants for commercial reprocessing. Still another alternative could involve diverting irradiated commercial fuel to a clandestine reprocessing plant from a temporary storage facility, where it might have been awaiting either commercial reprocessing or, in the case of a once-through fuel cycle, ultimate disposal.

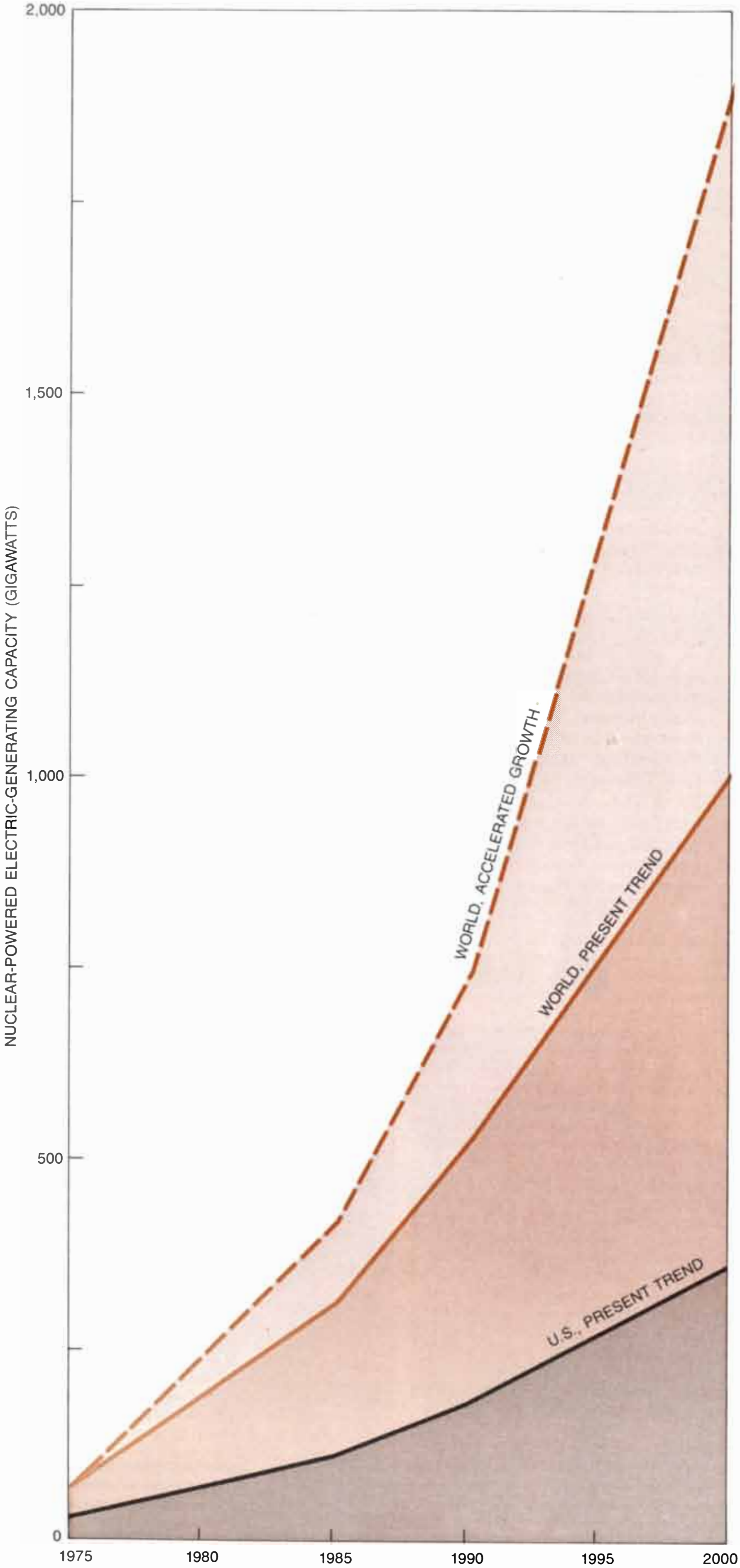
Rather than work with plutonium from spent fuel *Y* might attempt to extract the fissionable isotope uranium 235 either from natural uranium (which contains less than 1 percent U-235) or from the low-enriched uranium used for power-reactor fuel (about 3 percent U-235), concentrating it to, say, 90 percent. For this approach there are several candidate technologies at various stages of development. For the past 25 years practically all enrichment, either for power-reactor purposes or for the production of weapons-grade uranium, has been carried out in the huge gaseous-diffusion plants of the U.S. and the U.S.S.R. Gaseous-diffusion plants do not need to be built on such a heroic scale, however; furthermore, other enrichment technologies now challenge the dominant position of gaseous diffusion. Ultracentrifuge enrichment is being actively developed in several countries and is on the verge of commercialization in some of them. It requires less power than gaseous diffusion and has other advantages as well. These factors, together with the greater operational flexibility of centrifuge plants, suggest that this technology would offer a smoother path to weapons material, either with a plant built specifically for that purpose or through a facility built initially to fulfill civilian nuclear-power needs. It would be much easier to carry out the adjustments necessary to convert a gas-centrifuge plant from a low-enriched product to a high-enriched military-grade one than it would to similarly convert a gaseous-diffusion plant; alternatively a small string of centrifuges can be used over and over again progressively to enrich single batches of uranium.

Other enrichment technologies that are gaining in importance include the aerodynamic, or nozzle, approach, variations of which are being developed concurrently in West Germany and South Africa, and the laser technique for isotope separation, which is also being pursued in several countries. A pilot aerodynamic-enrichment plant in South Africa may have already been used to produce enough weapons-grade uranium for one or more explosive devices. Furthermore, although work on laser enrichment has so far been limited to laboratory research, enough information has been made available to suggest that the technique might ultimately provide a cheaper and more flexible route than any other enrichment process. All the known methods, at whatever stage of development, require sophisticated technology but nothing that is beyond the capability of many of the more advanced developing nations.

Turning to seemingly more bizarre ac-



percent in additions to domestic natural-gas reserves and a 100 percent increase in additions to domestic oil reserves compared with the preceding decade (*scenario No. 9*) the U.S. would still have to import close to 20 quads of energy by 1985, equivalent to almost nine million barrels of crude oil per day. On the other hand, if the rate of economic growth is greater than the 3.5 annual increase in gross national product projected in the base case and all areas of domestic energy supply turn out to be less productive than expected (*scenario No. 3*), the U.S. could be importing 17.7 million barrels of crude oil per day in 1985. The role of imported oil as a "swing" fuel (*color*) is evident.



tivities, country *Y* might employ agents to steal material from abroad or buy it in a black market, an open market or an intermediate "gray" market. It could receive such material as a gift or a loan from another government. It might steal an assembled weapon or even be given one. None of these activities can be excluded, and some may be more likely to occur than those we described above.

A really effective barrier to weapons proliferation would involve blocking all the lines marked with a black bar in the diagram. That is impossible, but the U.S. is only trying to make weapons proliferation substantially more difficult. Combining technological denial with an assortment of incentives related to the supply of enrichment services for light-water-reactor fuel (to be used by *Y* under tightly specified conditions) would, the Administration thought, significantly increase international stability.

Several other decision paths of a different nature exist, on which the U.S. Government has only an indirect influence; the principal ones are marked with gray bars. Why does *Y* decide to acquire a weapons capability in the first place? Why does it actually build nuclear bombs? Why might it use them? The answers to such questions depend on many things, and a long-term policy not to proceed at any one of these decision points makes all the technological elaborations irrelevant. The converse, however, is not necessarily true. As we implied when we were discussing reprocessing plants, a decision to proceed might be heavily influenced by the technological capability in place at that time.

When the U.S. policy described here is viewed from other countries or other domestic vantage points, it looks quite different. The main line of devel-

POTENTIAL TRENDS in the growth of nuclear power are shown through the end of the century for both the U.S. and the world (excluding China, the U.S.S.R. and the other countries of Eastern Europe). For the world the lower estimate is based on present trends in energy utilization and supply, including delays in the construction of new nuclear reactors, and assumes a continuation of these trends. The "accelerated growth" estimate assumes that the goals of ambitious nuclear programs will be met and that the world will return to higher rates of energy growth. The "present trend" estimate must be regarded as the more realistic of the two and may itself be too high. The data for these curves were obtained from a recent joint report by the nuclear-energy agency of the Organization for Economic Cooperation and Development (OECD) and the International Atomic Energy Agency (IAEA). U.S. projection is based on recent estimates by the Department of Energy. All such estimates must be viewed with caution, the authors point out, in view of the many uncertainties discussed in their article.

opment—in particular decisions by nation *Y*—can best be discussed in relation to a second diagram, which is similar to the first but starts with a U.S. decision to act restrictively and includes several additional logic paths. To understand its full significance we must start farther back, with the electric-power sector.

The U.S. electric-power industry abhors uncertainty about the future, for several reasons. One is the long time needed to construct new facilities (10 years is typical) and long expected life of these facilities (40 years, say). The industry also needs a stable fuel supply and is required by law to provide reliable service. In this regulated industry justifiable costs can be charged to the consumer. The present program of the Administration increases uncertainty about the future of nuclear power for several reasons. First, the decision against reprocessing spent fuel has raised fears, not yet completely alleviated by the Federal Government, that the electric utilities may in practice be left holding the spent fuel for a long time (for example by long-drawn-out court challenges to the Environmental Impact Statement for a Federal spent-fuel storage facility), a very unappealing prospect to them.

In addition, conflicting Federal opinions about the acceptability of nuclear power make the electric utilities both suspicious of Government motivations and better targets for anti-nuclear-power groups. Many electric utilities also fear that the Nuclear Regulatory Commission will order expensive, and in their view capricious, retroactive modifications to existing nuclear plants, in spite of current efforts to modify the Commission's legislative foundation. Many experts, concerned at the inherent uncertainty associated with the four-million-ton estimate for U.S. uranium resources that could ultimately be made available, feel that a "prudent planning estimate" for the purpose of setting nuclear-power policy should be appreciably lower. A National Academy of Sciences resource-evaluation group recently estimated that 1.8 million tons is all that is likely to be mined in the U.S. by the year 2010 at a cost of $30 per pound or less, even with a Government policy of maximum stimulation.

Furthermore, doubts have been expressed as to whether the U.S. uranium-supply industry, itself troubled by uncertainty about the size of the market for its product, will be prepared to invest in exploration, mining and ore-processing-plant construction at levels that will be sufficient to fuel a growing number of nuclear-power plants. Part of the uncertainty permeating the electric-utility sector stems from concern over the availability of nuclear-fuel supplies, so that the problem exhibits circular characteristics; it is also aggravated by the fact that the strength of the uranium supply industry's commitment to keeping power reactors adequately fueled is less than the utilities might find desirable. For example, it has been estimated that within a few years petroleum companies will own about 40 percent of all U.S. ore-processing capacity and as much as 50 percent of low-cost U.S. uranium resources. In short, the uranium suppliers do not constitute an industry "captive" to the electric-power sector. (Indeed, increasing corporate diversification in the various energy supply industries has led to the suggestion that the reverse might be true.) As a result uncertainty arising from the Administration's program may be compounded, unwittingly or otherwise. Similarly, each U.S. manufacturer of nuclear reactors has 75 percent or more of its business elsewhere (for example in other power systems), and the nuclear business is not essential to it. In a period of rising costs, large-scale cancellations of orders and excess production capacity, the business appears less than inviting.

The last point is worth further comment. The U.S. manufacturers of light-water reactors could turn out between 20 and 30 nuclear-power systems a year. The Administration has estimated that full implementation of its national energy policy would lead to an installed nuclear electric capacity of more than 300,000 megawatts by the year 2000. With 50,000 megawatts already installed, and a further 25,000 megawatts scheduled for completion by 1980, there would be, on the average, a dozen or so reactor systems completed each year for the last 20 years of the century, a situation that implies a large-scale restructuring of the reactor-manufacturing industry sooner or later.

Compounding these difficulties, the electric-utility sector suffers the additional one of raising enough capital. One cause is the host of uncertainties we have described. Another involves the general flight during recent inflationary times from long-term investment; the rate of economic return from the regulated utility industry has become unattractive, a circumstance that also affects other generating plants, particularly those that burn coal.

The upshot of all this is the paradoxical situation that although the existing nuclear reactors run pretty well and deliver economical electric power in many parts of the country, the nuclear industry in the U.S. may nonetheless be close to collapse. The proximate cause is a movement by the electric-utility industry and manufacturers away from nuclear power as they attempt to reduce their own institutional uncertainty, but deeper causes drive these changes and are coupled with the Administration's attitude toward nuclear power. This train of thought points toward the second possible motivation mentioned above: that the Administration, through internal indecision, is incapable of acting to prevent the nuclear industry from collapsing. The indications, however, are ambiguous.

What then is the U.S. electric-utility sector likely to do? The conventional option is coal, with the Administration's apparently enthusiastic backing. Oil and natural gas are expensive and in such uncertain supply that the Administration has submitted legislation prohibiting all new power plants from burning them, with only limited environmental and economic exceptions. Other legislative provisions would, through taxation and prohibitive clauses, encourage utilities not to burn oil and gas in existing facilities and to convert them to coal.

But what if the coal cannot be mined, transported and burned in time, and in socially accepted ways? Even before the Administration announced its national energy policy widespread doubts had grown about the wisdom, or even the possibility, of increasing coal production very quickly. In particular, can a goal of increasing coal production from its current level of 650 million tons per year to a projected total of 1.25 billion tons per year by 1985 actually be achieved? Industrial problems associated with such an expansion, land-use problems, states' policies and obstacles created by the Administration's environmental policy for coal have all been repeatedly raised as evidence to suggest that there will be hesitation on both the supply and demand sides of the coal industry. Even coal transportation, which currently accounts for about 30 percent of the U.S. rail tonnage, will be difficult for the disheveled railroad industry.

The environmental problems with coal appear to grow with time and increased understanding. The comparatively large amounts of disturbed land, the chemically and biologically active complex molecules present in coal and produced by the burning of it, and the ubiquitous nature of these effects create difficulties at local and national levels. On a global scale potentially the most serious long-range environmental impacts resulting from the large-scale burning of coal (or indeed of any fossil fuel) may arise from the effects of the increased concentration of carbon dioxide in the atmosphere. At this stage the problem is not well understood, and the potential contribution of planned U.S. coal-burning activities is therefore also shrouded with uncertainty. Nevertheless, in this problem area and others the prognosis looks more serious as more information accumulates. The U.S. electric-utility sector, generally aware of these difficulties, looks on coal with increasing anxiety.

Irresolution about nuclear power, increasingly apparent difficulty with coal, a partial ban against oil and a half-hearted attitude toward energy conservation make an impossible combination; something has to give. If the electric-utility industry waits a few years for public debate to resolve these issues, the concomitant pressure for rapid and comparatively pollution-free installations will drive it toward oil-burning plants. That would be doubly disastrous, because the conversion of some transportation, industrial, commercial and domestic systems from oil to efficiently used electric power, based on coal or nuclear fuels, is seen as a way of reducing oil imports. If the electric utilities are unable or unwilling to provide the means for this substitution, oil consumption will continue to exceed Administration targets.

Thus imported oil may once again fill the role the Administration has sought to prevent it from playing: the "swing" fuel that satisfies increased energy demands. Reinforcing this trend is the present concentration on increasing domestic production, which will have the effect of keeping up oil-based activities that must surely in the next decade or two be fed by imports.

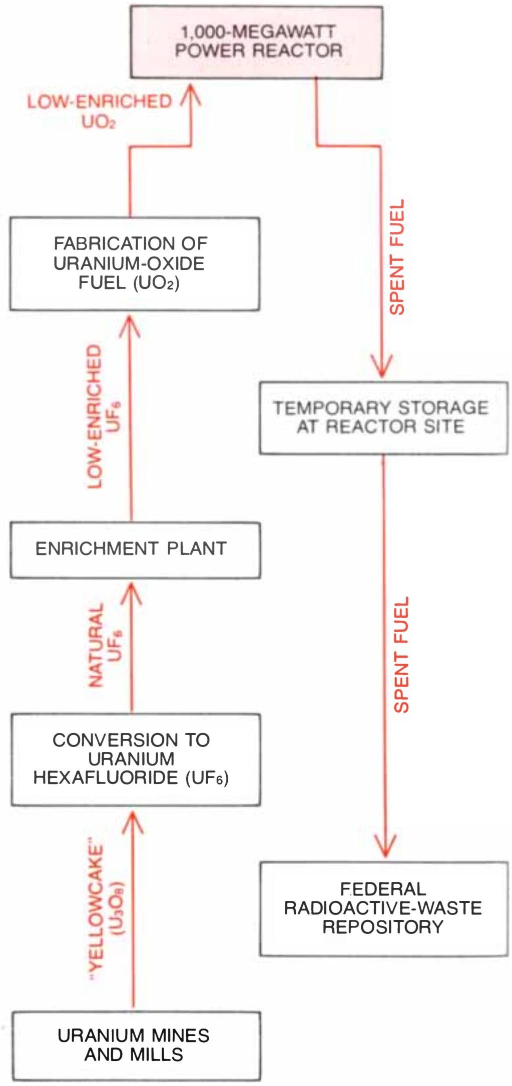
The international importance of President Carter's attempts to reduce oil imports, and the dangers implicit in failing to do so, cannot be overemphasized. Today the U.S. imports nearly half of its immense consumption of 17 million barrels a day, an amount equal to almost a quarter of all internationally traded oil. A reduction to the Administration's stated import target of less than seven million barrels a day by 1985 from levels that would otherwise be reached if present trends are allowed to continue unabated would save annually an amount of crude greater than the current oil imports of Japan or half those of Western Europe. If the target is not met and the U.S. imports increase, the competitive pressures for oil, particularly Middle East oil, may reach dangerous proportions even without another politically motivated interruption in supplies. No other nation in the world (with the possible exception of Saudi Arabia, with its vast potential production capacity and its role of "swing" producer) can exert such an influence on the world's energy outlook through its domestic policies.

After this analysis, we can now enter the logic diagram on pages 56 and 57 at the point where the U.S. states its nuclear-policy position and then explore the international consequences.

All the foregoing trains of thought have been quite apparent to both developed and developing nations. Japan and most of the advanced industrial nations of Europe have meager coal or oil supplies themselves, relatively speaking; even North Sea oil harvested at the maximum planned capacity will supply only about 20 percent of Western Europe's needs. Thus all those countries, facing their own difficulties and the distinct possibility of a continuingly gluttonous U.S., see an increased incentive to push ahead with their own nuclear programs, including reprocessing and breeder reactors. Further encouragement to do so seems to be arising from an unintended source. Along with the U.S., Australia (currently not a uranium producer but potentially one of the world's two or three major exporters of uranium within the next few years) and Canada (the world's biggest uranium exporter) have also recently imposed stringent proliferation-related controls on their exports. In all three cases these controls include the requirement that consent must be obtained from the exporting nation before any of the fuel can be reprocessed. In at least one case deliveries of fuel have been delayed pending agreement to these and other conditions by importing nations. Although the controls have been implemented with the intention of creating a more rigorous nonproliferation regime, for the uranium-deficient, fuel-supply-sensitive Japanese and Europeans such manifestations of external political involvement in their domestic fuel-cycle operations, together with the unsettling prospect of further mercurial behavior by their suppliers in the future, may ultimately act to increase the vigor with which these nations set about reducing their dependence on fuel imports by the use of plutonium and breeder reactors. According to statements of intent from many of them, that is already happening.

What will leaders in developing countries see? They will see a rich U.S. liable to forgo nuclear power, in spite of what it might say about preserving light-water reactors. They will see increased short-term pressure on the world's supply of oil, inevitably resulting in shortages and still higher prices. They will see pessimistic projections for world petroleum and gas resources that will allow neither continued profligate use by the industrialized nations nor any chance for their own countries to follow a development path anything like that followed by their predecessors. They will see an offer of nuclear-fuel supplies that binds them to the goodwill of the U.S. and other developed countries for even limited nuclear assistance. (The recent suggestion of an international nuclear-fuel bank is a partial response to this point.) They will see the growing appearance of a world divided into an oligopoly of developed states that turns into an oligarchy as nuclear power becomes more important throughout the world, and a coterie of less developed nations that must fall farther and farther behind. (This latter impression becomes reinforced by U.S. actions that treat certain of its industrialized trading partners as "exceptions" in view of their continued insistence on the need for reprocessing and breeder reactors in their countries without delay.) And they will see little promise of help from the U.S. to become truly independent in terms of nuclear power.

The inevitable result will be increasing distrust of the U.S. and a growing sensation of unwanted dependence. Both developed and developing nations share these feelings of insecurity. Considering only uranium enrichment as an example, would a U.S. offer of more enrichment services suggest an extension of U.S. market control of enrichment supply, which in turn would suggest increased dependence on U.S. whims? Troubled in large part by the somewhat fickle nature of the decision-making process governing U.S. exports, the out-

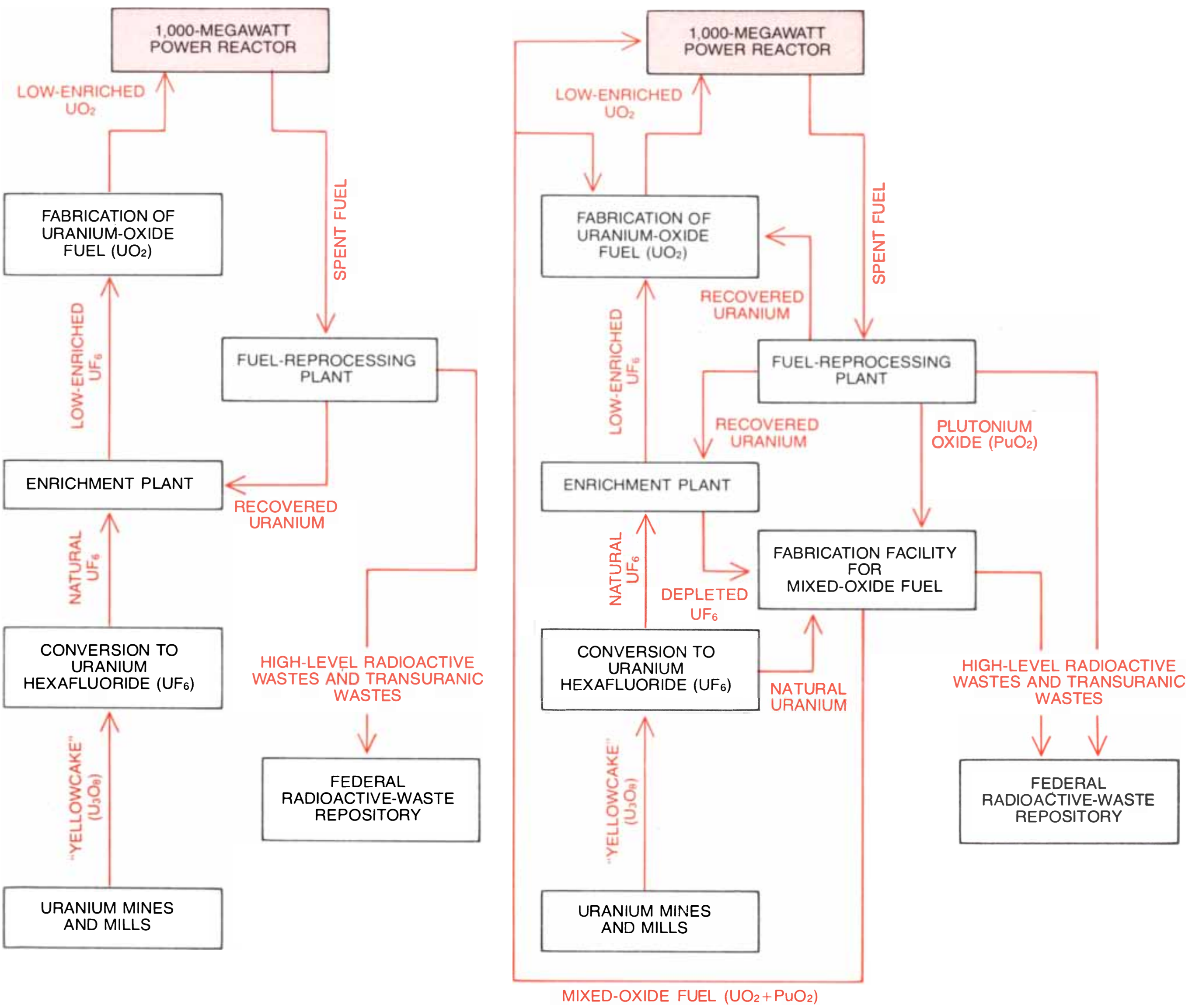


THREE NUCLEAR-FUEL CYCLES suitable with conventional light-water reactors are shown in these simplified diagrams, adapted from a recent report to the American Physical Society by its study group on nuclear-fuel cycles and waste management. In the prevailing "once through" approach (*left*) the spent-fuel rods, which still contain an appreciable amount of fissionable isotopes (principally "unburned" uranium 235 and plutonium 239 produced by the transmutation of the urani-

side world, including the less developed countries, sees more incentive to set up its own enrichment programs and to go nuclear with or without U.S. assistance.

Both developed and developing countries also share other reactions. Among both groups, for example, there are suspicions that the U.S. program is really designed to improve the sagging fortunes of U.S. nuclear exports: either to increase the attractiveness of U.S. light-water reactors or to curb the global movement toward plutonium and the breeder reactor until U.S. technology in these areas has caught up with capabilities in Western Europe. It has also been pointed out that the exemplary nature of the U.S. decision to defer indefinitely the reprocessing of its own commercial power-reactor fuel was compromised from the outset by the fact that fuel reprocessing in connection with the U.S. weapons program would continue as before. Related comments address the entire network of U.S. weapons-manufacturing activities and deployed weapons, suggesting that the scale and wide distribution of this system presents a more attractive target to would-be proliferators than a commercial plutonium fuel cycle would. Furthermore, some observers have speculated that the subsequent decision to use gas-centrifuge technology for the next increment of U.S. enrichment capacity rather than a more proliferation-resistant gaseous-diffusion plant compounds the already present inconsistencies.

To be sure, some of these reactions are mere rhetorical flourishes; nevertheless, they may still have wide-ranging international reverberations. Moreover, many of the reactions are contradictory. How can a rich U.S. liable to forgo nuclear power also be attempting to increase its share of nuclear exports? How can one explain the fact that some of the nations where complaints about the in-



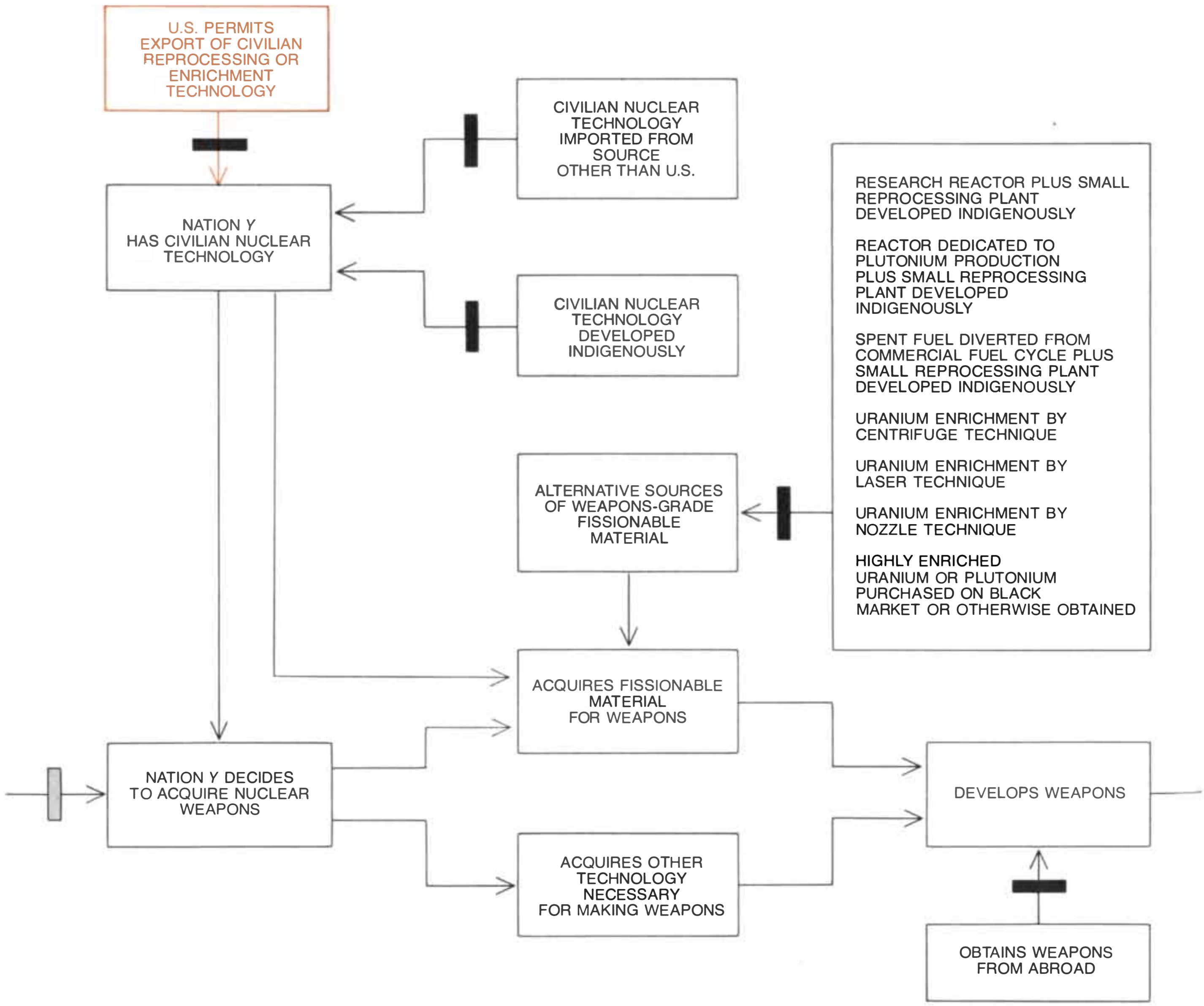
um-238 nuclei in the fuel are disposed of without reprocessing; disposition of the spent fuel can in principle be either temporary or permanent. In the uranium-recycle option (*center*) the spent fuel is reprocessed to recover only the residual uranium, which can then be enriched in the fissionable isotope U-235 or used as it is to replace some of the virgin natural uranium in the fabrication of new fuel assemblies. In the uranium-and-plutonium-recycle option (*right*) the spent fuel is reprocessed to separate both uranium and plutonium from the wastes. The recovered plutonium can then be combined with uranium having a very low concentration of U-235, in effect substituting the plutonium for some of the U-235 in the normally low-enriched fuel. Useful mixed-oxide fuels can be made by combining plutonium with uranium derived from a number of different sources, including the normal low-enriched uranium product from an isotope-separation plant, the uranium recovered from spent fuel or the depleted "tails" from a uranium-enrichment plant. It has been estimated that with both uranium and plutonium recycling the industrial operations required to supply enriched uranium could be reduced by about 20 percent in the year 2000 compared with what they would be for either the uranium-recycle or no-recycle options. This saving would of course require the introduction of the costly and complicated fuel-reprocessing and mixed-oxide fuel-fabrication operations.

consistency of U.S. policies toward domestic and military reprocessing have been heard are also those that rely most heavily on the presence of the U.S. nuclear deterrent for their defense? Such contradictions, however, do no more than mirror the ambiguities and contradictions we have recognized in the U.S. policy, as it attempts to strengthen the barriers between peaceful and violent uses of nuclear energy and simultaneously wrestles with an immense and growing demand for energy, both domestic and international.

The Nonproliferation Treaty, to which more than 100 nations are now parties, embodies an internationally negotiated agreement on the framework in which the energy v. proliferation enigma should be resolved. In it the non-nuclear-weapons states party to the treaty undertake not to develop or otherwise acquire any form of nuclear explosive and to accept international safeguards on all peaceful nuclear activities. In return for this commitment the right of all parties to the treaty to develop and use nuclear energy for peaceful purposes is affirmed, as is the right to participate in exchanges of equipment, materials and technology for the peaceful use of nuclear energy.

The restrictive export policies of the U.S. (and of other major nuclear suppliers) are viewed in many parts of the world as extending the inequalities that have always been inherent in the Nonproliferation Treaty between nuclear-weapons states and non-nuclear-weapons states. The new expression of these inequalities is the attempt to influence criteria for the international distribution of certain "sensitive" peaceful technologies, particularly reprocessing. Implied in this policy is a redrawing of the line separating peaceful uses of nuclear energy from violent ones, and therefore a redefinition of proliferation. Traditionally the latter had been defined as the acquisition of nuclear weapons. Now, however, the new U.S. position is being interpreted as an attempt to redefine proliferation as the capability of acquiring nuclear weapons. Had this always been the case, it is argued, negotiating the Nonproliferation Treaty would have been impossible in the first place.

We make no attempt to determine whether in fact the U.S. would be failing to comply with its international legal



NUCLEAR-PROLIFERATION SCENARIO currently perceived as being worrisome by the U.S. Government is illustrated in the form of this sequential logic diagram. The main horizontal decision path shows the series of steps that could be followed by a non-nuclear-weapons nation, designated *Y*, to acquire nuclear weapons and to use them. The vertical paths show several possible inputs to *Y*'s decision. For example, the U.S. sees the acquisition of a nuclear-fuel-reprocessing capability as a stimulant to weapons proliferation. Accordingly

obligations as a party to the Nonproliferation Treaty by implementing its proposed export criteria. We do observe, however, that the loss of confidence in the effectiveness of international safeguards that has taken place in the U.S. is reflected in many non-nuclear states by a corresponding loss of confidence in the ability of the Nonproliferation Treaty to provide an acceptable legal framework for the international distribution of peaceful and military applications of nuclear energy. In such circumstances the fabric of the global nonproliferation regime is inevitably weakened.

All the considerations we have discussed here show up as destabilizing routes in our second proliferation scenario. Not only does country *Y* find logical incentives to install domestic nuclear-fuel facilities, but also it perceives a world more fragmented and less secure. Feeling less secure itself, it naturally imagines others feeling the same way and hence it must increase its own security unilaterally. Escalation of uncertainty leads to escalation of international instability; a program originally intended by the U.S. to decrease the dangers of nuclear proliferation inadvertently has the opposite effect. Meanwhile the U.S. isolates itself from the mainstream of world nuclear policy, and its ability to favorably affect that policy diminishes.

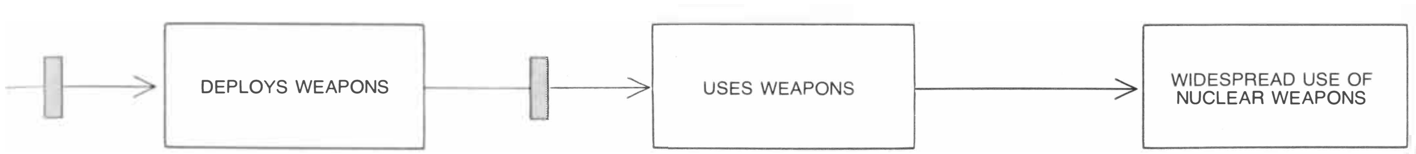
More caveats and auxiliary views remain to be displayed. None of this analysis is meant to overstate the role of nuclear power in solving the world's energy problems. That mistake has been made too many times before. Some of the problems currently facing the nuclear-power industry in non-Communist developed countries can probably best be understood in terms of a backlash against earlier technological overoptimism. For the majority of the less developed countries nuclear-generated electricity cannot play a significant part in meeting energy requirements for a long time. Costs have risen alarmingly, and besides, the type and scale of the energy supplied by currently available nuclear-power stations seem less compatible with the energy-demand structure in many of these countries than the output of other energy-supply systems.

Moreover, in some developing countries where nuclear power is intended to play a major role the overall development targets frequently appear overambitious and unlikely to be realized. Fears have been expressed that nuclear-power technology could critically exacerbate rivalries among the various political, industrial and technocratic elites and increase the gap between such elites and the remainder of the population. Until now suppliers have not discernibly modified any "hard sell" policies because they might ultimately contribute to domestic political instabilities with unpredictable international consequences, and it is unlikely that such self-restraint will be shown in the future. The point here is not, of course, that the industrialized supplier nations should decide what is good for the development of the poorer countries and impose export restrictions accordingly but that nuclear technology may be "sensitive" for many reasons other than the increased access to weapons-grade material it may provide.

We conclude, therefore, that the conventionally defended analyses have been inadequate. The various original motivations virtually disappeared from our discussion, and rightly so; they were more nearly goals than real policies. Both are necessary; to neglect the hard work of developing policy causes much trouble because then the original vision, however high-minded, is washed away by a sea of events, and only consequences remain.

What to do? In answering this question we have had to assume that many other issues related to nuclear power can be resolved. To us the three largest of those issues seem to be reactor safety, the management of nuclear wastes and the prevention of subnational nuclear felonies. Although we have not dealt with these matters, we realize their gravity. Our analysis and recommendations would be irrelevant, however, only if both the U.S. and almost all other countries opted out of nuclear power. Nuclear power might disappear in the U.S., but neither present reactors nor breeders will go away in many other places. If the activities of Western Europe and Japan are unconvincing, one need only consider the U.S.S.R. and its Eastern European allies; they also develop nuclear power, with the most sensitive activities being reserved to the U.S.S.R. The U.S. must stop acting from time to time as though nuclear power was about to go away, or as though its disappearance would have little consequence.

The Administration has begun to discuss some of the necessary changes: for example a gradual shift to a more flexible policy, more emphasis on providing an assured nuclear-fuel supply and a suggestion to set up an international storage facility for spent fuel. Owing to significant reductions in projected overall energy demand and particularly in forecasts of installed nuclear capacity, the next few years seem to us a period of grace, perhaps overperceived in the U.S. and underperceived abroad, during which fundamental repairs can be made to the fabric of nuclear goals and policies. The period may be a decade, but it can hardly be much longer, being limited by the consequences of the inexorable pressure on other energy resources. The time is technically long enough to develop variations in the current fuel cycle, perhaps long enough to devise a brand-new one and even reactors and other facilities to use the fuel, but such activity would require far firmer decisions and much prompter (and more expensive) implementation than we have seen. Both the preparation of fissionable fuel from new or reprocessed material (the "front end" of the fuel cycle) and the reprocessing of spent fuel must be carefully considered, not just the latter. As we have said, we consider the front end of the fuel cycle, the enrichment of natural uranium for example, to be a sensitive proliferation issue. The exis-



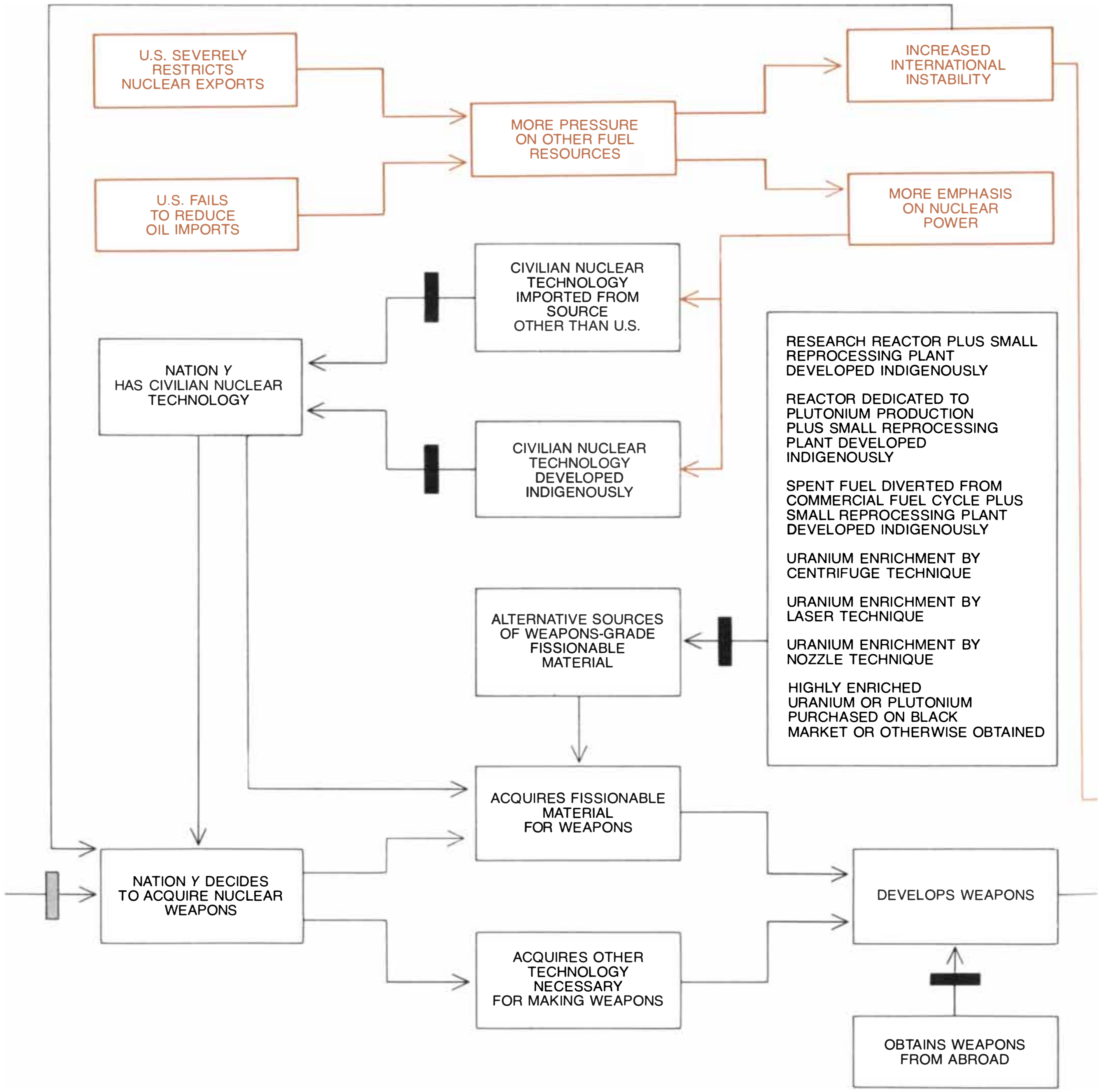
the Administration has sought to restrict the export of commercial reprocessing technology as a way of blocking this access route to the acquisition of nuclear weapons. Other routes exist, however, whereby *Y* could gain access to weapons-grade material; some of these possible routes are listed in the large box. In order to erect a really effective barrier to weapons proliferation, the U.S. would have to block all the lines marked with a black bar. Decision paths on which the U.S. can have only an indirect influence are marked with gray bars.

tence of about 200 power reactors around the world working on the present fuel cycle must also be considered, and the opportunities for technological innovation that might be applied to them are limited.

All these things make large demands on one decade or a little more. So do institutional accommodations among nations, not only with respect to the currently developed nuclear-fuel cycles but also to many other things. Time is short, whether for technological modification or for international institution-building. Whatever the outcome of the former, the latter is an indispensable part of efforts to deal with the problems of nuclear-weapons proliferation and energy scarcity. Although the Administration's proposals have created many problems, they have succeeded in injecting a new sense of urgency into the situation. It is essential that this asset not be allowed to evaporate.

Regarding present fuel cycles and other matters directly related to nuclear power in the next decade or two we have five main recommendations to offer:

1. Nuclear power should be kept alive in the U.S. at least as a long-term "insurance" option, and that means not only the continued development of light-water reactors but also progress toward de-



MORE COMPLEX SCENARIO is needed to portray more realistically the likely effects of an overly restrictive U.S. nuclear-export policy on the entire problem of international stability and the proliferation of nuclear weapons. The main decision path followed by nation *Y* toward the acquisition of a nuclear-weapons capability is the same as it is in the preceding diagram, as are a number of other elements in the diagram. In this case, however, the vertical inputs to *Y*'s decisions start with the new U.S. restrictions on nuclear exports,

veloping a viable breeder reactor. Central-station solar power and controlled fusion are only long-term possibilities, oil is only a short-term source of energy and we have little faith in coal for the long term.

2. To reduce uncertainty for the U.S. electric-utility industry and others the Federal Government should take several steps. First, it should reaffirm that the reprocessing of spent nuclear fuel is being delayed but not abandoned. Second, the Federal Government should assure the electric utilities of a review of national policy on reprocessing as the debate about it matures, and certainly within five years; that would include an assessment of the projected uranium supply, which would draw on the current national uranium-resource evaluation and other programs. Third, the complexity and prolixity of the licensing process for nuclear-power plants should be eased by making it more difficult for license applications to be recycled again and again to the Nuclear Regulatory Commission. The nuclear licensing bill currently under consideration appears to be facing formidable obstacles at several stages of the legislative process, and in any case it seems to address these issues only partially in its present form. Finally, the Federal Government should take on the entire burden of managing spent fuel, and guarantee to take responsibility for the fuel reasonably soon after its discharge from power reactors. That includes spent-fuel reprocessing, if and when a decision comes to do it. No other sector has an adequately long time perspective to plan and operate the appropriate facilities. In particular, the chemical industry, on which the task might otherwise be expected to fall, traditionally expects a payback on investment in a very few years, and therefore discounts far-future profits too much to match the long-term nature of the tasks, particularly waste handling and storage.

3. The U.S. should offer to explore with other nations the costs and benefits to the international community of completing the reprocessing plant now sitting idle at Barnwell, S.C., and operating it as an international facility. The principal objectives of such a project would be to gain experience with commercial reprocessing technology, to assess the effectiveness of international safeguards and to demonstrate the institutional viability of international cooperation in the provision of fuel-cycle services.

4. Efforts to increase the security of the international supply of uranium and enrichment services should be intensified. Domestically, differences among the various branches of the Government should not be allowed to interfere with the pivotal task of reestablishing the U.S. as a reliable supplier of enriched uranium fuel.

5. In all these activities we note the need for an international agency. We see none better prepared than the International Atomic Energy Agency (IAEA), and we believe it should be greatly strengthened so that it can continuously inspect sensitive facilities. The answer does not, however, lie in the mere strengthening or even proliferation of agencies. For example, restrictions on the export of appropriate nuclear systems may undermine the Nonproliferation Treaty, and the IAEA can do nothing about it.

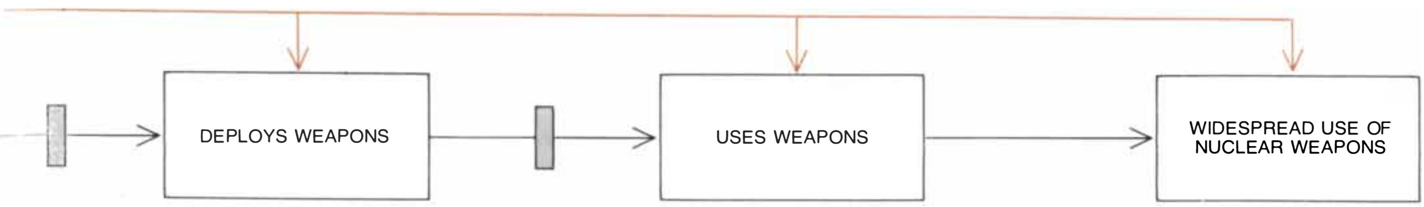
Beyond all these issues we see others seeming to stand out in the distance. First, it is noteworthy that the diagrams we have drawn differ from the conventionally discussed diagrams of causes and effects. Our discussion has been almost entirely international, as befits the problem. The U.S. approach has been too self-centered, insufficiently sensitive to the problems of other nations and lacking in awareness of its own potentially disruptive character.

Near the beginning of this discussion we mentioned that the Administration has attempted to breathe new life into the larger issues presented by nuclear weapons. If governments and people are so concerned with the risks of future proliferation, how much more should they worry about the huge numbers of nuclear weapons already deployed? One who lives on the edge of an abyss should not squander his effort avoiding small ditches. The real threat of nuclear weapons is seen once again, more clearly than before, in the illuminating perspective provided by the juxtaposition of thousands of existing megatons on the one hand and a few hypothetical kilotons on the other.

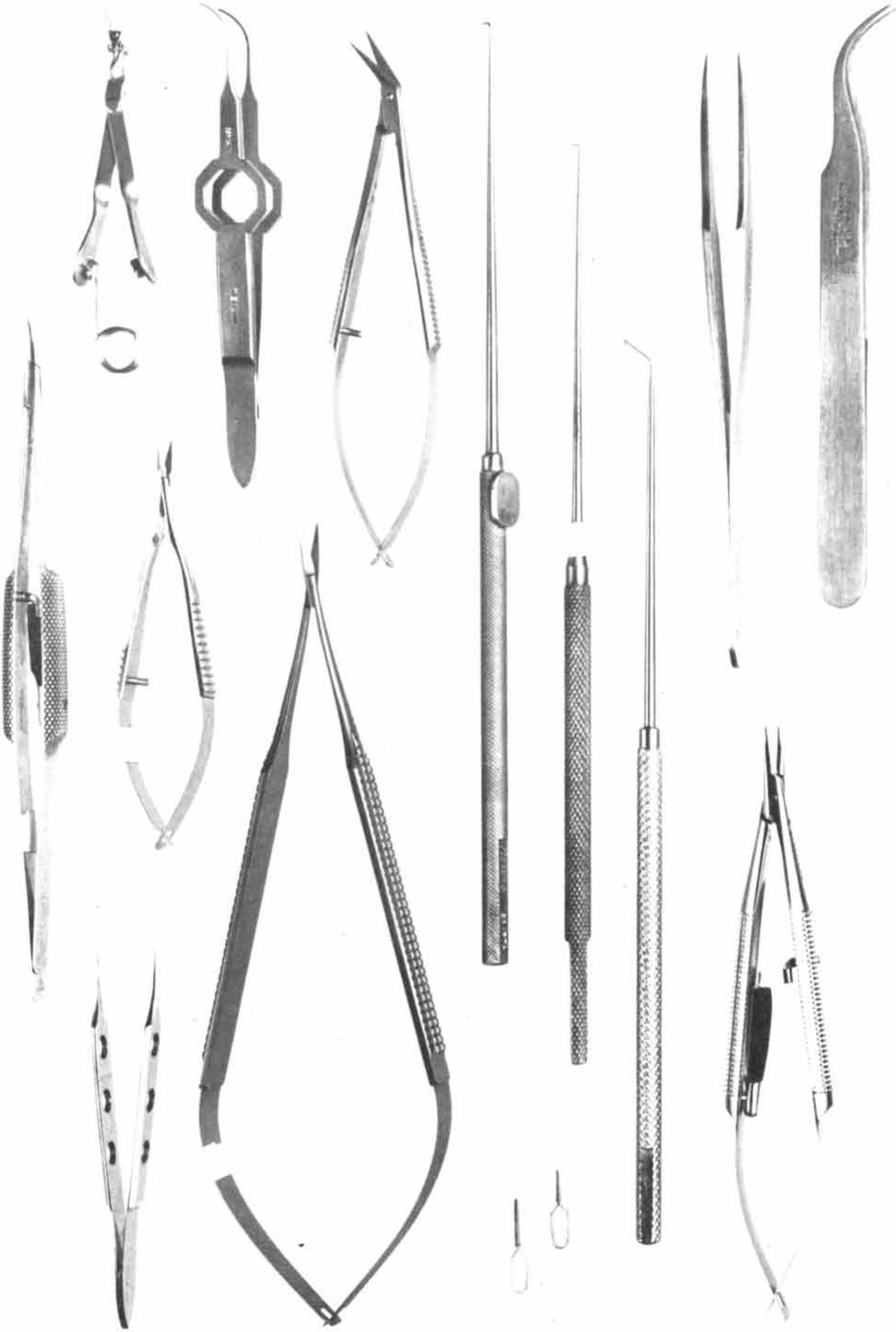
This brings us to the more general question of international peace and stability. In the worldwide search for routes to a juster and more sustainable society it has become clear to many observers that a peace in which the world is divided ever more rigorously into haves and have-nots is neither just nor likely to be very sustainable, whether the basis for division is social, economic or (as here) seemingly technological. Such a division not only defeats itself in the long run; even worse, it is wrong.

We propose that the real long-term solution both to the nuclear-power problem and to the larger problems of international instability lies not in fostering division but in its opposite: mutually cooperative international interdependence. Since nations must depend on one another, they lose more by going separately than by staying in partnership. Our analysis shows that this partnership must include the developing countries, since many of them, if they are excluded, are capable of upsetting the international order through the acquisition of nuclear weapons and other acts.

All of this will not be easy, but other approaches have yielded nothing but unstable arms escalation. The partnership should logically involve food, health care and many other sectors where the U.S. can make valuable contributions. Only in that way will we have a chance of answering constructively the question that can no longer be put aside: Why do people want to make nuclear weapons in the first place?



coupled with the failure of the U.S. to reduce significantly its imports of crude oil. Several additional logic paths, shown in color, represent the possible sequence of events that could result in added incentives for nation *Y* not only to push ahead with its own civilian nuclear-power program, including uranium-enrichment technology, spent-fuel reprocessing plants and breeder reactors, but also perhaps to respond to increased international instability arising out of growing competition for energy resources by joining the nuclear-weapons "club."



Microvascular Surgery for Stroke

An operation to increase cerebral blood flow by joining an artery in the scalp to one on the surface of the brain is made possible by new technology: an operating microscope and microinstruments

by Jack M. Fein

Diseases of the nervous system have become increasingly amenable to surgical intervention in recent decades, but the neurosurgeon's efforts are restricted in many instances by the limits of his ability to see clearly and to manipulate the very small structures with which he must deal. In the past five years new technology based on the operating microscope and microinstrumentation has made it possible to redirect blood flow from an artery in the scalp through the cranium into the arteries that supply the brain. This new bypass procedure can prevent a stroke in some patients who would otherwise be likely to have one; for many stroke victims it can prevent another stroke, and for some of them it may lead to at least a partial recovery of lost function.

A stroke is a sudden interruption of blood flow as the result of the obstruction or rupture of an artery supplying the brain. The brain's intricately organized electrophysiological activity calls for a high and continuous level of nutritional support. Any significant reduction in the supply of blood-borne oxygen and glucose impairs brain function; total interruption of blood flow for more than four minutes invariably gives rise to widespread and irreversible damage in the affected area. Stroke is by far the commonest serious disease of the nervous system, and it is a significant health problem in the U.S. Of some 500,000 new stroke victims each year, 30 percent die within 30 days; half of those who live require long-term institutional care.

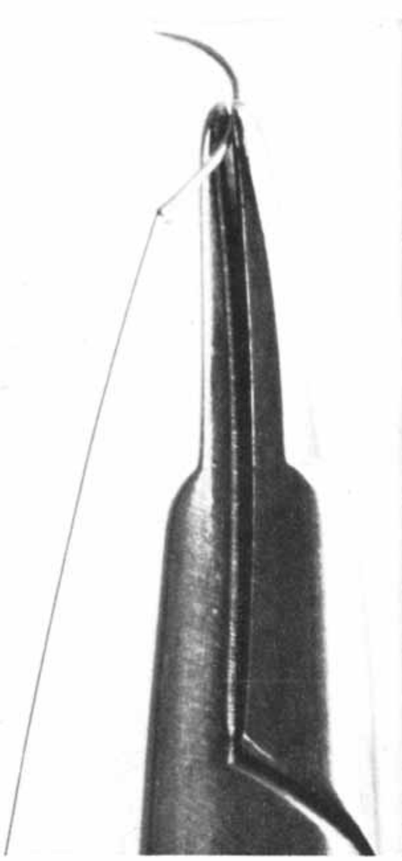
Most strokes (78 percent of them, according to the well-known Framingham study of cardiovascular disease) are the result of the obstruction of an artery by a blood clot: a thrombus or an embolus. It is these strokes that can sometimes be prevented or treated by microvascular surgery. (Hemorrhage due to rupture of an artery caused 17 percent of the Framingham study's strokes; various other events were responsible for the remaining 5 percent.) The clotting is usually precipitated by atherosclerosis, which narrows the lumen, or passageway, of the artery and otherwise predisposes the vessel to occlusion by thrombosis or embolism.

The innermost layer of a normal artery wall, the intima, is lined with endothelial cells attached to collagen and elastic connective tissue. The intima forms a smooth and continuous passageway for the blood. The next layer is the media, largely a wrapping (as in a radial tire) of smooth-muscle cells that constrict and relax to maintain a nearly constant blood flow as the blood pressure varies. The outer layer, the adventitia, contains supporting connective tissue and the nerves and small blood vessels supplying the artery wall itself. In atherosclerosis there is a proliferation of smooth-muscle cells into the intima and an accumulation of fats, fibrous material and cellular debris to form lumpy plaquelike lesions (atheroma) that disrupt the smooth endothelial surface and partially obstruct the lumen.

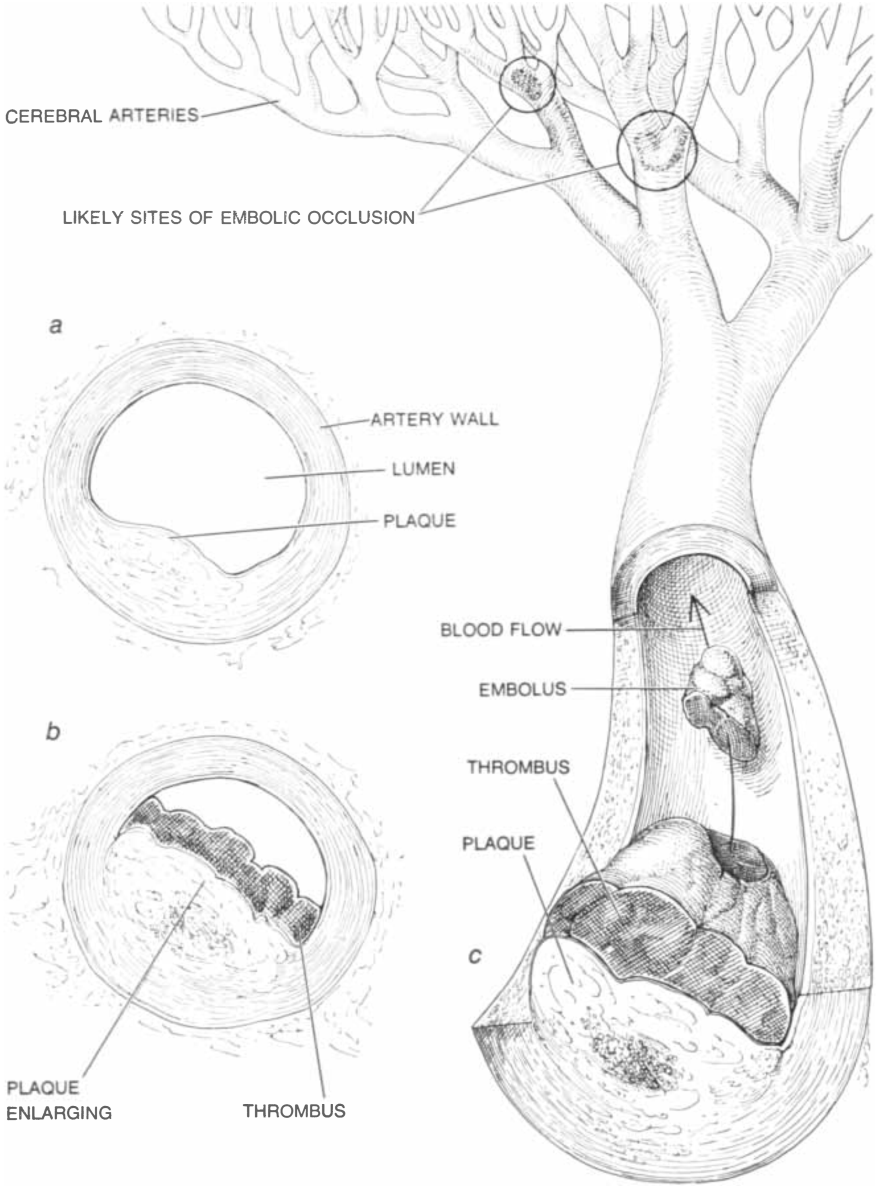
The exposure of platelets in the circulating blood to collagen fibers adjacent to the endothelial cells induces changes in the outer membrane of the platelet; the platelets become sticky and aggregate into a mass that initiates the formation of fibrin, a protein that participates in the clotting process, from the protein fibrinogen. The fibrin forms a weblike network that traps more platelets and red blood cells, building up a thrombus. The combination of thrombus and atheroma may narrow the lumen enough to reduce blood flow to the brain; as the disease progresses the artery may be completely occluded. Alternatively fragments of the thrombus may break off as mobile emboli, which are carried by the bloodstream until they become lodged in a smaller brain artery, usually at a branch point, and plug it.

Other organs (such as the heart and the kidneys) and the lower extremities are also prone to develop arterial insufficiency as a result of atherosclerosis. The effect on the brain is particularly ominous, however, because the organ is so susceptible to circulatory insult and because the results of a stroke can be so devastating. The brain is particularly sensitive to the interruption of blood flow because its energy reserves are meager. Even in the heart and in voluntary muscle some glucose can be stored as glycogen; not so in the brain. Nutri-

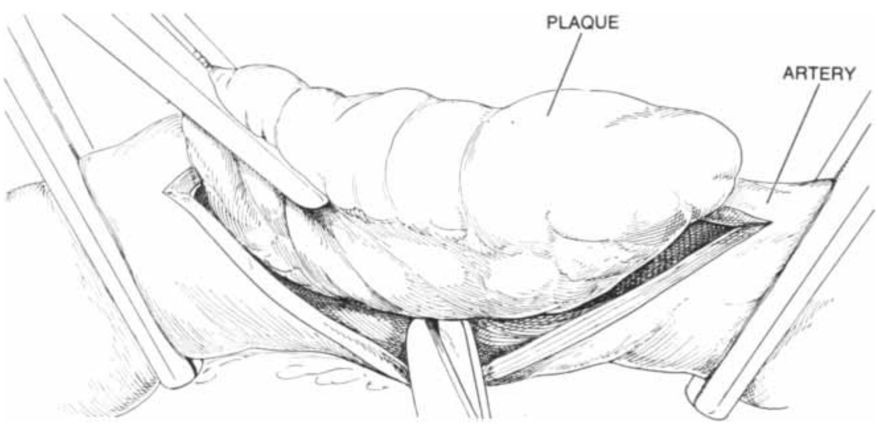
MICROINSTRUMENTS such as the ones shown at about actual size in the photograph on the opposite page are required for manipulating, cutting and suturing the small arteries that are joined in the microvascular bypass operation. Note microneedle holder (*far left*).



TIP OF NEEDLE HOLDER seen in the photograph on the opposite page is here enlarged 10.5 diameters. It holds a curved microneedle about two millimeters long carrying suture material that is 20 micrometers in diameter.



ATHEROSCLEROSIS can lead to a stroke. The atherosclerotic plaque, a thickening of the artery wall, narrows the lumen, or passageway (*a*). It may precipitate the formation of a thrombus, or blood clot (*b*), which can grow and occlude the artery. A portion of the clot may break off and move downstream, forming an embolus that occludes a smaller branch of the artery (*c*).



ENDARTERECTOMY is performed to remove an obstructing lesion from a relatively large artery, such as the common carotid. An incision is made in the artery wall, the thrombus is removed and the atherosclerotic plaque material is cut away with some of the adjacent intima.

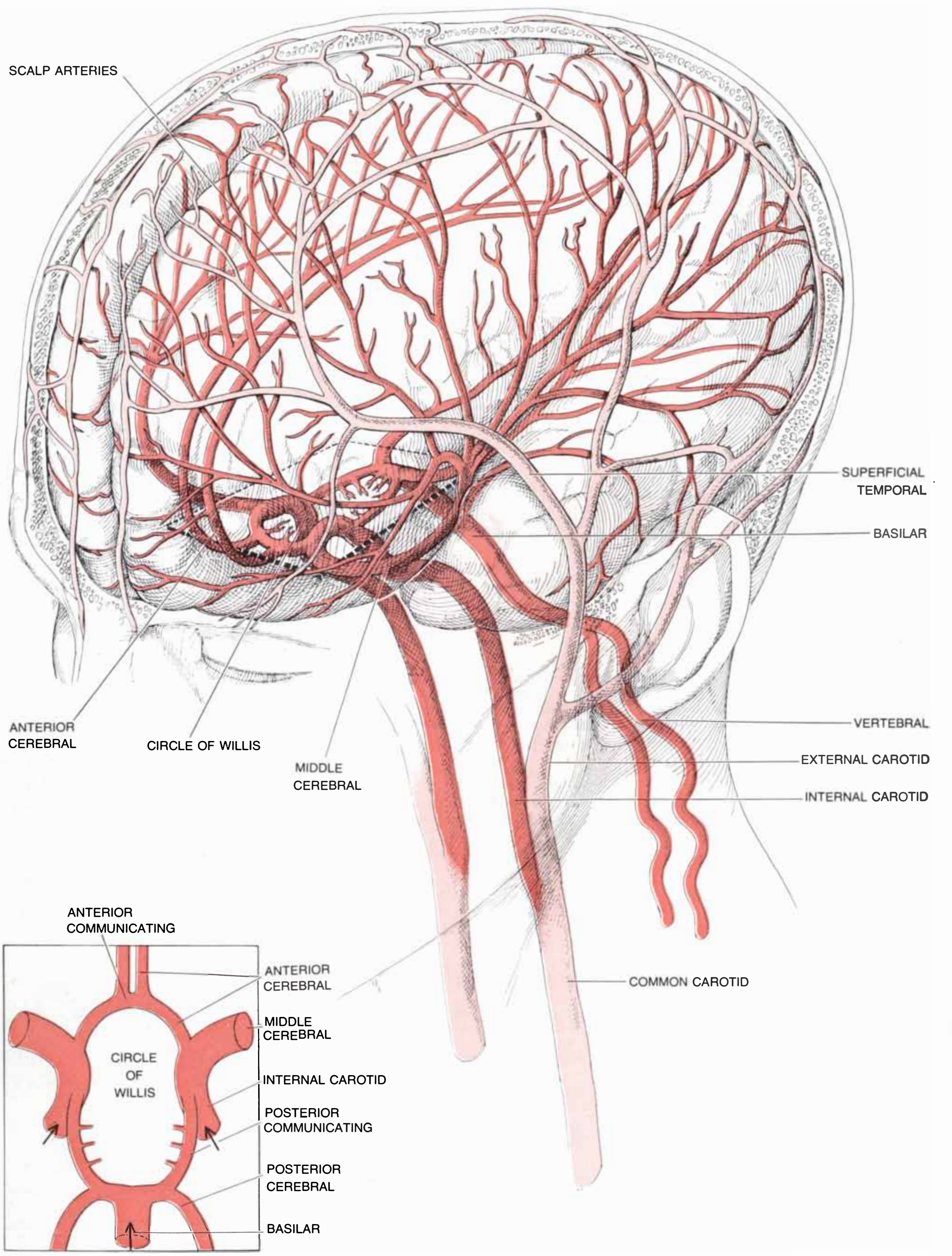
ents in the brain are rapidly metabolized to produce the energy-rich compound adenosine triphosphate (ATP) that powers the specialized functions of nerve cells and maintains their structure.

It is for this reason that the brain tolerates a major arterial occlusion very poorly. The brain can tolerate partial obstruction of flow through an artery supplying a localized region, depending on the adequacy of the "collateral circulation" through smaller interarterial channels, which may dilate rapidly as the blood pressure drops in response to the obstruction of a larger artery. It is the capacity of this collateral circulation to maintain blood flow around a localized obstruction, rather than the size or site of the obstruction itself, that most often determines how much brain tissue will survive a stroke.

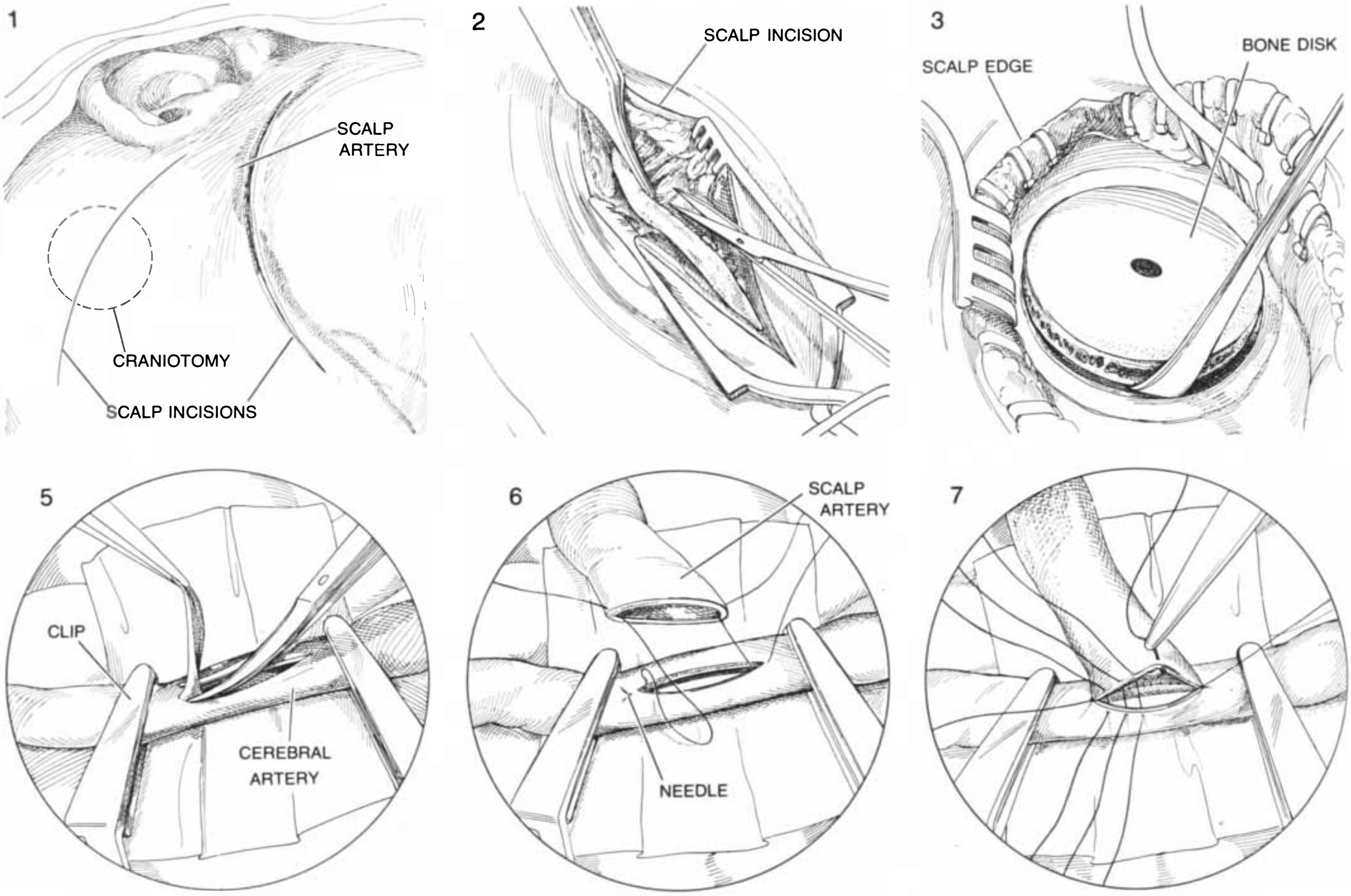
Blood is supplied to the head and brain by the right and left carotid arteries and the right and left vertebral arteries. The right common carotid and vertebral arteries arise from the innominate branch of the aorta, the left vertebral artery arises from the left subclavian artery and the left common carotid arises directly from the aorta. The common carotid artery on each side divides at the level of the larynx into the external and internal carotid arteries. The external carotid divides in turn into a dozen branches. One of them, the superficial temporal artery, takes (with its branches) a sinuous course through the scalp, providing nutrition for the hair, the skin and the subcutaneous tissues; its pulse can be felt at the temple, just in front of the ear, and it is often prominent enough to be visible through the skin.

The internal carotid arteries penetrate the base of the skull. Within the cranium their major branches (the anterior cerebral, middle cerebral and posterior communicating arteries) supply the two cerebral hemispheres; smaller branches supply, among other regions, the retina of the eye and the pituitary gland. The anterior cerebral arteries supply the portions of the brain near the midline that influence eye movement, memory, consciousness, genitourinary function and leg movements. Each middle cerebral artery supplies the central 60 percent of its cerebral hemisphere. This is the region that regulates movement and sensation (on the opposite side of the body), and on the dominant side of the brain it has an important role in the reception, interpretation and elaboration of language. (The left hemisphere is dominant in most right-handed people and in half of those who are left-handed.) The posterior communicating artery and the posterior cerebral artery, which it subserves, contribute to the supply of portions of the brain that are involved in the reception, interpretation and recall of visual information.

The brain stem and the cerebellum



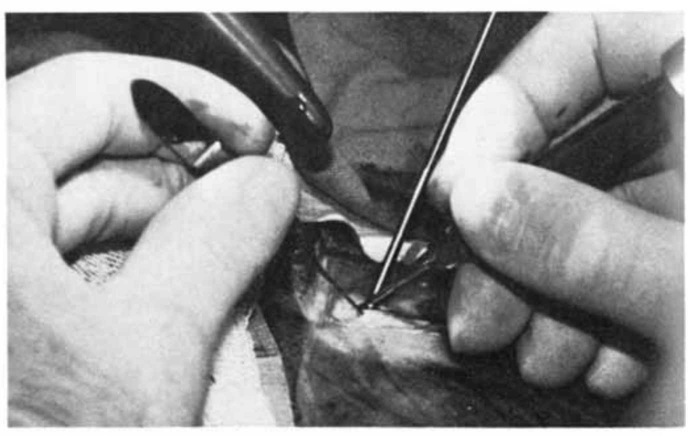
MAJOR ARTERIES of both sides of the brain and scalp are shown in this oblique view. The external carotid artery branches to feed the arteries of the scalp (*light color*); the internal carotid and the vertebral supply the vessels (*dark color*) that nourish the brain itself. The anterior and posterior communicating arteries, anterior and posterior cerebral arteries and a segment of the internal carotids form the circle of Willis (*broken black outline and inset drawing*), through which blood can flow in either direction to compensate for effect of an occlusion.



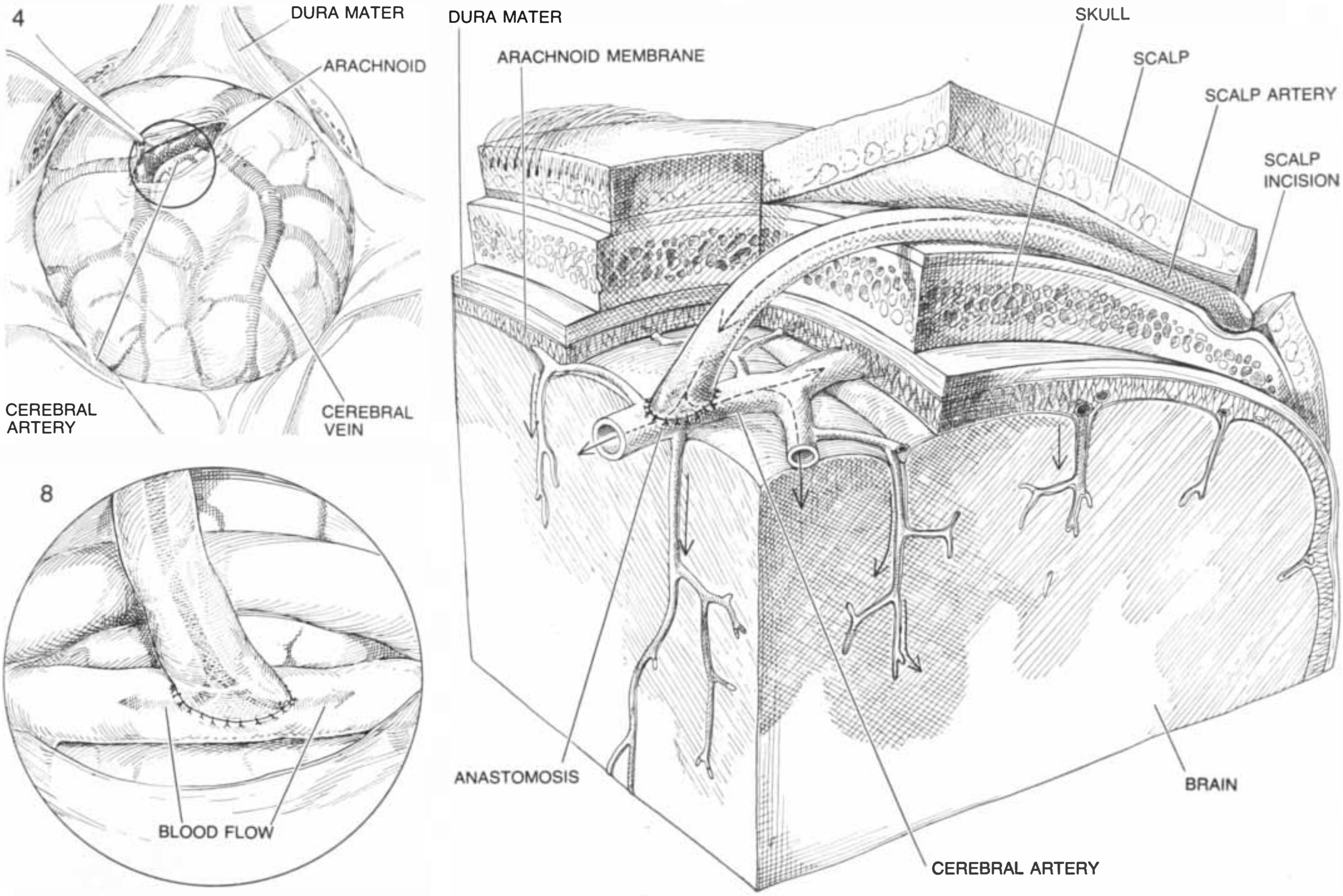
MICROVASCULAR BYPASS OPERATION connects an artery in the scalp to an occluded artery supplying the brain, at a point beyond the occlusion. The operation begins (*1*) with incisions in the scalp: one alongside of the donor scalp artery and one over the site of the craniotomy, where a disk of the skull will be removed. The scalp artery is dissected from underlying fat and muscle (*2*). The scalp is retracted and a disk about an inch and a half in diameter is removed from the cranium (*3*). The dura mater and the thinner

are supplied by branches of the two vertebral arteries and of the single basilar artery they join to form. The basilar artery is connected to the two internal carotid arteries by a segment of the posterior cerebral arteries and by the short posterior communicating arteries. The two anterior cerebral arteries are linked by the anterior communicating artery. The communicating arteries thus establish the circle of Willis: a continuous circular channel into which blood flows from both the carotid and the basilar arteries, and through which blood is redirected in response to a reduction in blood flow caused by a major occlusion between the heart and the circle.

An occlusion of any one of the major intracranial arteries usually produces symptoms related to the part of the brain that is affected, thus suggesting which artery is involved. The symptoms cannot, however, pinpoint the site of the occlusion. Moreover, lesions in a carotid or vertebral artery outside the cranium can give rise to symptoms similar to those caused by occlusion of a cere-



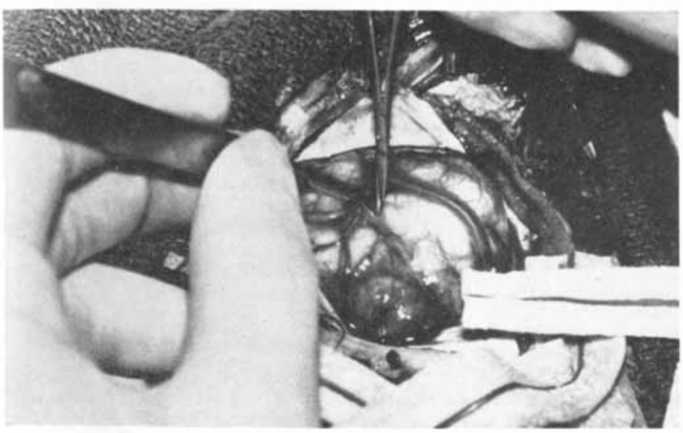
FOUR STEPS IN OPERATION are seen in photographs made at the Albert Einstein College Hospital in New York. The surgeon (*seated*) and an assistant are at the operating table (*left*). The operating microscope is draped in plastic for asepsis. The scalp artery is dis-



arachnoid membrane are peeled back, revealing the blood vessels at the top of the cortex (*4*). Working under the operating microscope, the surgeon cuts an opening in the recipient cerebral artery (*5*). The scalp artery, having been divided and led under the scalp to the cerebral artery, is brought into apposition with it by two anchoring sutures (*6*). More sutures are placed at equal intervals (*7*) and then the sutures are tied (*8*). The end result of the surgery is shown in the large drawing (*right*), with arrows indicating the reestablished blood supply.

bral artery. Before any surgical treatment can be considered it is therefore necessary to localize the site of occlusion. This can be done by means of cerebral arteriography, which was originally developed in 1927 by the Portuguese neurologist Antônio de Egas Moniz and has since been modified and improved. A long, flexible cannula is introduced into the femoral artery in the leg and passed up through the arch of the aorta. By means of the cannula an iodinated contrast agent that appears opaque on an X-ray film is then injected into both common carotid and both vertebral arteries in turn. In each case the passage of the contrast material through the brain is followed with serial X-ray exposures. This technology has been so refined that now a significant abnormality can be visualized in arteries as small as 200 micrometers (two tenths of a millimeter) in diameter.

The surgical procedure that first proved to be effective in stroke was endarterectomy, in which a longitudinal



sected from its underlying tissue (*second from left*). With the microscissors the arachnoid is opened to bare the recipient cerebral artery (*third from left*). The anastomosis is performed as the two blood vessels are joined by sutures placed with the microneedle holder (*right*).

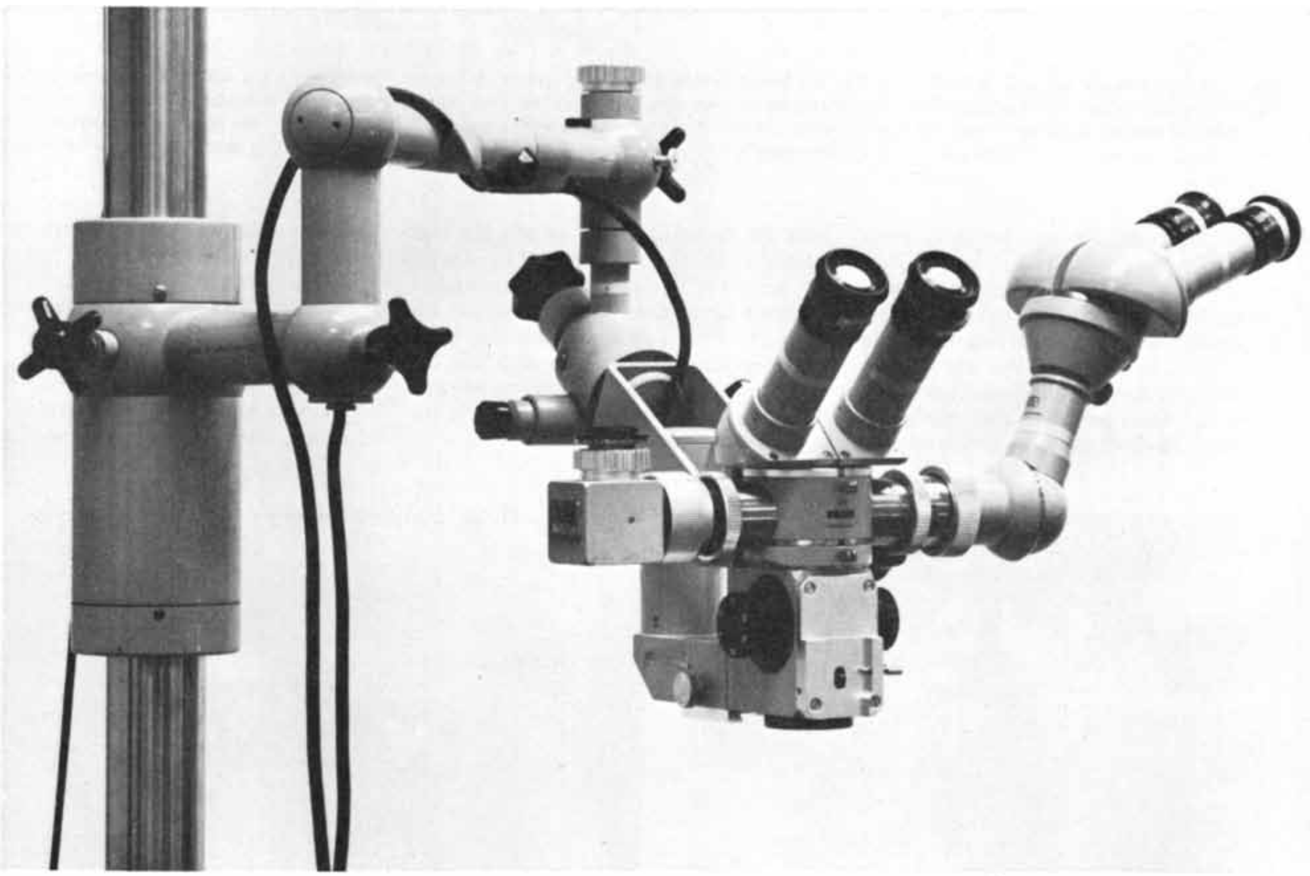
incision is made in an artery at the site of an occlusion and the obstructing clot is removed. The first endarterectomy of the common carotid artery was performed at the Montefiore Hospital and Medical Center in New York on January 28, 1953. The clot extended too far up the artery to be removed completely, and so only a small increase in blood flow was achieved, but the operation did establish the basic principles for surgical treatment of lesions obstructing the carotid artery. Endarterectomy is now a standard method of reestablishing blood flow through the extracranial arteries of the neck. Occlusions of the carotid and vertebral arteries can also be treated by enlarging the lumen of the blocked segment with a patch graft or bypassing the blocked region with a prosthetic tube or a vein graft.

As neurosurgeons accumulated experience with these extracranial procedures it seemed clear that such treatment was of most benefit to patients who had suffered a relatively minor stroke and were at risk of an impending major stroke. A collaborative study at medical centers throughout the U.S. indicated that patients with a history of minor stroke and with a clearly defined lesion of the carotid artery were indeed more likely than other people to have a major stroke and were most likely to be helped by surgery. The same study showed, however, that although the bifurcation of the common carotid artery in the neck was a frequent site of atherosclerotic lesions in stroke patients, it nonetheless accounted for less than half of the lesions. Most patients with occlusive disease of the cerebral vessels have obstructive lesions at more than one site, often in the smaller arteries of the brain as well as (or rather than) in the large arteries of the neck, and conventional procedures such as endarterectomy and grafting cannot be carried out on these smaller vessels.

A way to reestablish blood supply in patients with intracranial occlusions had been sought for some time. In 1944 a German surgeon, G. Henschen, tried attaching a flap of muscle (the temporalis) from the side of the head to the surface of the brain, hoping that the muscle's blood vessels would join the cerebral arteries and supply them. He reported that the patient's condition improved, but no arteriogram was made to document the result. In the 1950's C. Miller Fisher of the Harvard Medical School proposed that anastomosis, or joining, of cerebral arteries beyond the point of occlusion might be appropriate in some stroke cases, but such manipulations were then still beyond the state of the art.

At about the time the first carotid endarterectomy was performed a few neurosurgeons did attempt to remove obstructive atherosclerotic lesions from intracranial arteries. In 1955 W. Keasley Welch reported removing two occlusive lesions of the middle cerebral artery by means of conventional techniques. In 1960 two surgeons at the University of Bologna reported a procedure in which they milked a small "gunshot pellet" embolus from the middle cerebral artery back into the internal carotid.

Endarterectomy of intracranial vessels proved to be an unsatisfactory procedure, however. Surgeons described the difficulty of operating at the base of the brain, where the lesions were situated, and reported on complications associated with the procedure. The small size of the arteries presented major problems. Techniques for manipulating and repairing vessels one or two millimeters in diameter had not been worked out. Instruments and suture material adapted to large vessels caused damage in small intracranial arteries, which led to thrombosis. Moreover, the surgeon was inevitably hampered by the limits of his visual acuity and by inadequate

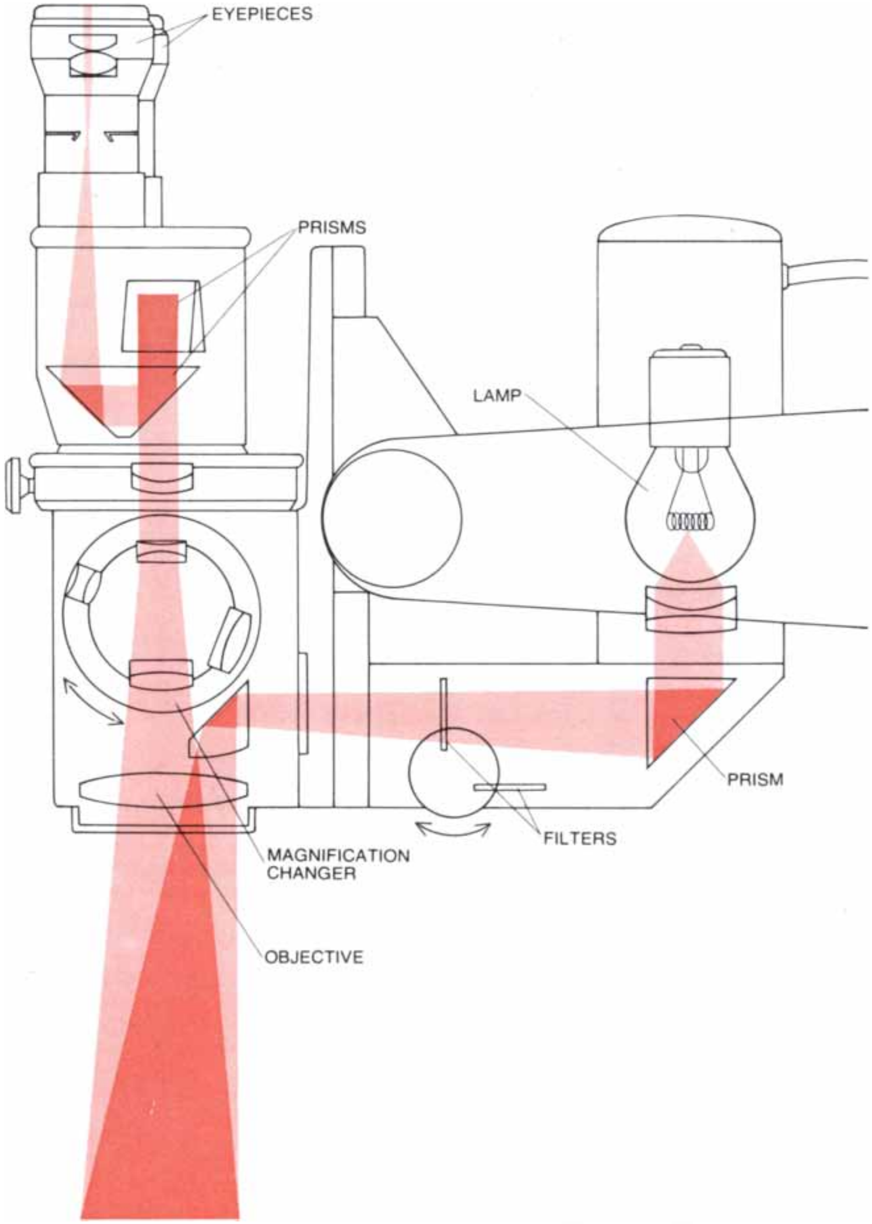


BINOCULAR OPERATING MICROSCOPE made by Carl Zeiss, Inc., gives the neurosurgeon at the Albert Einstein College Hospital a three-dimensional view of the field of operation at magnifications ranging from about six to about 40 diameters. This version, in which the eyepieces are inclined at an angle of 45 degrees from the axis of the objective, has secondary viewing tubes that can be fitted with eyepieces for an assistant, as here, and a camera attachment. The microscope is suspended from a crossbar that rides on a floor stand.

illumination deep in the brain. Intracranial lesions came to be considered off limits for surgical treatment. The development of a practical operation awaited not only a new concept but also improved technology.

A milestone was reached with the development of the operating microscope and its modification for neurosurgery, which required in particular that direct illumination be provided through the same optical system that magnifies the operating field. It was the ear, nose and throat surgeons who pioneered in applying the microscope to surgery. In the early 1920's the Swedish surgeon G. Holmgren used a binocular microscope, which affords three-dimensional vision, in treating otosclerosis of the middle ear. Ophthalmologists were the next to adapt the microscope for surgery. Eventually an operating microscope was designed that met the requirements for neurosurgery. It is constructed so that the field of operation is illuminated through the objective lens by a light source incorporated within the microscope, eliminating the shadows that would otherwise be cast by structures above the field. The objective lens and the eyepiece of the microscope provide the basic magnification (16 diameters in our instrument at the Albert Einstein College Hospital). A rotating magnification changer carries two pairs of refractors that can be interposed above the objective; one pair enlarges the image and the other reduces it, so that the surgeon can select from among four more magnifications (about six, 10, 25 and 40 diameters in our case) by adjusting a control knob. A beam splitter above the objective directs the image to two additional portals as well as the main binocular tube. While the surgeon looks through his eyepieces one supplementary tube can accommodate an assistant and the other can be fitted for motion-picture or still photography or for television.

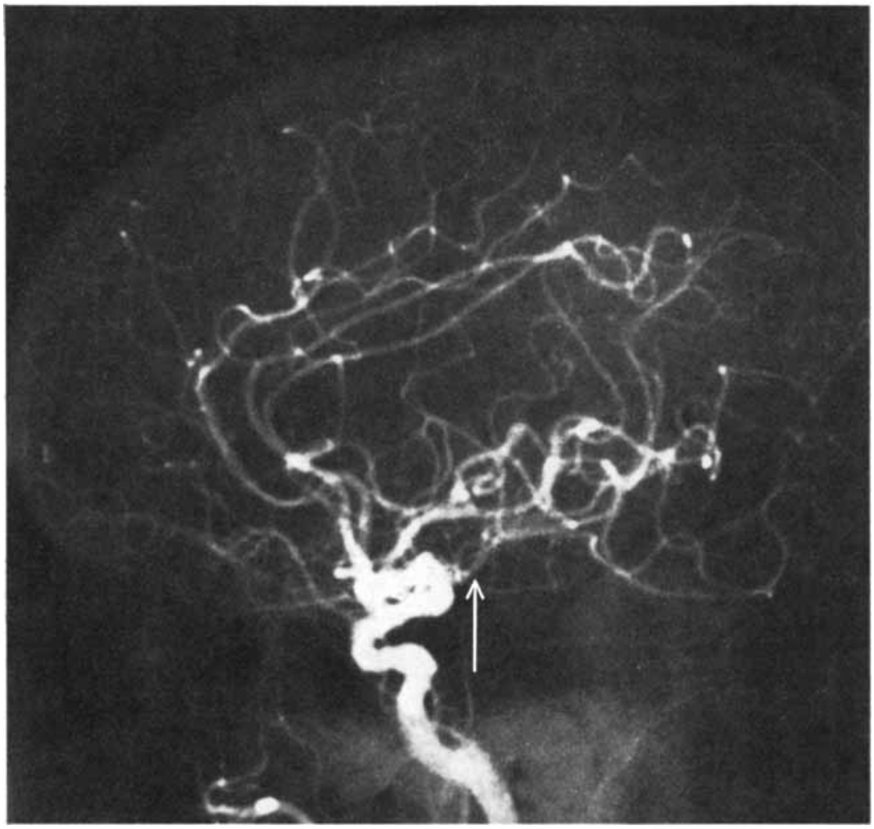
The development of microinstrumentation in the course of the past 10 years has greatly extended the range of the neurosurgeon's manipulations and his ability to deal with very small blood vessels. Microscissors and corresponding versions of probes, hooks, blood-vessel clips and suture-needle holders have been designed so that they can be activated by small pinching movements of the fingers rather than by hand and wrist action. Nylon microsuture material 20 micrometers in diameter, a quarter or a fifth as thick as a human hair, is now available, with which the surgeon can join arteries and veins a millimeter in diameter.



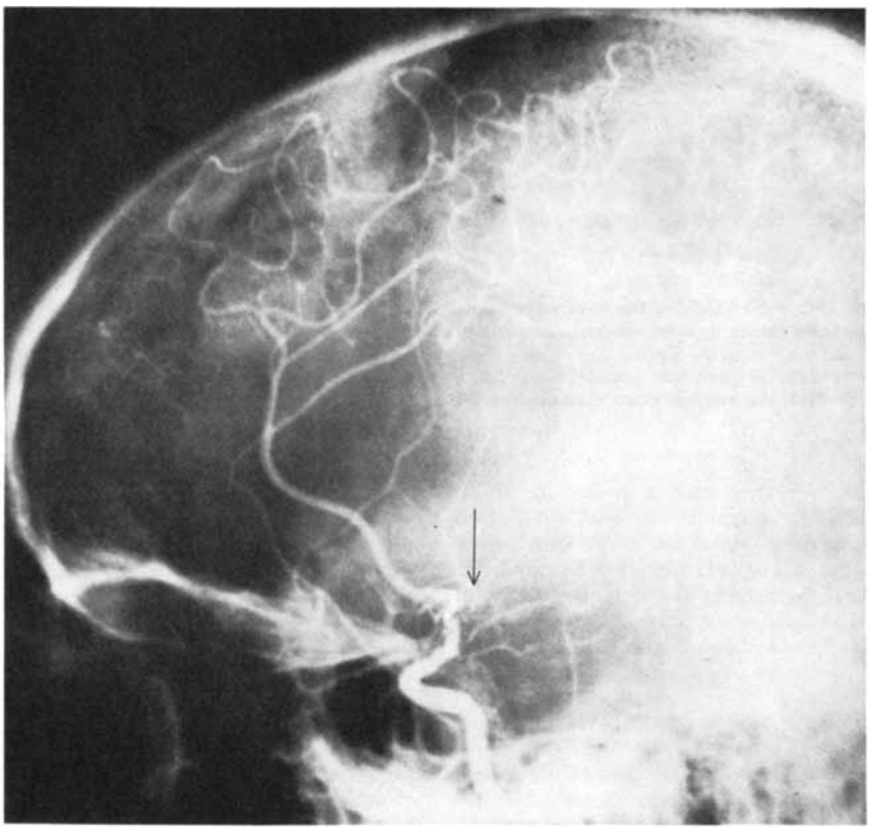
OPTICAL SYSTEM of the basic operating microscope made by Zeiss, with a straight binocular body rather than an inclined one and with no secondary tube, is shown. The illuminating beam is directed (by a prism just outside the line of sight) down into the field of operation along a path that is essentially parallel to the line of sight. The reflected light passes up through the objective, the magnification changer and the prisms of the binocular body to the eyepieces.

In the late 1960's M. Gazi Yazargil of the Cantonal Hospital in Zurich was the first surgeon to link a scalp artery with a branch of the middle cerebral artery on the surface of the cerebral cortex. Whether such a procedure could actually augment cerebral circulation enough to make the effort worthwhile remained to be established. Beginning in 1972 our group at the National Naval Medical Center was able to simulate cerebrovascular occlusive disease in experimental dogs and then to treat the condition by performing an anastomosis of the superficial temporal artery to a cortical branch of the middle cerebral artery. The increased flow of blood and the brain's improved function in response to the bypass suggested that the procedure could supply enough blood to the brain to prevent the development of a stroke in a susceptible human patient. By 1975 our group, which was now at the Albert Einstein College of Medicine, and surgeons at other medical centers were performing the bypass surgery on selected patients.

The major goal of stroke surgery is to prevent a debilitating stroke. The best candidates for the microvascular bypass procedures are therefore patients who are likely to develop a stroke because of atherosclerotic lesions; the operation can bypass obstructive lesions of the internal carotid, middle cerebral or vertebral arteries. In most patients there is a stepwise progression of the disease before a major, crippling stroke. Careful questioning of a stroke victim or his rel-



ANGIOGRAM of the cerebral arteries is made by injecting a radiopaque contrast medium and following its course with serial X-ray exposures. This angiogram shows the branching of a normal internal carotid artery. The middle cerebral artery and its branches (*arrow*) are visible.



OCCLUSION of the middle cerebral artery is evident in this angiogram. The blood flow can be seen to stop abruptly at the site of the occlusion (*arrow*); some collateral circulation has been established, as is indicated by the fine light traces in the region beyond the occlusion.

atives reveals that in more than 75 percent of the cases the major stroke was preceded by transient episodes of decreased blood flow to the brain. These transient ischemic attacks, or warning strokes, last anywhere from several minutes to 24 hours, after which there is a gradual return of normal blood flow and a resolution of the neurologic deficit.

Some patients have only a few such attacks prior to a major stroke; others may have more than a dozen a day. The symptoms depend on the part of the brain that is affected. There may be weakness, numbness or heaviness on one side of the body, difficulty in speaking or writing, weakness in both legs, temporary blindness or dizziness. It is still not clear whether the decrease in blood flow is caused directly by an obstructing thrombus in the carotid artery or by a mass of platelets from the thrombus that forms an embolus and temporarily occludes a smaller vessel. In the case of the temporary blindness known as amaurosis fugax there is evidence that platelet emboli originating at plaques in the carotid artery travel through the ophthalmic artery to the arteries of the retina. Such emboli have actually been observed through an ophthalmoscope within the retinal arteries (the only arteries subject to such direct observation) during an attack.

Both prospective studies of patients who have transient ischemic attacks and retrospective studies of stroke patients indicate that people who have had even one transient attack are 10 times more likely than people in the normal population to have a stroke. Clearly they are potential candidates for microvascular bypass surgery. Three other categories of potential candidates should be mentioned. There are patients who experience a more severe version of the transient ischemic attack: a reversible ischemic neurologic deficit, in which the neurologic symptoms persist for more than 24 hours but there is still little permanent brain damage. Then there is a group of patients who suffer from a stepwise dementia (commonly called senile dementia but often seen in middle age) caused by an insufficient blood supply to the brain. Finally there are patients who have had a stroke that leaves them only partially disabled.

Bypass surgery may be appropriate for a patient who is established, by clinical history and physical examination, to be in one of the four categories described, provided that his condition is such that he still has a lot to lose in the event of a stroke and provided that the lesions are at a favorable site. The computerized tomographic (CT) scan, a soft-tissue X-ray technique, helps determine the extent of permanent brain damage; extensive damage is a relative contraindication for the operation. The electroencephalogram, the isotope brain scan and examination of the spinal fluid help the surgeon rule out causes other than vascular insufficiency (a brain tumor, for example) for the patient's neurologic symptoms. Finally it is necessary to confirm directly by arteriography that the clinical problem is caused by an obstructive lesion and to determine the precise site of the lesion.

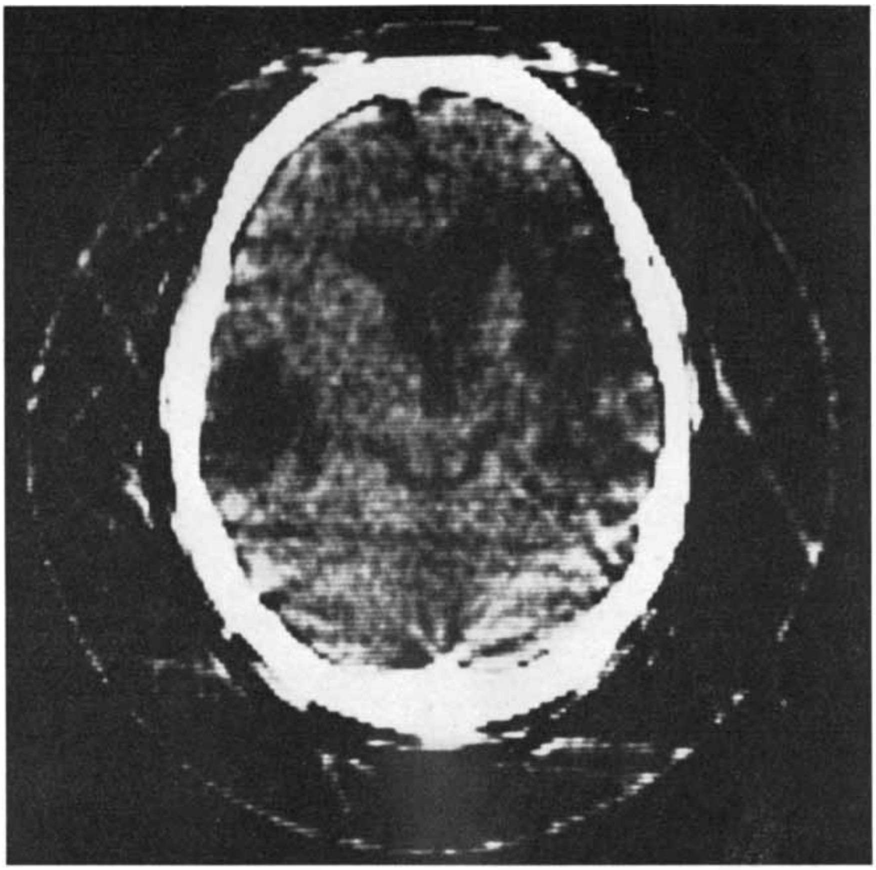
The patient selected for operation is placed under general anesthesia on an operating table fitted with a device that keeps the head fixed rigidly in place. A scalp artery of the proper size and in the right place to serve as the donor vessel is selected by palpation through the shaved skin and an incision is made alongside the artery. (Since the scalp vessel provides the blood supply for a region of the scalp, diverting its flow into the brain may impair the skin's blood supply, but there is enough collateral circulation for this not to have been a permanent or serious problem.) The scalp artery is freed from the skin and the underlying muscle is opened, exposing the bone of the skull. A disk four centimeters (about an inch and a half) in diameter is removed from the skull with a drill-like cylindrical saw called a trephine. Then the dura mater, the protective membrane that covers the brain, is opened to expose the cerebral cortex and the arteries and veins on its surface.

The surface of the cortex is examined through the operating microscope and an appropriate vessel, typically the angular branch of the middle cerebral artery, is chosen as the recipient vessel for the graft. The donor artery from the scalp is dissected free of subcutaneous tissue, its branches are ligated and its blood flow is interrupted with a small noncrushing clip; the artery is divided and a length of it is mobilized for the bypass. Working under the operating microscope at a magnification of 16 or 25 diameters, the surgeon then prepares the end of the donor artery, cutting it diagonally to make an oval opening that maximizes the surface area of the anastomosis. Noncrushing microclips are placed on the cortical artery to interrupt its blood flow temporarily, an oval opening is made in its side and the segment of artery between the clips is emptied of blood; a soft Teflon tube called a stent is placed in the vessel to keep the lumen open and facilitate the suturing. The microscope is turned up to high magnification (25 or 40 diameters), the end of the scalp artery is brought into apposition with the opening in the side of the cortical artery and the two vessels are joined by a series of from 18 to 24 carefully placed microsutures.

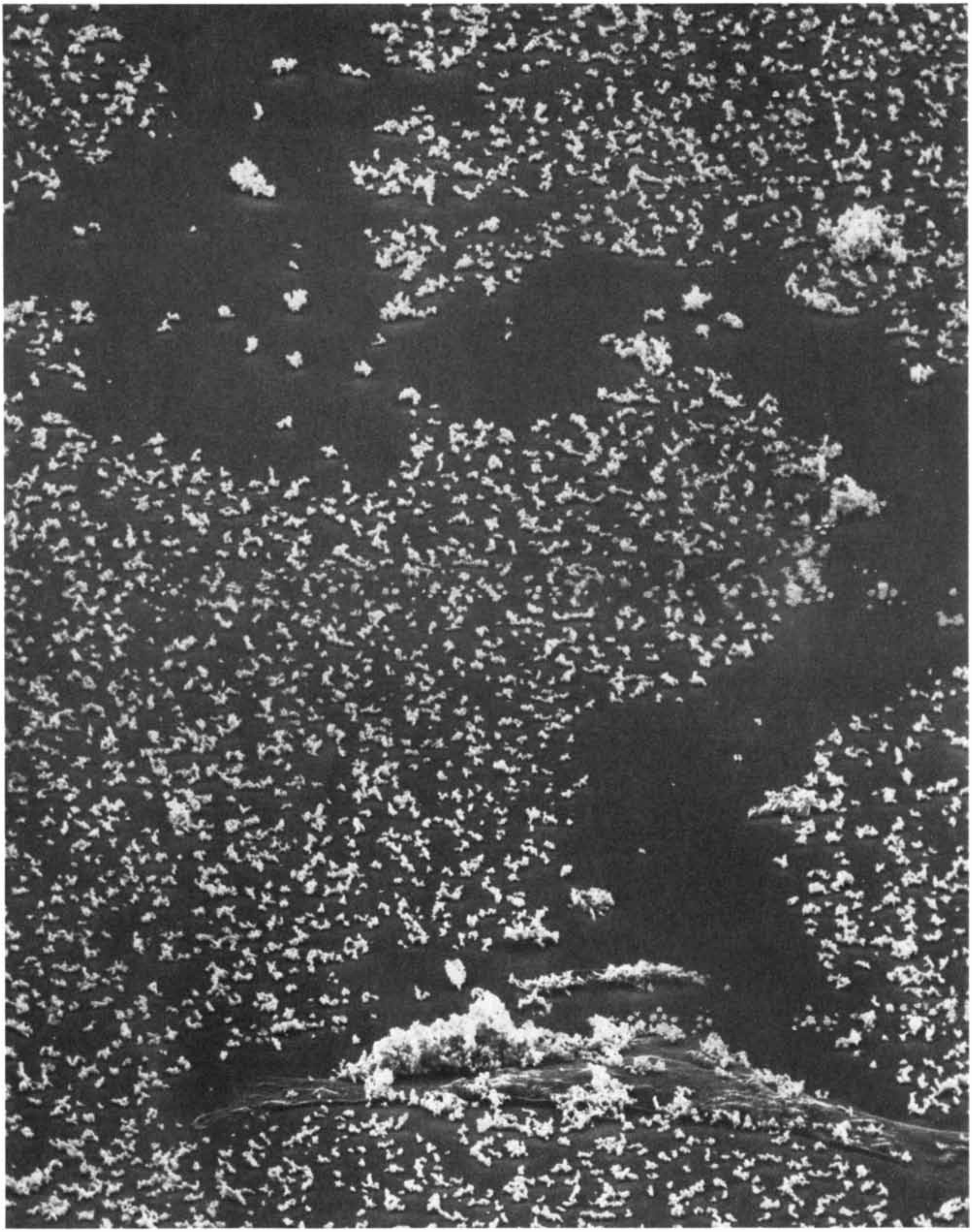
When the anastomosis is completed and the temporary clips are removed from the donor and recipient arteries, there is an immediate increase in blood circulation to the brain—as much as 100 cubic centimeters per minute, or between a third and a half of the normal flow through the internal carotid artery. Some surgeons have estimated that the blood flow through the graft may in some cases be as much as 200 cubic centimeters per minute. The increment in circulation can sometimes be observed visually in the superficial temporal artery. The brain requires more blood than the scalp, and so the diameter of the artery, which ordinarily supplies the scalp, is enlarged when its flow has been diverted into the brain. Blood flow varies with the fourth power of a vessel's radius, and so it appears that the anastomosis may eventually carry enough blood to supply most of the involved cerebral hemisphere. Clearly the new technique can make a large difference in a patient with borderline blood flow to the brain and thus in the amount of nutrition available to the brain.

The results of microvascular bypass surgery have been gratifying and the risks to the patient appear to be minimal. It has been established that in the absence of treatment some 50 percent of the patients in the high-risk group from which bypass-surgery candidates are selected will have a stroke within three and a half years. After a follow-up time averaging two and a half years, some 400 patients who underwent the operation were reviewed at the Second International Symposium on Microvascular Anastomosis in 1974. Only three had suffered a stroke. There was a significant reduction in the frequency of transient ischemic attacks for all patients who had previously experienced such episodes, and in most patients they were completely eliminated. Of some 30 patients operated on by our group at the Albert Einstein College Hospital only one has subsequently had a stroke, and that was related to sudden loss of overall blood pressure from a pulmonary embolism a month after surgery. We have not had a single procedure-related death (defined as any death within 30 days of surgery).

The impact of stroke on the patient and his family is evident. The economic cost of stroke in the U.S. in 1972 was assessed at $6.2 billion: direct expenditures for medical and nursing care and rehabilitation and indirect costs such as earnings lost through disability and premature death. It seems reasonable to expect that microvascular bypass surgery, along with other innovative approaches it has stimulated, should reduce the personal, social and economic toll of this dreaded disease.



COMPUTERIZED TOMOGRAPHY (CT) can reveal the extent of the brain damage, one element in selecting candidates for stroke surgery. The CT scan is generated by computer analysis of the attenuation of a series of X-ray beams, recorded as the X-ray source and detector move around the head. Butterfly-shaped dense region near middle of the brain is the ventricular system. Dense areas at left and right indicate tissue damaged by inadequate blood supply.



PARTICLE-FREE TRACK made by a cultured tissue cell (*foreground*) as it migrates across a glass substrate coated with tiny particles of colloidal gold is shown in this scanning electron micrograph made by the author at the Cold Spring Harbor Laboratory. The cell ingests the gold particles in its path until its capacity to do so is saturated; the particles then accumulate on top of the cell in large clumps that are occasionally pinched off and deposited near the track. In this way the migrating cell creates a graphic record of its wanderings that can later be examined at low magnification under the light microscope. Because this new approach to the study of cell migration involves both phagocytosis (the ingestion of particles) and kinetics (the study of motion), the author has proposed to call it phagokinetics.

The Tracks of Moving Cells

Migrating tissue cells will remove particles of gold from a particle-coated substrate, creating a graphic record of their movements that reveals a surprising degree of order

by Guenter Albrecht-Buehler

The ability of animal cells to move from one place to another plays a crucial role in many biological phenomena. Embryonic development depends on the migration of cell populations to precisely defined places at exact times. The defense of the body against bacteria and other invaders requires the movement of white blood cells through the walls of blood vessels to defend the besieged tissues. Wound healing involves the migration of cells that bridge and eventually close the wound. And most cancers might be considered relatively harmless growths if it were not for the migration of tumor cells throughout the body, forming myriad new tumors.

Cells can be observed moving around inside organisms only in rare cases such as the developing embryo of the sea urchin, and then only at relatively low microscopic resolution. Cell migration is therefore better studied outside an organism with cells cultured in glass or plastic dishes. At first the movement of cultured animal cells appears to be largely random, at the whim of the powerful forces that dominate the microscopic world. The virtually weightless moving cells are exposed to violent Brownian motion, strong electric fields, surface tensions, osmotic pressures and the forces released by molecular rearrangements. Nevertheless, the analysis of the paths of moving cells has revealed a surprising logic and order in the cells' behavior.

The technique of isolating animal cells from embryonic tissue and growing them in dishes in a special nutrient medium has been known since 1907 and is a common practice in biological laboratories. When the cells have been removed from the tissue, separated from one another with a protein-cleaving enzyme and suspended in a nutrient fluid, they are tiny spheres. When they are transferred to a glass or plastic dish, they attach themselves to the bottom, flatten out and grow and divide until within a few days they have filled the available surface area. If they are to be given more surface area for further growth, they are distributed among several new dishes. If after repeated transfers the cells do not change their morphology or stop growing, they are considered to be an established cell line.

Most of the cells I have studied while working at the Cold Spring Harbor Laboratory are from a cell line established some 15 years ago by George J. Todaro and Howard Green of the New York University School of Medicine. Because the establishing procedure was to transfer 3×10^5 cells into new dishes every three days, the cells are designated 3T3. They resemble fibroblasts, the cells that connect other cells in animal tissue. When the movement of the cells is to be studied, they are placed on the horizontal glass window of an observation chamber, which contains a nutrient fluid maintained at 37 degrees Celsius. The chamber is thin enough to place under the light microscope, and the living cells are observed through the window. When the morphology of the cells is to be more closely examined without movement, the cells are grown on other supports and are processed for scanning or transmission electron microscopy.

The movement of animal cells can be defined as a continuous change in cell form, and migration can be defined as a sequence of such changes. The changes are so slow, however, that time-lapse photography is required to bring them into the observer's accustomed range of movement perception. The task of studying animal-cell motility is thus to analyze the logic and mechanisms of a slow-moving, self-deforming object.

There are two basic approaches to the analysis of cell movement. One is to study the structure and modes of action of the skeletal elements of the cell and the extensions of the cell surface that determine the morphology of the cell from moment to moment. The other approach is to focus on the biological functions of these elements, including their responses to the environment of the cell, their interactions with one another and the rules underlying the time-sequence of their actions. In my experiments I have taken a primarily functional approach by making the paths of actively moving cells visible and then searching for regularities in the cells' movements.

The paths of the cells are made visible by exploiting the curious tendency of certain types of cells (including 3T3 cells) to phagocytize, or ingest, tiny particles of colloidal gold (or any other particles). A particle-coated substrate for the cells is prepared by reducing a gold salt in a water solution; the gold particles precipitate out of the solution and form a dense and even layer over the surface of a glass cover slip. The solution is then replaced with culture medium, and the cells are seeded onto the particle-coated cover slip. As the cells migrate they remove and ingest the particles in their path, leaving particle-free tracks behind them that can be clearly seen as black lines under the microscope with low magnification and dark-field illumination.

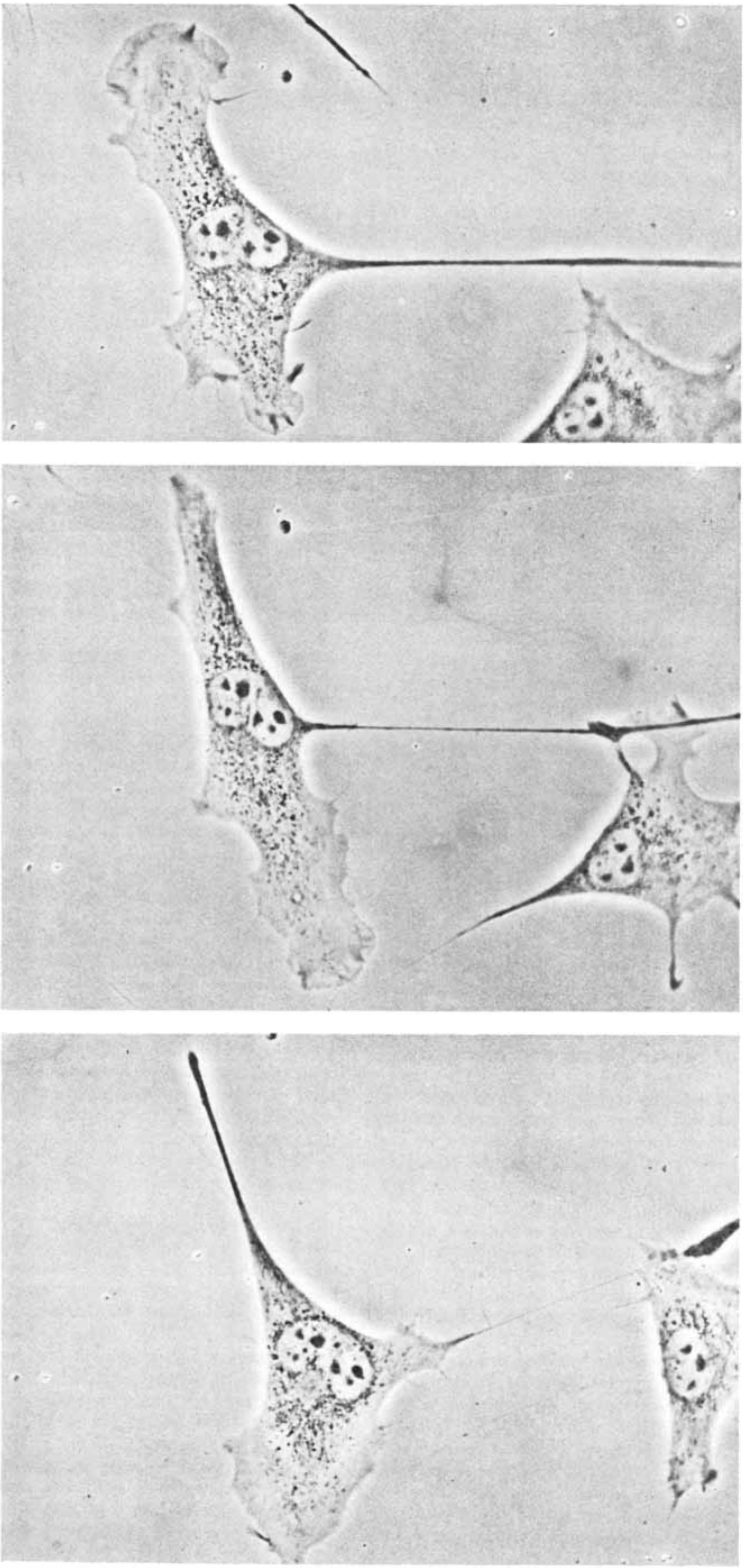
Because the formation of such tracks involves a combination of phagocytosis and kinetics (the study of motion) I have proposed that it be called phagokinetics. The advantage of phagokinetic patterns for the study of animal-cell movement is that large numbers of individual cells automatically record their own migrations, without the need for the investigator to observe them continuously. Of course, compared to the inside of an organism, the particle-coated environment is quite unnatural, but not much more so than uncoated substrates of glass or plastic. I therefore believe that the experimental advantages offered by phagokinetic tracks for the study of cell behavior outweigh the misgivings one might have about the exposure of the cells to the gold particles.

When spherical 3T3 cells are seeded onto a particle-coated substrate, one of their first actions is to send out various surface extensions, primarily needlelike rods called filopodia. The filopodia wave about like rigid sticks until they attach themselves to the substrate a few micrometers away from the cell body. If gold particles are present, the filopodia come in contact with them and stop waving, and after a minute or so they retract into the cell body, dragging the gold particles with them until the particles are ingested by the cell. The process continues until the cell has cleared away almost all the particles from a ring surrounding it. Then newly projected filopodia attach themselves to the glass surface and the cell flattens into the cleared area.

Once the cell has flattened it begins to migrate, clearing particles out of its way and leaving a track through those that remain. It clears the particles with various surface extensions, such as the sheetlike lamellipodia or the hemispheric blebs, transporting them over the surface of the cell toward the cell nucleus. The particles are engulfed and assembled in a ring around the nucleus, and when the capacity of the cell to swallow them is apparently saturated, they accumulate in clumps on the cell's upper surface. The clumps are occasionally pinched off and deposited near the track.

The fate of the ingested particles is not yet known. If they are reextruded, they must be replaced by newly ingested particles, because the ring of particles around the cell nucleus never disappears. One might think that overfeeding cells with gold particles might harm them, but I have found no detrimental effects even after exposures of several weeks. On cell division the internalized gold particles are neatly split between the two daughter cells. If the cell divides after forming a track, the track will branch because both daughter cells migrate along separate paths. Thus the three major types of animal-cell movement one can observe in cultured cells take the form of three types of graphic patterns on the gold-particle-coated substrate: flattening of the cell is represented by a particle-free ring, migration of the cell by a particle-free line and division of the cell by the branching of such lines.

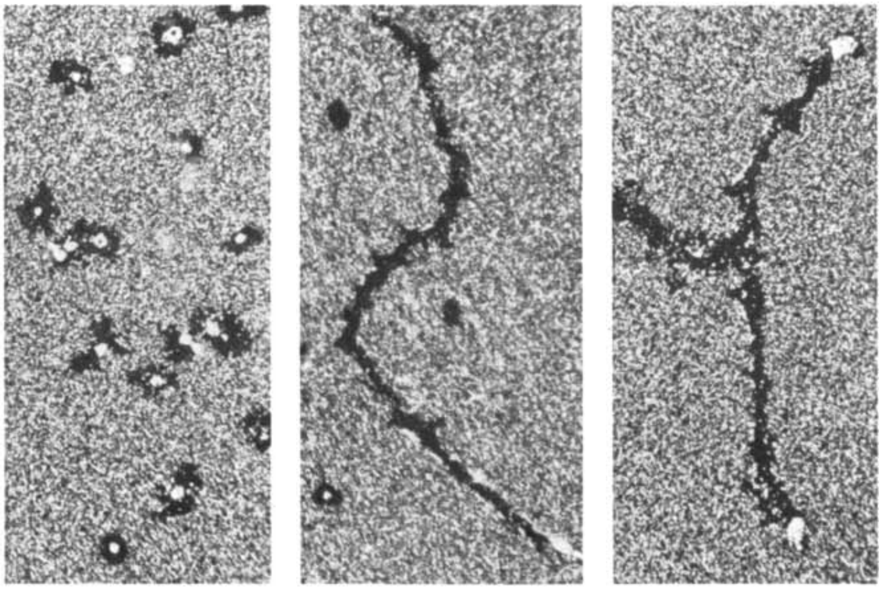
Why do the cells collect and ingest the gold particles? It appears that the filopodia probe the firmness of a substrate



DIRECTIONAL CHANGE made by a migrating cell in tissue culture is shown in this sequence covering an interval of two hours. The cell possesses a "tail" (*dark structure in top frame*) and a leading edge that resembles a ruffled sheet. To change direction the cell shifts the location of its leading edge to the lower left while absorbing its tail; it then develops a new tail out of the former leading edge. The orientation of the cell nucleus (*indicated by the dark spots*) changes very little as the cell rotates, suggesting that the nucleus moves in parallel with the surrounding cytoplasm. The magnification is 850 diameters.

by tentatively pulling at the first solid object with which they come in contact, rather like a mountaineer testing a foothold before trusting it with his life. When the particle is loose, the retractile mechanism that generates the testing force apparently pulls the particle all the way back into the cell. This interpretation is an attractive one because it provides a simple way for a single animal cell to determine whether or not a given object with which it comes in contact is suitable for anchorage.

To further test this hypothesis I seeded cells onto a substrate to which gold particles had been firmly attached at some distance away from the cell. Once the filopodia had attached themselves to the gold particles, they apparently found them suitable for anchorage, and the cell body began to extend in the particles' direction. Only the tips of the filopodia had to come in contact with the fixed gold particles before the cell body began to move, a finding consistent with the notion that the function served by the filopodia is primarily to probe the environment.



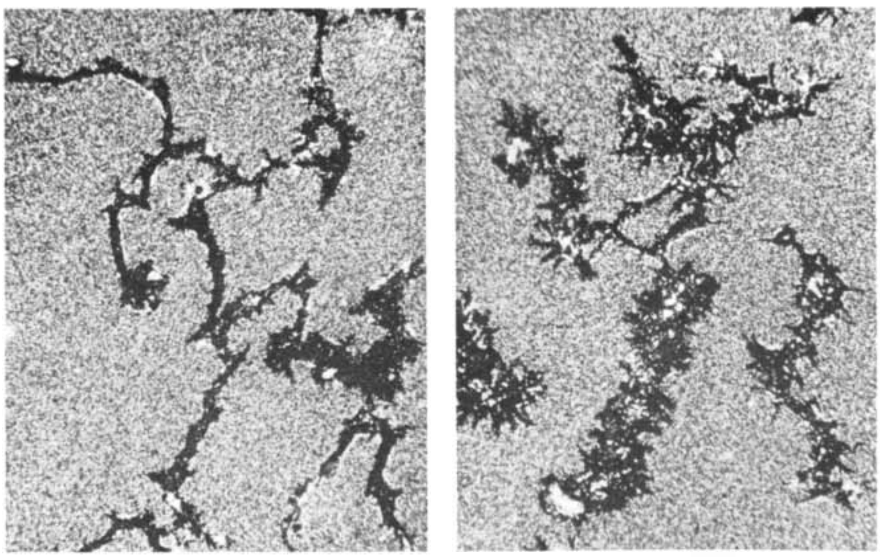
GRAPHIC PATTERNS made by migrating tissue cells on a substrate coated with gold particles are shown in these light micrographs. Because of the dark-field illumination the tracks appear black and the particle-filled cells at the ends of the tracks glow brightly. The tracks are of three basic types: the flattening of cells is represented by particle-free rings (*left*), cell migration by a continuous linear track (*center*) and cell division by a branching track (*right*).

On observing the phagokinetic tracks made by different cells from the same cell line I found that they differed not only in their gross form but also in certain of their details. For example, some tracks were smooth and others were decorated with "thorns" that jutted out on either side, suggesting that those cells had repeatedly moved sideways as they migrated. There were even more striking differences in overall track patterns between different cell lines, such as virus-transformed 3T3 cells, human fibroblasts and monkey epithelial cells, suggesting that contrary to the prevailing belief, the migration of cultured animal cells might be nonrandom. After all, random tracks should look generally alike, no matter what generates them.

A still stronger indication that cell migration was nonrandom came from the observation of branching tracks laid down by dividing cells. Comparing the major directional changes along the track made by one daughter cell with the track made by the other daughter cell, I found that one track is often a slightly distorted mirror image of the other! Occasionally the tracks are almost perfect reflections of each other; more often one has to expand the track of the "slower" daughter cell in one's mind to recognize the relation. The mirror symmetry of the tracks is therefore topological rather than literal.

I have found mirror symmetry in about 40 percent of the branching tracks, identity in about 20 percent of the tracks and no correlation in the remaining 40 percent. Although one can think of many outside forces acting on moving cells that might cause the branching tracks of daughter cells to appear unrelated, there are only two explanations for the existence of mirror-image or identical branches: either the daughter cells are actually migrating in this closely related fashion or the phenomenon is the result of random coincidence. One way to test for the role of chance in such a situation is to devise a Monte Carlo computer program, which generates random numbers to determine the distances and angles between consecutive directional changes of simulated tracks. Because such simulations inevitably involve a large number of assumptions a more satisfactory way to approach the problem is to search for additional experimental evidence or theoretical explanations for why two daughter cells might move in mirror-symmetrical or identical ways.

In seeking such a theoretical explanation let us look more closely at the process of mitosis, or cell division. In this process the chromosomes in the cell nucleus are duplicated and arranged along the plane of division passing through the center of the cell. The duplicated halves of each chromosome are then pulled



CHARACTERISTIC TRACKS are laid down by different tissue-cell lines, suggesting that cell migration is nonrandom. The smooth tracks at the left were made over a two-day period by cultured 3T3 cells; the "decorated" tracks at the right were made by 3T3 cells that had been transformed into cancer cells by infecting them with polyoma virus. The "thorns" on these tracks suggest that polyoma-transformed cells repeatedly moved sideways as they migrated.

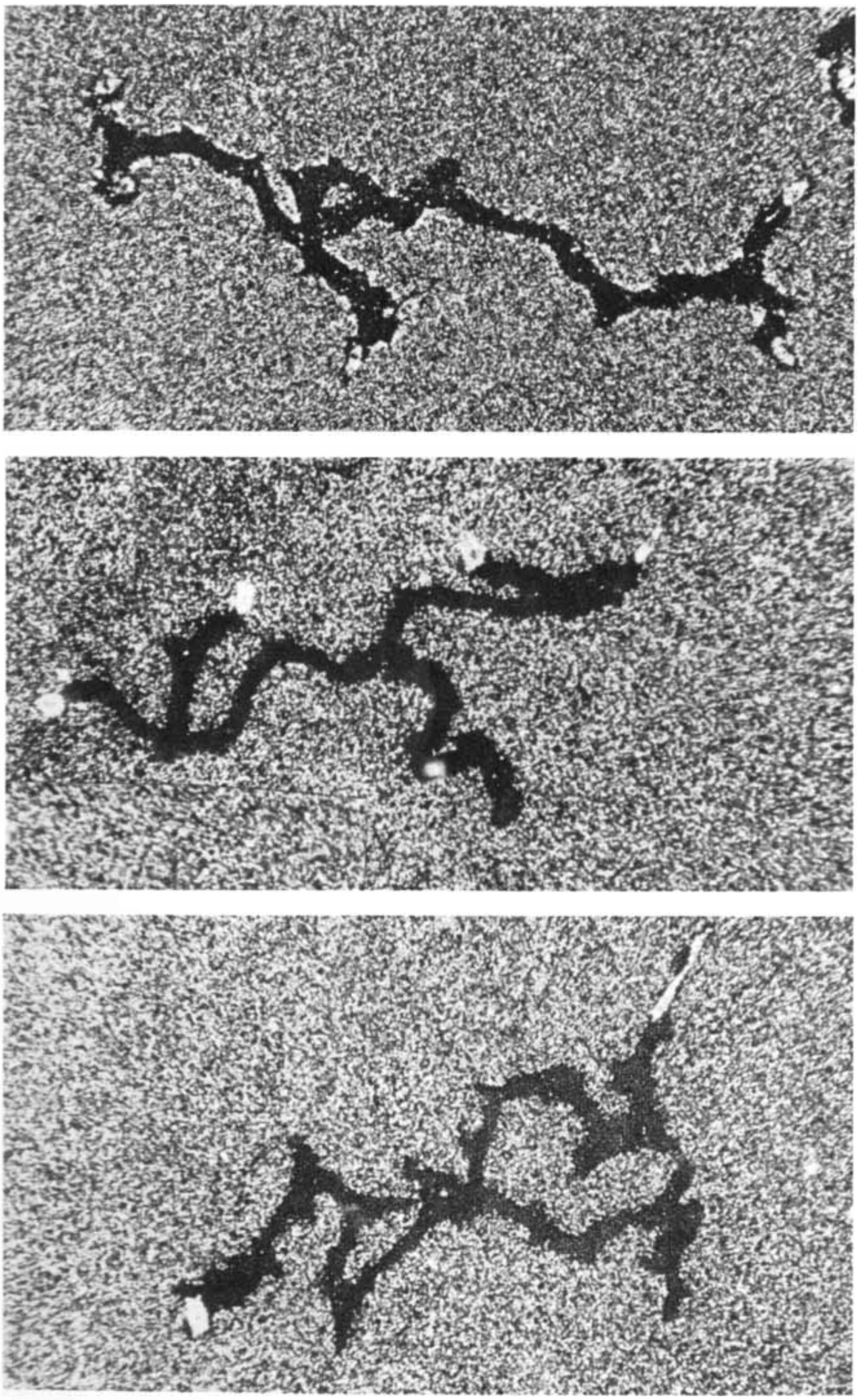
apart along the fibers of the mitotic spindle, so that as each daughter cell divides it receives an equal share. By itself, however, the separation of chromosomes is not enough to provide the daughter cells with everything they need to be functional. Each cell also needs copies of vital cellular organelles and other components, and these too must be duplicated and distributed into the two compartments of the mother cell that are destined to become the daughter cells.

Although we can assume that the cellular organelles are equally partitioned between the two daughter cells, it does not follow that these components are arranged in an identical way. On the contrary, because the mitotic spindle is an almost perfect mirror-symmetrical structure the duplicated pairs of organelles might be assembled and separated along the plane of cell division in the same way that the chromosomes are. After the cell components had been distributed in this manner the three-dimensional intracellular organization of one daughter cell would be the mirror image of the other.

For the two daughter cells to then become identical, as it is generally assumed they are, there would have to be an extremely complex redistribution of the cellular components, a task analogous to transforming a right hand into a left hand by rearranging the bones, the muscles, the blood vessels and the fingernails. Applying Occam's razor to the situation (that is, eliminating all but the simplest explanation suggested by the available evidence), we can assume that the daughter cells avoid such complications and remain mirror images of each other. If this assumption is correct, it seems logical that the structural complementarity of the daughter cells might cause them to migrate in mirror-symmetrical ways. Indeed, it is interesting to extend this reasoning beyond the field of cell movement and speculate that the mysterious bilateral symmetry of all animals and most plant parts, such as leaves and petals, can be traced back to the mirror symmetry of certain dividing cells in the early stages of embryonic development.

Evidence to support the mirror-symmetrical organization of daughter cells was provided by looking at the pattern of fibers that makes up the internal skeleton of the cells. Two years before I discovered the phagokinetic tracks of 3T3 cells Elias Lazarides and Klaus Weber at the Cold Spring Harbor Laboratory developed an antibody specific for the protein actin, which makes up one type of intracellular structural fiber (the microfilaments). When the antibodies are applied to fixed cells whose membranes have been made permeable to large molecules, they pass through the cell membrane and coat the microfilaments inside. A fluorescent label is then attached to the coated fibers, making them glow brightly when they are observed in the fluorescence microscope. This technique, called indirect immunofluorescence, is presently utilized by many investigators to make visible a variety of subcellular fibers.

When I applied the indirect immunofluorescence technique to revealing the pattern of microfilaments within pairs of daughter cells, I found that the arrangement of fibers within one daughter cell was indeed the mirror image of the arrangement in the other. Rigorous symmetry could not be expected because the daughter cells were usually caught at different stages of movement at the moment of fixation. Once again, however, I found cases where the inter-



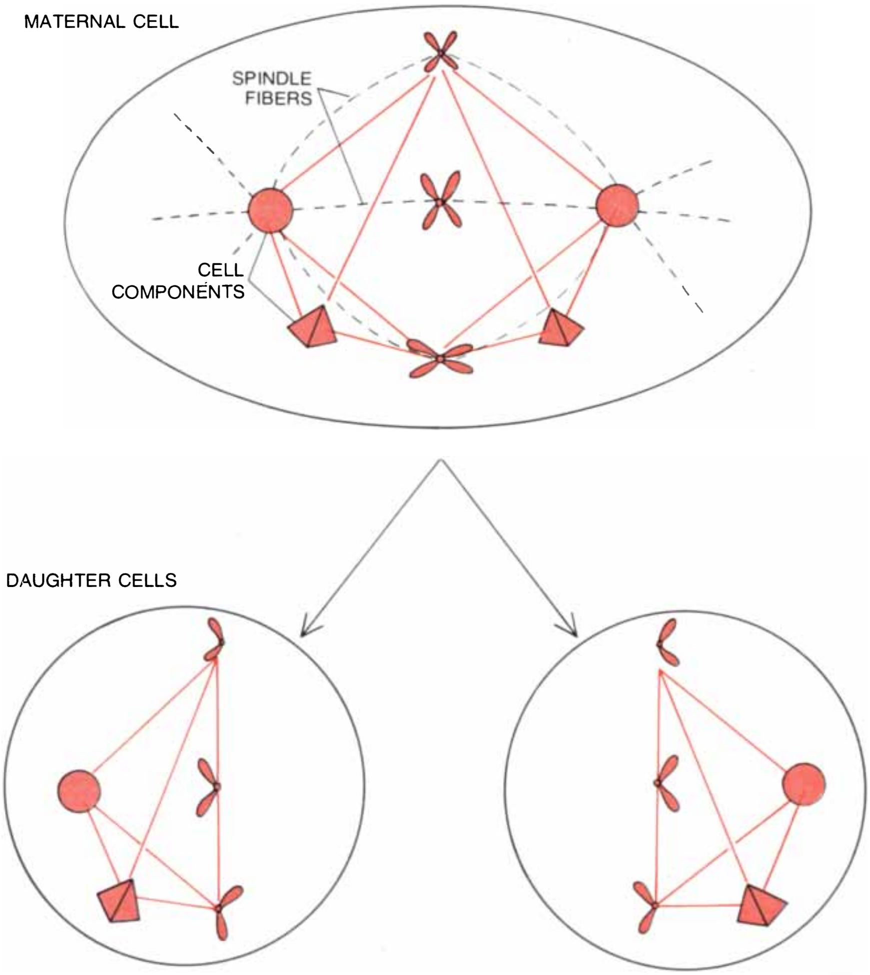
STRIKING SIMILARITIES are apparent in the branching phagokinetic tracks of daughter cells derived from the division of a parental cell. In the two micrographs at the top the tracks of two daughter cells and four granddaughter cells are mirror-symmetrical. In the micrograph at the bottom the tracks are identical: they are related by rotation rather than mirror symmetry.

nal organization of the pairs of cells appeared identical or unrelated.

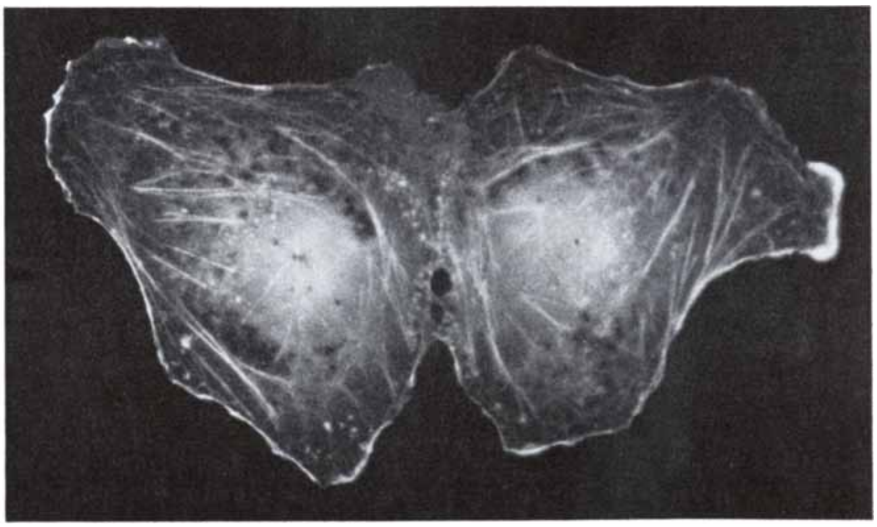
The cases of identity or unrelatedness in the phagokinetic tracks and the microfilament patterns of daughter-cell pairs can perhaps be explained by simple analogy. Observe the shadows made by your hands as you keep one stationary and rotate the other about its axis. Each hand is an almost perfect three-dimensional mirror image of the other, yet the shadows may appear unrelated when one hand is in certain rotational positions with respect to the other. And when one hand is rotated 180 degrees with respect to the other, the shadows will appear identical. A similar argument may well apply to tracks and microfilament patterns that are not mirror-symmetrical. Like shadows, the tracks and fluorescent patterns are two-dimensional projections in the plane of migration of the postulated three-dimensional mirror symmetry of the daughter cells. If immediately after cell division the two still-spherical daughter cells rotate with respect to each other and subsequently flatten back onto the substrate, then the subcellular components of one daughter cell will be organized with respect to the plane of migration in such a way that its mirror-symmetrical relation to the other daughter cell can no longer be recognized. As a result the phagokinetic patterns of the two cells may turn out quite different. In the special case of the rotation of one daughter cell through 180 degrees, the tracks and microfilament patterns of the daughter cells may appear identical.

The fact that daughter cells that are far away from each other still make directional changes in closely related ways suggests that these changes may be predetermined in some way by the mother cell. Indeed, certain characteristics of the tracks of a particular cell can be found repeated by its four granddaughter cells. Since one area of a plastic culture dish containing a well-stirred nutrient solution would seem to be about the same as any other, normal culture conditions do not appear to offer any outside inducement for cell displacement. Thus the impulse of cells to migrate appears to be predominantly due to intracellular influences. Perhaps the cells "remember" and carry out parts of instructions that would have been appropriate within the embryo from which they were originally obtained. The finding of symmetrical or identical track branches suggests that the predetermination of movement includes even the directional changes made by the cell during displacement.

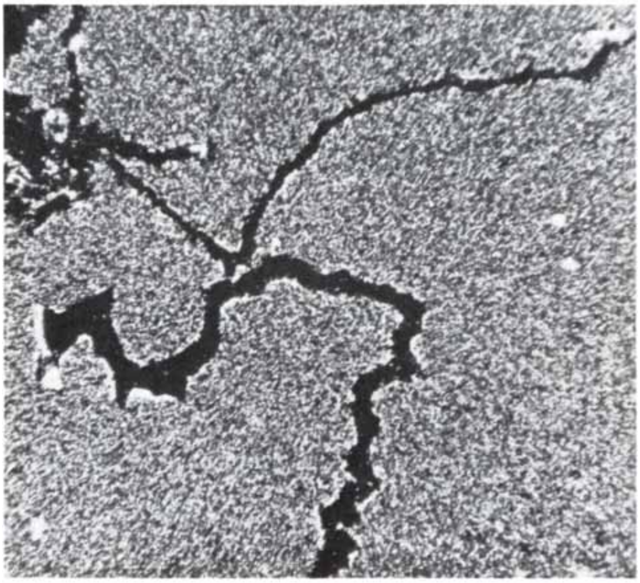
The collision of migrating cells provides another striking example of nonrandom behavior in animal cells. When a 3T3 cell comes in contact with another 3T3 cell or with a nonmigrating cell, the

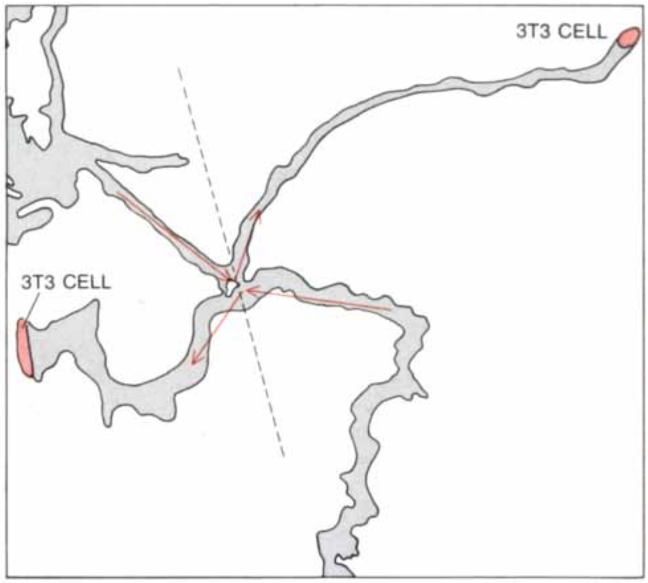


MODEL OF CELL DIVISION illustrates schematically how the mirror symmetry of the mitotic spindle (*broken lines*) might provide an explanation for the tendency of daughter-cell pairs to migrate along mirror-image paths. According to the model, chromosomes and other cellular components are duplicated and distributed to opposite sides of the spindle during cell division, so that the organization of one daughter cell becomes the mirror image of the other.



ACTIN-CONTAINING FIBERS GLOW within a daughter-cell pair, revealed by the technique of indirect immunofluorescence. Although rigorous symmetry between the actin patterns of the two cells cannot be expected because each cell is in its own state of movement, the left- and right-handedness of the patterns is striking. This finding suggests that the organization of the two daughter cells is indeed mirror-symmetrical. Magnification is some 1,400 diameters.

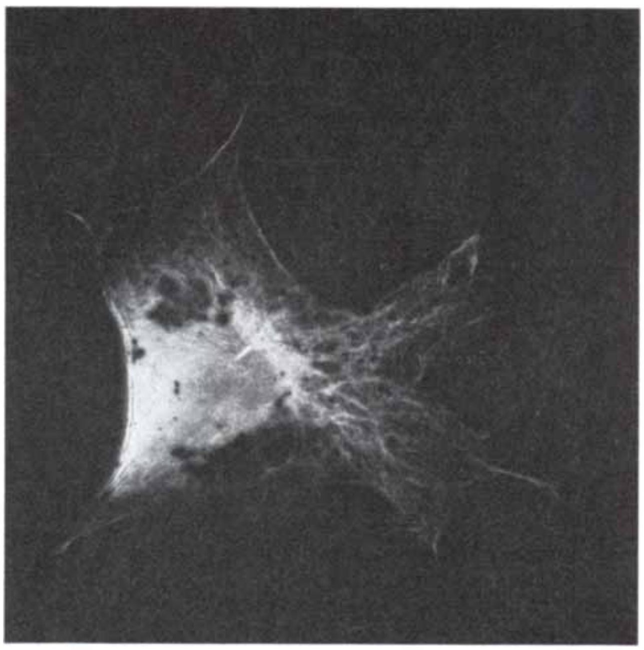




TRACKS OF COLLIDING CELLS are curiously reminiscent of the elastic collisions of solid objects such as billiard balls. In the micrograph at the left two migrating 3T3 cells, magnified 94 diameters, collide in the center of the frame. After two hours of contact the two cells reorient and move away from each other as if they were rebounding elastically. The map at the right shows how each cell preserves one component of its movement while inverting the other. This strategy would be a logical one for a 3T3 cell migrating inside an embryo because it is unlikely that the cell would be pushed back to its starting point by multiple collisions with other migrating embryonic cells.

sheetlike lamellipodium at the front end of the cell ceases its ruffling movement. The cell then slowly backs away from the collision area, leaving needlelike extensions called retraction fibers that bridge the widening gap between the two cells. Finally, after two or more hours have passed, new lamellipodia form elsewhere along the cell's circumference and the cell moves off in a different direction.

The phagokinetic tracks left by colliding animal cells reveal a remarkable feature. When a 3T3 cell collides with another cell, the new lamellipodia do not form just anywhere on the cell circumference; they form at specific locations so that the cell moves away from the impact area as if it were elastically reflected from the target. Specifically, if the direction of cell migration is represented as a vector having a tangential component and a perpendicular (impacting) component, then after collision the tangential component of the cell's movement remains unchanged and the perpendicular component is inverted.



CELLULAR STRUCTURES implicated in the control of cell migration can be made visible by the technique of indirect immunofluorescence. The micrograph at the left shows the pattern of actin-containing fibers (microfilaments) in an actively moving cell. The parallel arrays of microfilaments extending toward the rear of the cell are aligned with the current direction of cell movement. They appear to serve as "rails" along which the bulk of the cell slides during displacement. The micrograph at the right shows the pattern of the cellular components and fibers consisting primarily of the protein tubulin. The spaghettilike fibers are microtubules. The brightly glowing rod located close to the cell nucleus is the "primary cilium," which grows out of a cylindrical structure called a centriole. Although the function of the primary cilium is unknown, its orientation (and hence that of the centriole) is precisely correlated with direction of cell movement.

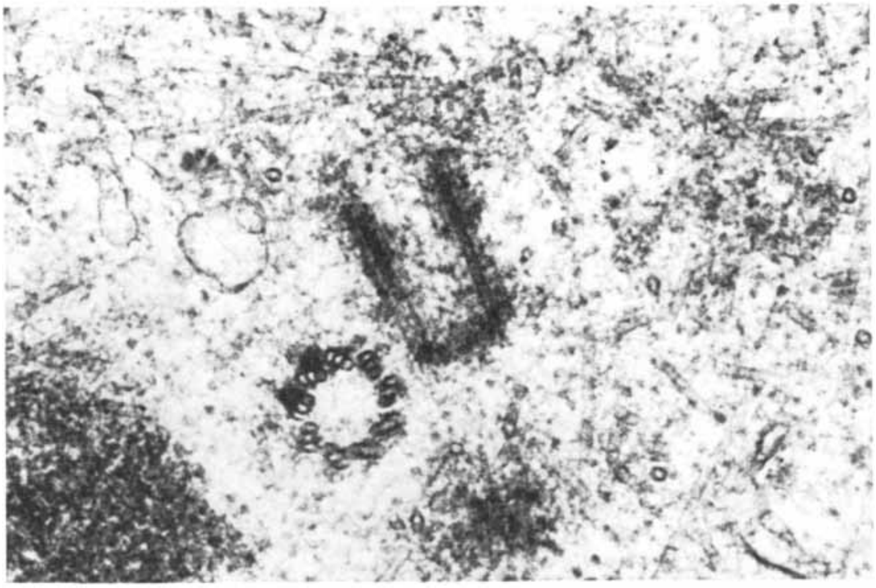
Although the interaction of migrating cells obviously has little in common with the physics of colliding billiard balls, the strategy of preserving at least one component of a cell's direction after a collision is logical in that it reduces the chances that a cell migrating inside an embryo will be pushed back to its starting point after multiple collisions with other cells. In sum, it appears that although the displacements of 3T3 cells are largely predetermined by intracellular influences, extracellular signals such as cell-cell collision are capable of overriding the movement instructions that the cell may have obeyed until then.

The ability of migrating cells to predetermine movements and to reorient migration after collision implies the existence of specialized cellular components that are capable of storing and revising movement programs and of controlling their execution in time by the motile apparatus of the cell. I began to search for such components by looking for cell structures that were conspicuously oriented with respect to the phagokinetic tracks. Working with the technique of indirect immunofluorescence, I found that there are indeed structures within the cell that fulfill this criterion.

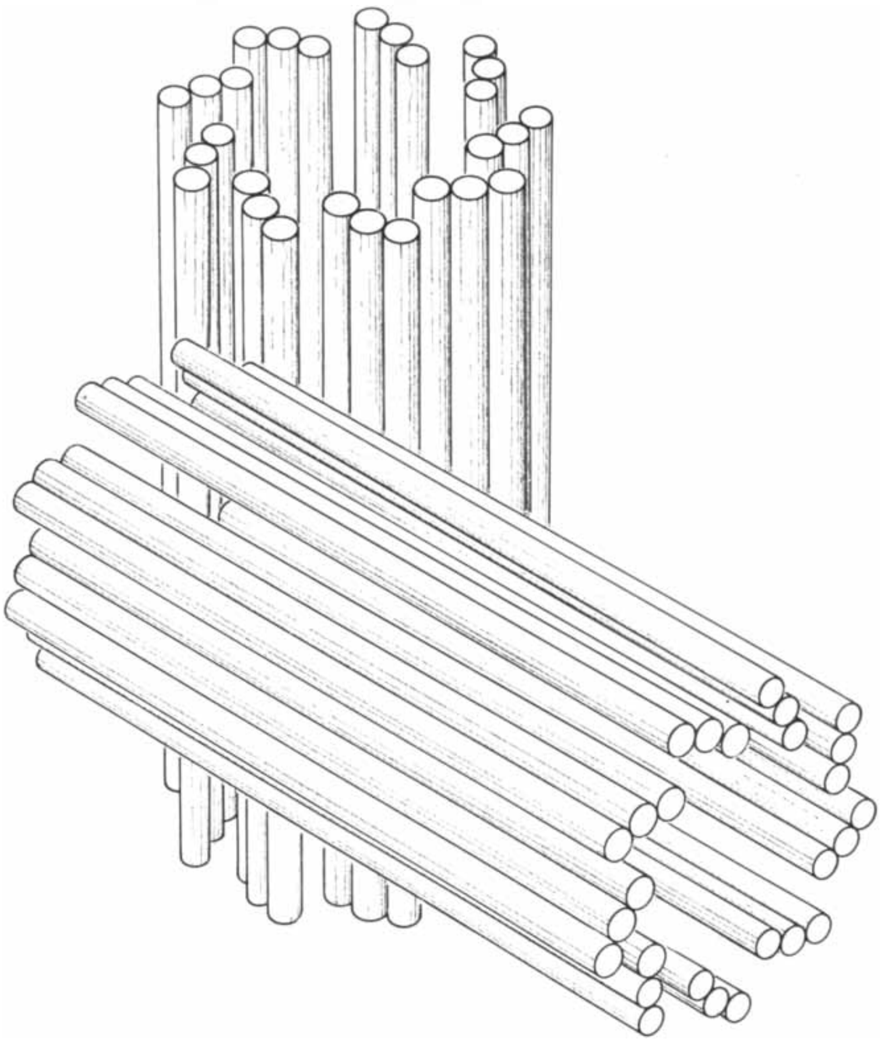
In migrating 3T3 cells "bundles," or arrays, of microfilaments are oriented parallel to the current direction of cell movement. In most instances the microfilaments extend from underneath the nucleus to the rear of the cell, where they often converge into the elongated "tail" typical of migrating fibroblasts. The microfilament bundles are located primarily near the substrate, and they appear to serve as "rails" along which the bulk of the cell slides during displacement. According to this model the used portions of the rails converge into a tail in the rear and are dismantled as the tail retracts into the cell by a mechanism similar to that of muscle contraction. Sliding on the rails apparently allows the nucleus and the cytoplasm of the cell to move in parallel, thereby reducing the danger of knotting up the delicate intracellular organization. Such parallel movement is illustrated by the fact that the nucleus of a migrating cell often remains in its initial orientation even after the cell has rotated its leading edge [*see illustration on page 70*].

My search for a cellular component that might be involved in directing the movement of migrating cells led me to investigate the centrioles, the cylindrical bodies toward which the fibers of the mitotic spindle converge during cell division. Centrioles are made up of a strikingly geometric arrangement of nine triplets of the tiny tubular structures called microtubules. They are generally found in pairs, with one centriole lying at a right angle to the other. Each daughter receives one pair at cell division and subsequently manufactures another.

In addition to serving as the poles of



TWO CENTRIOLES are shown at right angles to each other in this transmission electron micrograph. The centriole at the left is in cross section; the one at the right is cut parallel to its long axis. Centrioles are cylindrical structures made up of a strikingly geometric arrangement of nine "blades." Each blade consists of three microtubules. Magnification is 70,000 diameters.

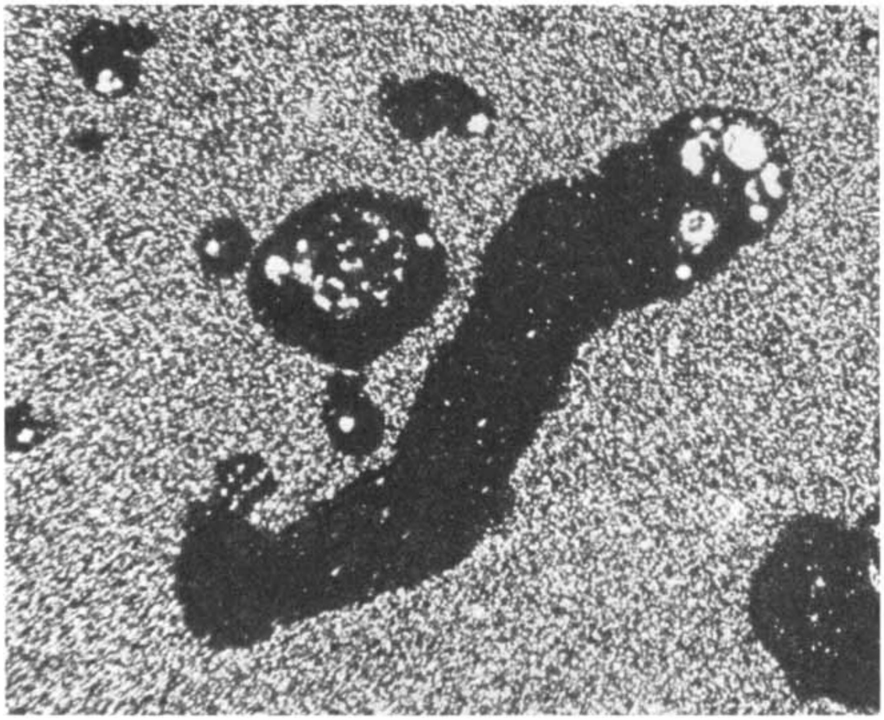


THREE-DIMENSIONAL MODEL of a pair of centrioles is somewhat schematic; additional proteins (not shown) are associated with the inside and outside of the cylinders. The ability of centrioles to organize microtubules into the leading edge of a moving cell suggests that they play a major role in guiding cell migration, although the precise mechanism is not known.

the mitotic spindle, centrioles are found in a variety of cellular structures, such as cilia and flagella. Many cultured animal cells also form what is known as a "primary," or internal, cilium by elongating one of the two centrioles within the cell body. Once assembled, the shaft of the primary cilium is enclosed in a vacuole, or membranous sac, and stored close to the cell nucleus. Its biological function has long been a mystery.

To make visible the primary cilium inside migrating cells I again applied the indirect immunofluorescence method, but this time I used an antibody against tubulin, the major structural protein of cilia, flagella, centrioles and microtubules. In nearly every case the primary cilium of migrating 3T3 cells was pointed along the current direction of the track. This relation stimulated a number of conjectures, ranging from the possibility that the cilium shaft is passively oriented by cytoplasmic streams that arise during cell displacement to the possibility that the primary cilium, or at least its basis, the centriole, might play a direct role in guiding the direction of the migrating cell.

This latter speculation was supported by the work of other investigators, who had found that centrioles in animal cells may act as organization centers for microtubules. The microtubules in turn determine the polarity of the cells, that is, they determine which end of a cell is the front and which is the rear. Moreover, some additional evidence for the role of centrioles in guiding migration is provided by the fact that plant cells, which are unable to migrate, do not possess centrioles.



MIGRATING CELL GROUP from the cell line designated PtK1 formed this phagokinetic track over a four-day period. A nearby cell group did not migrate, and a number of individual cells (*visible as small white dots inside black rings*) displaced themselves very little. How cells that migrate poorly by themselves can associate and begin to move in a seemingly coordinated fashion is an unanswered question. Such interactions are crucial, however, for embryonic development, which depends on the migration of groups of cells to precisely defined locations.

CLOSEUP VIEW of a group of PtK1 cells migrating together is shown in this scanning electron micrograph magnified 820 diameters. The gold particles coating substrate are also visible.

The precise mechanism by which migration in animal cells is programmed is not yet known, but a few general remarks about the postulated program can be made. First, the successive execution of timed instructions in the program requires the existence of a "clock." Such a clock mechanism has long been presumed to exist because cells have a precisely timed cell cycle. Second, the program is likely to be a physical structure that can exist in mirror-image forms because that would explain the mirror-symmetrical track patterns of daughter-cell pairs. Third, the program structure is probably associated in some way with the centrioles.

One of the remarkable properties of animal cells is their ability to cooperate to form multicellular organisms, and indeed certain types of cultured animal cells have recently been found to coordinate their migrations. I have observed small groups of the cells designated PtK1 migrate together in culture, and other cell lines isolated by Green form sheets of cells that can deform and flow with respect to one another. The establishment of such cell lines has now made it possible to study in tissue culture the collective migration of cell populations, a phenomenon essential to embryonic development.

Taken together, the new perceptions of animal-cell migration provided by phagokinetic analysis strongly suggest that the movements of such cells are largely predetermined by a hypothetical program structure within each cell. Contrary to the prevailing view of cell migration, phagokinetics has revealed geometry in what was thought to be chaos, and signs of determinism where many had expected blind randomness.

Attention: **Hawaii, Alaska, Washington, Oregon, California, Texas, Louisiana, Mississippi, Alabama, Florida, Georgia, North Carolina, Virginia, Maryland, Delaware, New Jersey, New York, Connecticut, Rhode Island, Massachusetts, New Hampshire, and Maine**

As you are well aware, your states have a special asset and a special responsibility. You have seacoast. A lot of your coast or a crucial bit of it, as the case may be, is marshland. Once upon a time marsh (where ladies' and gentlemen's shoes get all muddy) was considered *ugh!* Since property values reflected this, well-meaning folk proceeded to upgrade their possessions. Thereby, as you have since learned, they downgraded your coast's ability to sustain the sea life without which the ocean is merely a wet, chilly, and dangerous waste of valuable space.

Now you are probably spending a fair amount to halt further deterioration. You have biologists surveying your coastal marshes and lawyers trying to back them up against continued destruction. Before there can be a master plan, it is necessary to ascertain where there is room for compromise and where compromise would be tragic. So your biologists spend much time in travel in order to take inventory.

We have omitted South Carolina from the above roster because our purpose here is to call to the attention of the other coastal states how that state in two years for $65,000 has done as much inventorying of tidal wetlands as has cost four years and $1.2 million elsewhere.

South Carolina's method depends on the differences in color with which a certain Kodak aerial film renders different vegetative communities in accordance with their reflectivity into the near infrared. From a minimal amount of ground truth, aerial photography within an hour of low tide has proved right from 6,000 feet for distinguishing eight types of land. These are then entered on the topographic maps. Data are digitized for storage.

The name of the film is Kodak Aerochrome infrared film 2443.

More important to you, the name of one of the South Carolinians who would be proud to give colleagues in other states details on how they did it is Robert H. Dunlap, Jr. His address is Office of Conservation and Management, South Carolina Marine Resources Center, P.O. Box 12559, Charleston, S.C. 29412.

Kodak

SCIENCE AND THE CITIZEN

Law of the Sea

For the past five years the delegates of 148 nations have sought in intermittent meetings of the Third United Nations Conference on the Law of the Sea and its committees to forge a comprehensive system of governance for the oceans. The conference has made progress on a number of issues, including the safeguarding of the traditional freedom of the high seas within the 200-mile offshore limit of jurisdiction claimed by an increasing number of nations, arrangements for foreign vessels to pass through straits that are under national jurisdiction and provisions on environmental control. The conference is, however, bogged down on several points, among them one paramount issue: the regime that is to be established for exploiting the extensive mineral resources of the deep-sea bed.

Here a polarization has arisen between on the one hand a few developed nations (including the U.S.) and on the other an assemblage of developing nations known as the "Group of 77" (although it has more than 100 members). The disagreement centers on three principal issues: the system of mining, the degree to which the production of seabed minerals should be limited and the governance of the international institutions created to manage seabed mining. On mining the developed nations favor access to the seabed minerals by qualified companies and state enterprises, whereas the Group of 77 would put mining solely in the hands of the proposed International Seabed Authority and its operating arm, known as "the Enterprise." On controlling production the developed nations favor a generally liberal policy and the developing nations seek restrictions. On governance the developed nations want the control of the seabed authority to reflect not only the desires of member governments but also the interests of such groups as producers, investors and consumers; the developing nations favor a one-nation, one-vote principle.

The Conference is now holding (in Geneva) its seventh session, which began on March 28 and is expected to continue into May. Before it is a bulky "informal composite negotiating text," consisting of a preamble, 303 articles and seven annexes. The text was issued last summer by the president of the Conference and the chairmen of its three committees. Its provisions on the deep-sea bed have been described by the U.S. as "fundamentally unacceptable."

Before the session started Ambassador Elliot L. Richardson, who is chairman of the U.S. delegation, reviewed the situation in a speech in Cincinnati. "In the life of international conferences," he said, "a moment arrives when the elements of negotiations are so arranged after exhaustive and exhausting debate that a fair compromise comes within reach.... Such a moment has been reached by the [Conference]." Pointing to the Conference's agreement that the resources of the deep-sea bed are the common heritage of mankind, that their exploitation should be managed on behalf of all mankind by an appropriately designed international authority and that the developing countries should share in the proceeds of mining the deep-sea bed, he said: "If fair compromises are reached on the first two of the key seabeds issues—exploitation and the resource policy—it will be on the basis of accommodations fairly reflecting legitimate interests in due proportion to their true weight. It follows that a fair compromise on the issue of governance must rest on the creation of a Council which operates as the executive branch of the Authority with membership, powers and voting procedures that can achieve the same kind of accommodation." He added that although seabed mining will go forward with or without a treaty, a "successful conclusion of the Conference is so important as a precedent for constructive international cooperation that the Conference cannot be allowed to fail."

Pseudoscience on the Campus

A seminar on reincarnation is being conducted this semester at Iowa State University by a professor who takes a "phenomenological" approach to psychology, lectures on astrology and teaches his students to see the "visible color of the energy field around people." Twin issues are thus raised once again at a major institution of higher learning: where does the boundary lie between science and pseudoscience and to what extent should academic freedom protect the right of eccentric faculty members to present material that most of their colleagues believe is at best outside the mainstream of their discipline and is at worst either religious or mystical special pleading or even fakery? Last fall it was revealed that a course in the biblical creationist theory as an alternative to Darwinian evolution was being taught at Michigan State University [see "Science and the Citizen," December, 1977].

David A. Weltha is associate professor in Iowa State's department of family environment, which offers courses in consumer management, housing and household equipment as well as on "the individual and the family." Weltha teaches a course titled "The Individual and Family Development" that he says emphasizes "man's potential" and the "awareness of consciousness." In it he includes instruction in astrology and its "relation of the cosmic to the mundane" and on the "aura" that surrounds an individual, which, he teaches, can be seen and can be recorded by Kirlian photography. In his seminar on "Your Former Lives" he intends "to explore the meaning of life through the reincarnation theory and regression techniques through guided fantasies." Regression in this sense means becoming aware of life before one's birth, both in utero and in possible previous incarnations. Weltha says, "I'm not a researcher. I'm simply a conveyor of information to the students and it is very controversial."

John W. Patterson, professor in the Iowa State department of materials science and engineering, has challenged Weltha precisely on the ground that because what he teaches is controversial it should not simply be "conveyed" to students but should be supported by evidence and subjected to the challenge of conflicting opinion. He believes, along with some other faculty members who have not taken a public position, that Weltha is accepted by many of his students not simply as a conveyor but as a kind of oracle with special access to the truth. He has tried unsuccessfully to arrange a public debate in which Weltha's ideas could be challenged. This semester he is giving a seminar on "Critical Judgment in the Age of Aquarius" that will examine such subjects as astrology, aura visualization and reincarnation.

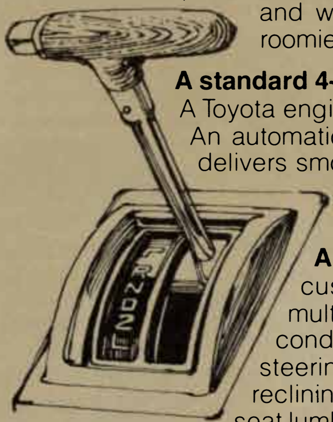
The Iowa State administration has held that the Weltha issue is one for the faculty to handle. Patterson's request for faculty action was considered last fall by a subcommittee of the faculty council, which recommended that the matter not be included on the council's agenda during this academic year. Patterson believes the faculty's avoidance of the issue is hypocritical and represents a perversion of scholarship and academic freedom. Iowa State is probably not alone in tolerating teaching on the fringe of psychology. Paul Kurtz of the State University of New York at Buffalo, chairman of the Committee for the Scientific Investigation of Claims of the Paranormal, believes that courses and material of the kind taught by Weltha are being presented at many colleges and universities in an effort to cater to student interest in parapsychology and the occult in general.

Supermass

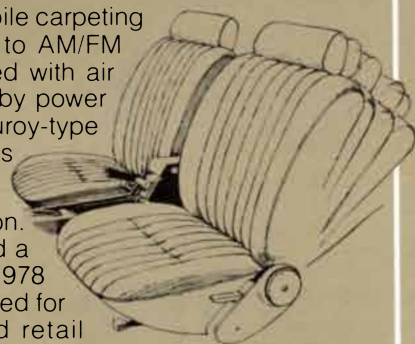
The giant elliptical galaxy M87, which at a distance of only 50 million light-years is virtually a neighbor of our own galaxy, interests astronomers

THE EXPENSIVE TOYOTA.

Introducing the limited edition Cressida. For a Toyota it may seem expensive. Compared to its competition it's a bargain. This new European sized Toyota luxury sedan and wagon best exemplify Toyota's automotive state-of-the-art. A larger, roomier car built with Toyota quality.



A standard 4-speed overdrive automatic transmission, the only one of its kind. A Toyota engineering advancement not even found as an option on any other car. An automatic transmission that reduces engine RPMs by 31% in overdrive and delivers smooth, quiet, powerful performance. Attributes further enhanced by a 2.6 liter, 6-cylinder overhead cam engine. A combination unique to the Toyota Cressida.



A small price to pay for comfort. Cut pile carpeting cushions the feet. Your ears respond to AM/FM multiplex sounds. Your cool is maintained with air conditioning. Steering response is aided by power steering. And you're comforted in rich corduroy-type reclining bucket seats with adjustable driver's seat lumbar support. All standard.

Traditional Toyota values. Toyota economy of operation. Toyota dependable. The 1978 Cressida. "If you can find a better built small car than a Toyota…buy it." The new 1978 Cressida. Proof a well-built, luxury sedan can be acquired for only $7,059. Based on manufacturer's suggested retail price. Price does not include tax, license, transportation, California emissions or optional equipment.

for a number of reasons. It is the brightest member of the Virgo cluster of more than 130 galaxies and one of the brightest galaxies known. It also emits strongly in the radio and X-ray regions of the spectrum. A peculiar jet extends outward from the nucleus of the galaxy for 5,000 light-years. All this energetic activity suggests a close similarity between M87 and the more distant superluminous quasars. In addition, like many other giant galaxies, M87 seems to be deficient in mass: in order for a cluster of galaxies such as the Virgo one to be held together by gravitational forces, M87 and its companions should have from 10 to 30 times more mass than can readily be accounted for on the basis of their luminosity. One supposition is that the missing mass is in the form of an extended "halo" of stars too faint to show up in ordinary photographs.

A recent study of M87, conducted with ultrasensitive photon-counting techniques, has uncovered evidence of perhaps the most remarkable feature yet: a nonluminous concentration of mass in the nucleus of the galaxy equivalent to about five billion times the mass of the sun. Although the nuclear-mass concentration falls short of the putative missing mass by several orders of magnitude (which as the investigation proceeds may yet be discovered in the halo around the galaxy), the mass detected in the core of M87 could be a black hole enormously more massive than any of the half dozen or so black holes (probably having less than 10 solar masses each) postulated to exist as dark companions in double-star systems. The discovery of the large central mass in M87 will be reported this month in *The Astrophysical Journal* by a group of six astronomers: Wallace L. W. Sargent and Peter J. Young of the Hale Observatories and the California Institute of Technology; A. Boksenberg and Keith Shortridge of University College London, Clarence R. Lynds of Kitt Peak National Observatory and F. D. A. Hartwick of the University of Victoria in Australia.

The study of M87 was conducted in parallel with observations of a "normal" elliptical galaxy, NGC 3379, which served as an observational control. Although similar in outward appearance to M87, NGC 3379 is not a radio emitter and is otherwise unremarkable. M87 was observed with the four-meter (158-inch) Kitt Peak telescope, NGC 3379 with the 200-inch Hale telescope on Palomar Mountain. Both telescopes were equipped with new photon-counting systems developed at University College London that use sensitive photocathodes in place of photographic emulsions to record spectra of the object under study. The great advantage of the all-electronic system is that it can record about 20 percent of all the photons it intercepts. Photographic emulsions capture only about .5 percent. Thus the observing time is reduced by a factor of 40. Even so the observations of the two galaxies took 16 hours.

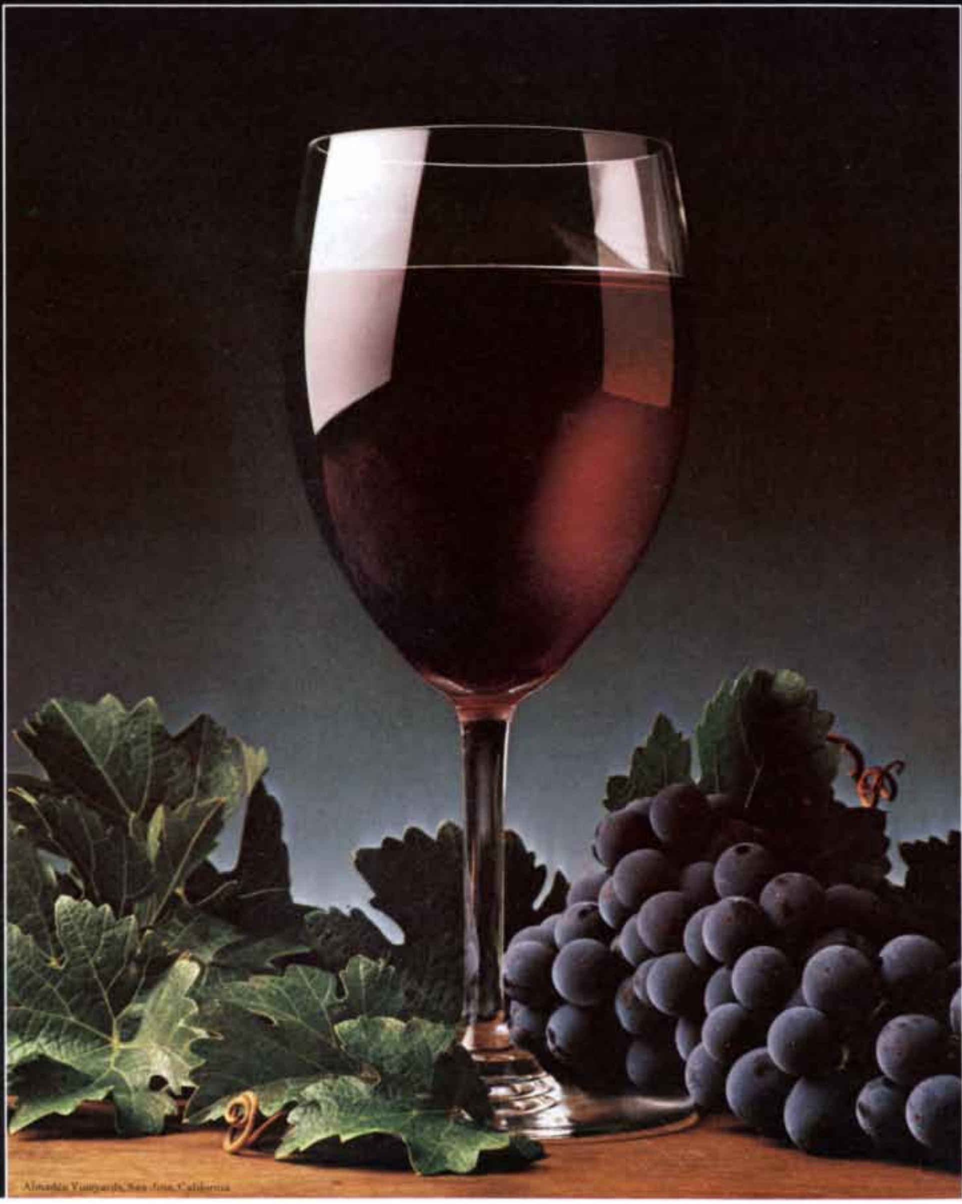
The design of the experiment called not only for a search for galactic halos but also for an examination of the nuclei of the two galaxies, which had heretofore been studied only with photographic techniques. The investigators were confident the photon-counting system could improve greatly on previous measurements of the velocities of the stars in the nuclei of the galaxies. The dispersion, or range, of the velocities also serves as a sensitive indicator of the amount of mass present in a given volume; the higher the average velocity, the greater the aggregate mass that is needed to keep the stars confined to that volume. The computed mass can then be compared with the light issuing from the same volume of space to obtain a ratio of mass to luminosity, both commonly expressed in units of the sun's mass and luminosity.

For the normal galaxy NGC 3379 the mass-to-light ratio in the nucleus was found to be about six to one. The comparable ratio for M87 is roughly 10 times higher. In other words, in the nucleus of M87 there appear to be about 60 solar units of invisible mass for every solar unit of luminosity. When this ratio is computed over a central volume whose radius is 110 parsecs (360 light-years), the quantity of nonluminous mass comes to five billion solar masses, of which no more than two percent can be in the form of normal stars of the type that populate the rest of M87. This works out to 25 solar masses per cubic light-year. Such a density is not exceptional for galactic cores, but the implication that most of the mass must be dark is unique. The authors conclude: "The presence of a central, supermassive black hole must be [considered] a serious possibility."

The hypothesis can be tested in a few years with the Large Space Telescope, which will have a resolution 10 times higher than ground-based instruments and therefore should be able to resolve the central 10 parsecs of M87's nucleus. If the dark mass is found to be confined within that limited radius (a volume 1,000 times smaller than that set by the present observations), the case for a galactic black hole will be compelling.

Cultured Masses

The difficulty of growing cells of a particular type in large quantities has long been a stumbling block in biological research. Many molecules of biological interest, such as peptide hormones, antigens and cell-surface receptors, are manufactured in minute amounts and must therefore be extracted from an extremely large number of the cells that make them. Such cells can-



When you look for a fine red wine, here's what to look for.

Experience. It's essential in making a fine red wine. Because most fine red wines need years of careful aging in small oak barrels. That demands constant attention. And the experience to know the right moment to bottle.

This is why you should look for a well-established winery first. One that has demonstrated its dedication by consistently producing fine wines year after year.

When selecting a fine red wine, you'll find many to choose from. Pinot Noir, Gamay Beaujolais and Zinfandel—to name a few. In each you'll find something unique. Some will be delicate. And some robust. Like Cabernet Sauvignon—perhaps the most complex fine red wine.

What should you look for in a Cabernet Sauvignon? A clear, deep ruby-red color. A bouquet that is distinctive and assertive. And a rich, complex taste that fills your mouth.

To fully enjoy this complex taste uncork the bottle and let it "breathe" at least one hour before serving. This will help the flavor fully develop.

And finally, remember a fine Cabernet Sauvignon always begins with fine Cabernet Sauvignon grapes. This is why at Almadén we care for our grapes as if they were our children. And you'll taste this tradition of quality in every Almadén wine.

To learn more about wine, send a postcard for a free copy of "The grapes and wines of Almadén" to: Almadén, P. O. Box 24249 SC, San Jose, CA 95154.

ALMADÉN

CABERNET SAUVIGNON

not usually be obtained in sufficient quantities from animals (particularly if the molecules of interest are found only in human cells), and must therefore be cultivated in the laboratory.

Although many virus-transformed and cancer cells grow unattached in suspension culture, normal tissue cells grow and divide only when they are anchored to a solid substrate. Hence they must be cultured in flat plastic dishes or roller bottles: a difficult and costly process when large volumes of cells are involved. Now, however, a research group at the Massachusetts Institute of Technology has perfected a new technique that involves growing large batches of substrate-dependent cells on tiny synthetic beads suspended in a nutrient medium. This approach promises to greatly reduce the tedium and expense of large-scale cell culture.

The idea of using synthetic beads as a substrate for cell growth was first proposed in 1967 by A. L. van Wezel of the Dutch National Institute for Public Health, who grew cells on commercially available beads designed for ion-exchange chromatography. These beads are made of cross-linked chains of dextran (a polymer of glucose) to which positively charged amine groups have been attached. Because the surface charge of animal cells is predominantly negative, they are attracted electrostatically to the beads and adhere to them. Although the beads are only about .15 millimeter in diameter, they provide a remarkably large surface area for cell growth: one gram of beads yields a surface area of approximately 6,100 square centimeters (about 6.5 square feet), equivalent to about 110 plastic dishes or 12 roller bottles.

In practice, however, Wezel found that the commercial beads eventually became toxic to the cells and limited their growth. This problem remained unsolved until David W. Levine, Daniel I. C. Wang and William G. Thilly at M.I.T.'s Cell Culture Center found that if they prepared dextran beads with fewer amine groups than the commercially available beads, the modification resulted in attachment without toxicity, making for rapid growth of substrate-dependent cells. The cells cultured to date have been primarily fibroblasts (connective tissue cells) from chick embryos and human tissues, although several other types have been cultivated with comparable success. Utilizing the modified beads suspended in a standard culture medium, the M.I.T. group has obtained concentrations of more than 7.6 million cells per milliliter. The cells are harvested by removing them from the beads with protein-cleaving enzymes, after which the beads settle to the bottom of the vessel, leaving the lighter cells suspended in the medium.

Levine and his colleagues believe that bead culture could be potentially scaled

up to industrial-size fermenters, replacing present-day cell-culture facilities that require vast numbers of small culture vessels. A cell-culture plant comprising 100,000 roller bottles, for example, might be replaced by four or five bead-suspension tanks with a total capacity of 250 gallons. Such a changeover to large batch cultivation would result in dramatic decreases in the cost of growing animal cells—perhaps as much as 80 to 90 percent.

The technological implications of cheap large-scale cell culture are considerable. Viral vaccines for human beings and animals that are now made either in eggs or in small-scale tissue culture could be prepared much more cheaply and easily with the bead-suspension system. Large-scale cell culture should also allow the economical production of peptide hormones, such as growth hormone, without the need to collect large quantities of animal glands or to risk the possible hazards of recombinant-DNA technologies. Finally, the production of interferon—the powerful antiviral agent manufactured by the body's own cells—could become economically feasible through the use of large-scale cell culture. According to Joel A. Huberman, a member of the Human Cell Biology section of the National Science Foundation, "At this point it is difficult to imagine all the possible implications of the easy culture of cell lines in large quantities."

Metallic Glass

When a solid is formed by the freezing of a liquid, one might imagine all the molecules simply stopping in their tracks as the temperature falls. The real process of freezing is somewhat different: the molecules come to rest only as they take up positions in an orderly, crystalline array. Nevertheless, a few liquids do commonly solidify in a more-or-less random configuration. That process is not properly called freezing, however, and the material that results is not an ordinary solid; it is a glass.

The familiar glasses are silicates, or compounds of silicon and oxygen with various impurity elements. Indeed, it was long thought that only silicates and compounds with similar chemical properties could form glasses. In the past several years, however, methods have been found for creating glassy, or amorphous, solids from a variety of other substances. Most notably, alloys of certain metals can now be solidified as glasses. The physics of the metallic glasses was reviewed recently in *Science* by Praveen Chaudhari of the IBM Research Center at Yorktown Heights, N.Y., and David Turnbull of Harvard University.

Unlike freezing, which takes place abruptly and at a particular temperature, the transition from a liquid to a glass is continuous. As the temperature falls the viscosity of the liquid increases smoothly, until at some temperature—defined arbitrarily as the glass temperature—the material flows so slowly that it must be considered a solid. In general, the glass temperature is far below the freezing point; it is for this reason that most materials do not readily form glasses. The liquid crystallizes long before the glass temperature is reached.

The principal technique for creating a metallic glass consists in the extremely rapid cooling of a molten alloy. In this way time spent in the unstable region below the freezing point but above the glass temperature is reduced to a minimum; essentially the atoms of the metal do not have time to arrange themselves in a crystalline array before they are immobilized by a low temperature. One such method is called "splat cooling." Droplets of molten metal are flung onto a plate held at low temperature. The droplets spread out into a thin film that cools to a glassy solid in a few hundred microseconds; the rate of cooling can reach a million degrees Celsius per second. A similar procedure developed recently produces continuous ribbons of metallic glass by directing a thin stream of liquid onto a cold, rotating cylinder.

So far only alloys, such as compounds of iron and boron, have been induced to solidify in the glassy state, and it seems unlikely that elemental glassy metals can be made. Alloys have a slightly higher glass temperature than pure metals and some alloys, called eutectics, have a much lower freezing point, so that the interval between the two temperatures is greatly reduced.

Glasses are said to be amorphous solids, but they almost certainly have some degree of order, at least over a range of a few atomic or molecular diameters. Determining the topology of this short-range order, Chaudhari and Turnbull write, is the central problem in the study of metallic glasses. Two main classes of models have been formulated. In one the solid is represented as the densest possible random packing of spheres, much as some crystals represent the densest orderly packing of spheres. In the other class of models the structure is built up from randomly oriented microcrystallites, each about five atoms across. Certain predictions of the models can be tested through X-ray-diffraction studies of the glasses. In general, the random-packing models seem to be more accurate than the microcrystallite models, and good results have also been obtained with models that fall somewhere between these extremes. In these "amorphous-cluster" models the basic structural element is an assemblage of atoms that is not crystalline but that nonetheless has close to minimum potential energy.

Many of the properties of a crystalline metal are determined by its periodic long-range structure, and so it comes as no surprise to learn that amorphous metals differ significantly in some properties. For example, the tensile strength of crystalline metals and of silicate glasses is greatly diminished by even minute flaws, and so the theoretical maximum strength is approached only in carefully prepared small samples. Glassy metals can exhibit very high tensile strength even in bulk materials, and the strength is comparatively insensitive to small imperfections. This combination of properties gives the amorphous metals a toughness that may lead to applications as structural materials.

The electrical and magnetic properties of metallic glasses are also distinctive. Electrical conductivity is lower than it is in the crystalline state, but it is also less sensitive to temperature, a fact that might be exploited in the manufacture of thin-film electronic devices. Magnetic properties depend strongly on composition; of particular interest are amorphous solids that are magnetically "soft," that is, metals that are readily magnetized and demagnetized. One possible application of such materials is in the storage of information as magnetic "bubbles," or small controlled domains of reversed magnetic polarity. In metallic glasses the bubbles can be made smaller than they can in the crystalline materials now employed, and as a result it may be possible to store as many as a billion bits of information per square centimeter. The magnetically soft glassy metals might also be useful in transformers and similar devices.

Finally, in several glassy alloys superconductivity has been observed. Certain properties of the glassy superconductors, such as the transition temperature where resistanceless flow begins, are significantly different from those in crystalline materials, and the differences have not been explained. Nevertheless, practical applications are at least conceivable. Quantized eddy currents, which can be induced in all superconductors, are mobile over long distances in amorphous ones. The eddy currents might be employed for information storage in much the same way as magnetic bubbles, but with a potential information density 100 times greater.

Fifth Wheel

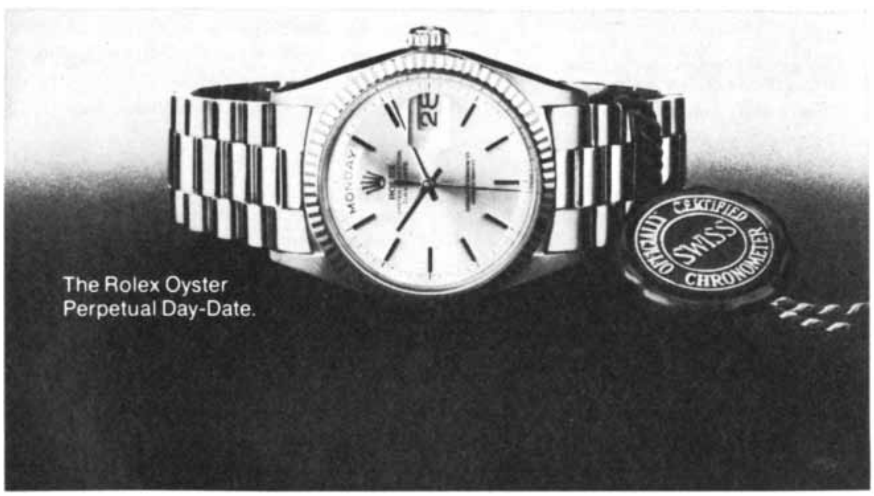
Government demands for greater automobile fuel economy have induced Detroit to pare the weight of its vehicles in a variety of ways, for example by substituting plastic for metal and aluminum for iron. One prime target has been the automobile tire, because improvements in tire design offer potential fuel savings of two kinds. Design changes leading to a reduction in rolling resistance, one approach under active development, can result in fuel savings on the order of 10 percent. Changes that allow a punctured tire to run flat without danger or that prevent a punctured tire from going flat even when it contains no air offer a windfall to the automobile makers: the elimination of the spare tire. Eliminated with the tire are of course the spare wheel and the jack. These changes effect an economy in manufacturing and can subtract as much as 60 pounds from the weight of the car.

The move toward getting rid of the spare tire has already led to half-measures such as providing miniaturized spares. Some come collapsed, along with a can of pressurized inflatant; others are preinflated but are smaller than the tire they replace. All have met with customer apathy or hostility. With a number of run-flat tires and at least one no-flat model now being tested, however, the time for half-measures may be coming to a close. Most private car owners will continue to demand the psychological comfort of a fifth wheel for a while, but fleet operators such as car-rental companies and government agencies will welcome the elimination of the spare.

Announcing The Winners of The Rolex Awards for Enterprise

In September 1976, to celebrate the 50th anniversary of the Rolex Oyster, we launched the Rolex Awards for Enterprise.

"To provide financial help for projects which seek to break new ground in their particular sphere and which capture the spirit of enterprise shown by Rolex, and Rolex owners, over the last 50 years."



The Rolex Oyster Perpetual Day-Date.

We offered a total of five awards of 50,000 Swiss francs each, together with a specially inscribed Rolex Chronometer.

The response from around the world was overwhelming and over 3,000 projects were finally considered.

The Winners

This spring, Dr. Billy Lee Lasley, Kenneth Lee Marten, and Francine Grace Penelope Patterson from the U.S.A., Luc Jean-François Debecker of Belgium, and Georges Marcel André Delamare of France were invited to Geneva, where they were each presented with a check for 50,000 Swiss francs, together with a specially inscribed gold Rolex Chronometer.

The Book

"The Rolex Awards for Enterprise," by Gregory B. Stone, will be published later in the year as a lasting tribute to the Rolex Spirit of Enterprise. The book will give a full account of the winning projects along with the details of hundreds of other entries. In the meantime, if you'd like more information on the winning entries, write to Enterprise, Rolex Watch U.S.A., Inc., Rolex Building, 665 Fifth Avenue, New York, New York 10022.

ROLEX

The Food-sharing Behavior of Protohuman Hominids

Excavations at two-million-year-old sites in East Africa offer new insights into human evolutionary progress by showing that early erect-standing hominids made tools and carried food to a home base

by Glynn Isaac

Over the past decade investigators of fossil man have discovered the remains of many ancient protohumans in East Africa. Findings at Olduvai, Laetolil, Koobi Fora, the Omo Valley and Hadar, to name some prominent locations, make it clear that between two and three million years ago a number of two-legged hominids, essentially human in form, inhabited this part of Africa. The paleontologists who have unearthed the fossils report that they differ from modern mankind primarily in being small, in having relatively large jaws and teeth and in having brains that, although they are larger than those of apes of comparable body size, are rarely more than half the size of modern man's.

The African discoveries have many implications for the student of human evolution. For example, one wonders to what extent the advanced hominids of two million years ago were "human" in their behavior. Which of modern man's special capabilities did they share? What pressures of natural selection, in the time since they lived, led to the evolutionary elaboration of man's mind and culture? These are questions that paleontologists find difficult to answer because the evidence that bears on them is not anatomical. Archaeologists, by virtue of their experience in studying prehistoric behavior patterns in general, can help to supply the answers.

It has long been realized that the human species is set apart from its closest living primate relatives far more by differences in behavior than by differences in anatomy. Paradoxically, however, the study of human evolution has traditionally been dominated by work on the skeletal and comparative anatomy of fossil primates. Several new research movements in recent years, however, have begun to broaden the scope of direct evolutionary inquiry. One such movement involves investigations of the behavior and ecology of living primates and of other mammals. The results of these observations can now be compared with quantitative data from another new area of study, namely the cultural ecology of human societies that support themselves without raising plants or animals: the few surviving hunter-gatherers of today. Another important new movement has involved the direct study of the ecological circumstances surrounding human evolutionary developments. Investigations of this kind have become possible because the stratified sedimentary rocks of East Africa preserve, in addition to fossil hominid remains, an invaluable store of data: a coherent, ordered record of the environments inhabited by these protohumans.

The work of the archaeologist in drawing inferences from such data is made possible by the fact that at a certain stage in evolution the ancestors of modern man became makers and users of equipment. Among other things, they shaped, used and discarded numerous stone tools. These virtually indestructible artifacts form a kind of fossil record of aspects of behavior, a record that is complementary to the anatomical record provided by the fossil bones of the toolmakers themselves. Students of the Old Stone Age once concentrated almost exclusively on what could be learned from the form of such tools. Today the emphasis in archaeology is increasingly on the context of the artifacts: for example the distribution pattern of the discarded tools in different settings and the association of tools with various kinds of food refuse. A study of the contexts of the early African artifacts yields unique clues both to the ecological circumstances of the protohuman toolmakers and to aspects of their socioeconomic organization.

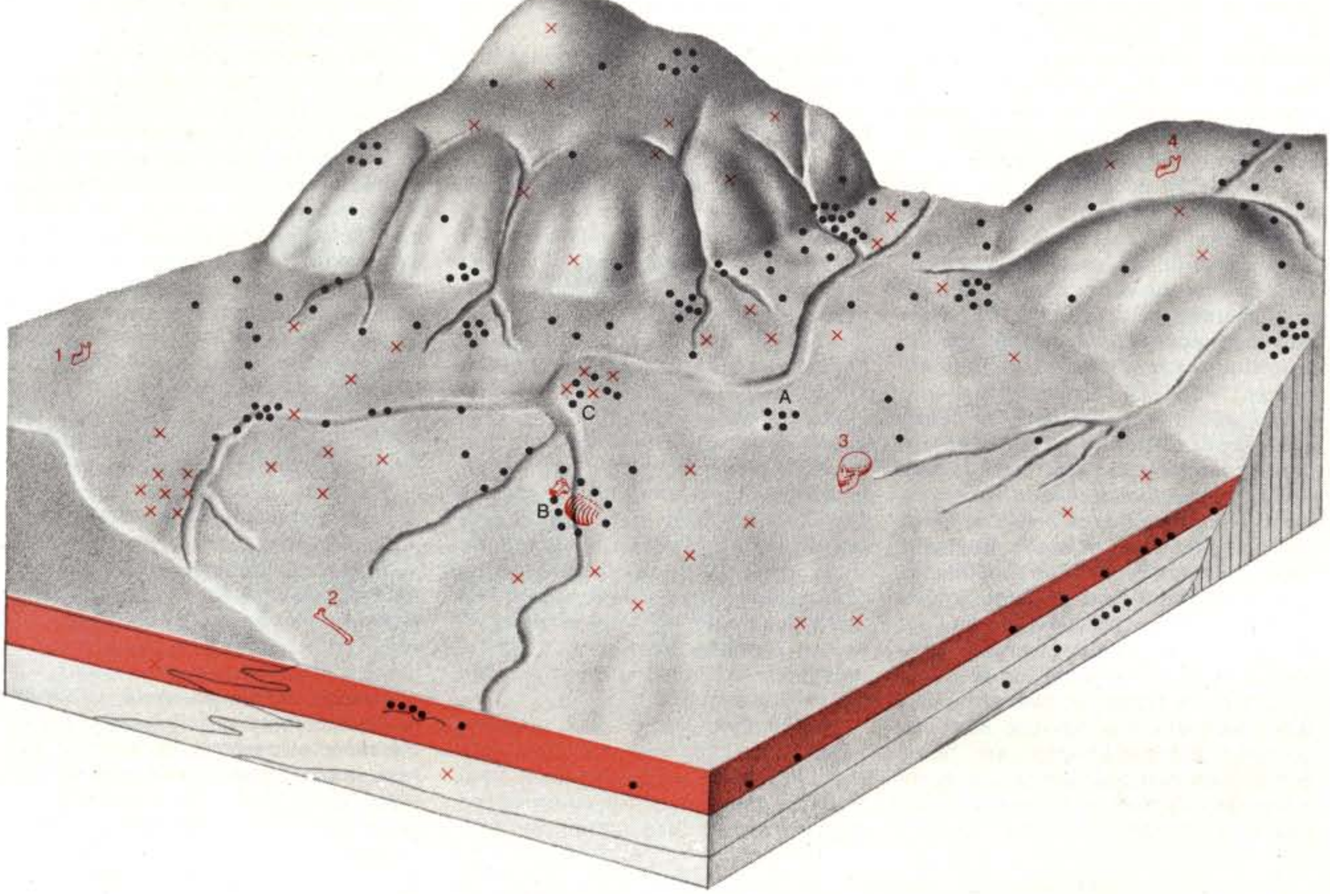
Comparing Men and Apes

What are the patterns of behavior that set the species *Homo sapiens* apart from its closest living primate relatives? It is not hard to draw up a list of such differences by comparing human and ape behavior and focusing attention not on the many features the two have in common but on the contrasting features. In the list that follows I have drawn on recent field studies of the great apes (particularly the chimpanzee, *Pan troglodytes*) and on similar studies of the organization of living hunter-gatherer societies. The list tends to emphasize the contrasts relating to the primary subsistence adaptation, that is, the quest for food.

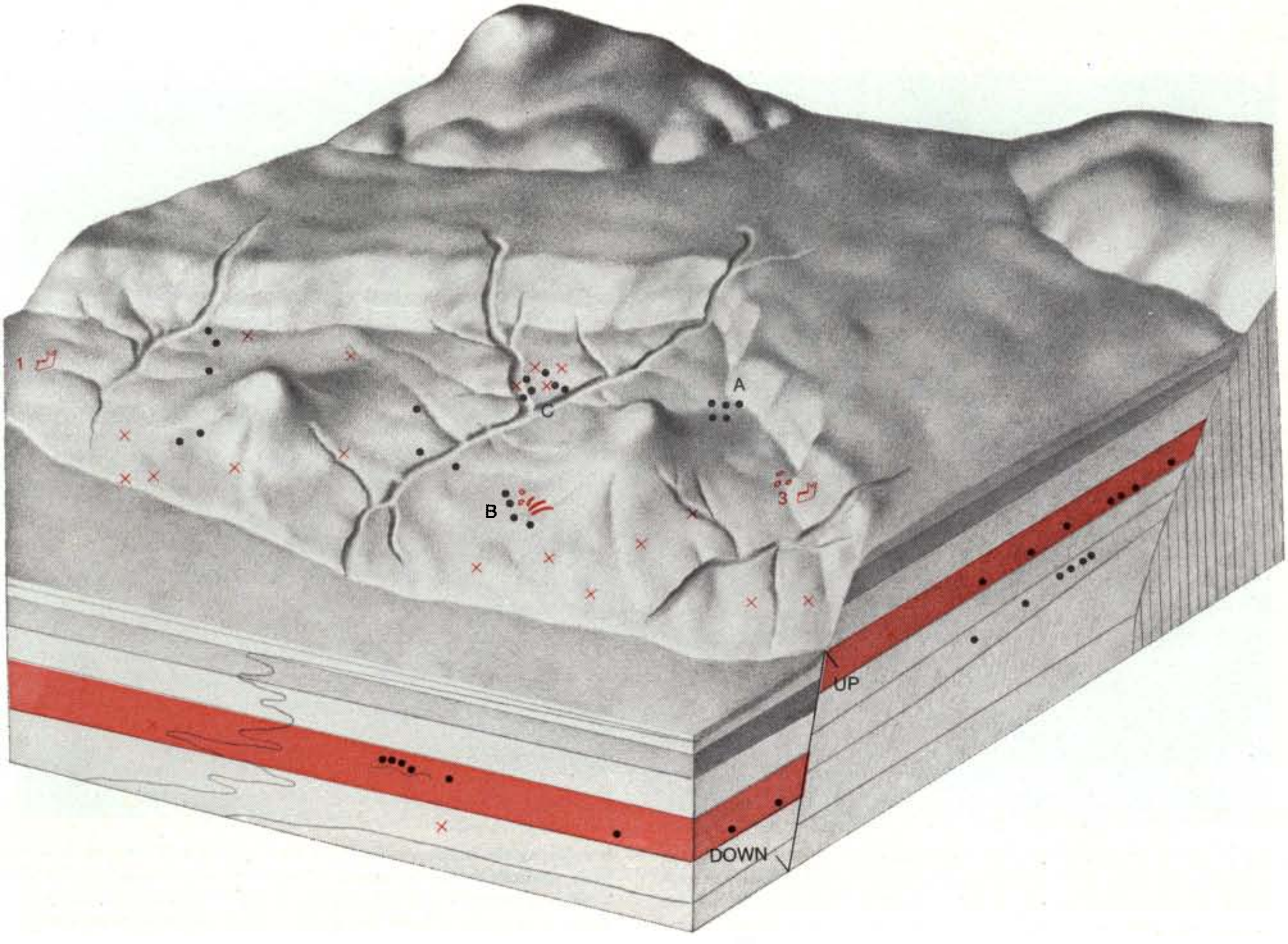
First, *Homo sapiens* is a two-legged primate who in moving from place to place habitually carries tools, food and other possessions either with his arms or in containers. This is not true of the great apes with regard to either posture or possessions.

Second, members of *Homo sapiens* so-

PAST AND PRESENT LANDSCAPES in the Rift Valley region of East Africa, shown schematically on the opposite page, summarize the geological activity that first preserved and later exposed evidence of protohuman life. Two million years ago (*top*) the bones of hominids (*1–4, color*) and other animals (*x's, color*) were distributed across hills and a floodplain (*foreground*) adjacent to a Rift Valley lake. Also lying on the surface were stone tools (*black dots*) made, used and discarded by the protohumans. Layers of sediments then covered the bones and tools lying on the floodplain; burial preserved them, whereas the bones and tools in the hills were eventually washed away. Today (*bottom*), after a fault has raised a block of sediments, erosion is exposing some of the long-buried bones and clusters of tools, including the three types of site shown on the surface in the top block diagram (*A–C*). Sites of Type A contain clusters of stone tools together with the leftover stone cores that provided the raw material for the tools and waste flakes from the toolmaking process, but little or no bone is present. Sites of Type B contain similar clusters of tools in association with the bones of a single large animal. Sites of Type C also contain similar clusters of tools, but the bones are from many different animal species.



1
2
3
4
A
B
C



1
3
A
B
C
UP
DOWN

cieties communicate by means of spoken language; such verbal communication serves for the exchange of information about the past and the future and also for the regulation of many aspects of social relations. Apes communicate but they do not have language.

Third, in *Homo sapiens* societies the acquisition of food is a corporate responsibility, at least in part. Among members of human social groupings of various sizes the active sharing of food is a characteristic form of behavior; most commonly family groups are the crucial nodes in a network of food exchange. Food is exchanged between adults, and it is shared between adults and juveniles. The only similar behavior observed among the great apes is seen when chimpanzees occasionally feed on meat. The chimpanzees' behavior, however, falls far short of active sharing; I suggest it might better be termed tolerated scrounging. Vegetable foods, which are the great apes' principal diet, are not shared and are almost invariably consumed by each individual on the spot.

Fourth, in human social groupings there exists at any given time what can be called a focus in space, or "home base," such that individuals can move independently over the surrounding terrain and yet join up again. No such home base is evident in the social arrangements of the great apes.

Fifth, human hunter-gatherers tend to devote more time than other living primates to the acquisition of high-protein foodstuffs by hunting or fishing for animal prey. It should be noted that the distinction is one not of kind but of degree. Mounting evidence of predatory behavior among great apes and monkeys suggests that the principal contrast between human beings and other living primates with respect to predation is that only human beings habitually feed on prey weighing more than about 15 kilograms.

The gathering activities of human hunter-gatherers include the collection of edible plants and small items of animal food (for example lizards, turtles, frogs, nestling birds and eggs). Characteristically a proportion of these foodstuffs is not consumed until the return to the home base. This behavior is in marked contrast to what is observed among foraging great apes, which almost invariably feed at the spot where the food is acquired.

Still another contrast with great-ape feeding behavior is human hunter-gatherers' practice of subjecting many foodstuffs to preparation for consumption, by crushing, grinding, cutting and heating. Such practices are not observed among the great apes.

Human hunter-gatherers also make use of various kinds of equipment in the quest for food. The human society with perhaps the simplest equipment ever observed was the aboriginal society of Tasmania, a population of hunter-gatherers that was exterminated in the 19th century. The inventory of the Tasmanians' equipment included wood clubs, spears and digging sticks, cutting tools made of chipped stone that were used to shape the wood objects, and a variety of containers: trays, baskets and bags. The Tasmanians also had fire. Although such equipment is simple by our standards, it is far more complex than the kind of rudimentary tools that we now know living chimpanzees may collect and use in the wild, for example twigs and grass stems.

In addition to this lengthy list of subsistence-related behavioral contrasts between human hunter-gatherers and living primates there is an entire realm of other contrasts with respect to social organization. Although these important additional features fall largely outside the range of evidence to be considered here, they are vital in defining human patterns of behavior. Among them is the propensity for the formation of long-

DESOLATE LANDSCAPE in the arid Koobi Fora district of Kenya is typical of the kind of eroded terrain where gullying exposes both bones and stone tools that were buried beneath sediments and volcanic ash more than a million years ago. Excavation in progress (*center*) is exposing the hippopotamus bones and clusters of artifacts that had been partially bared by recent erosion and were found by Richard Leakey in 1969. The site is typical of the kind that includes the remains of a single animal and many tools manufactured on the spot.

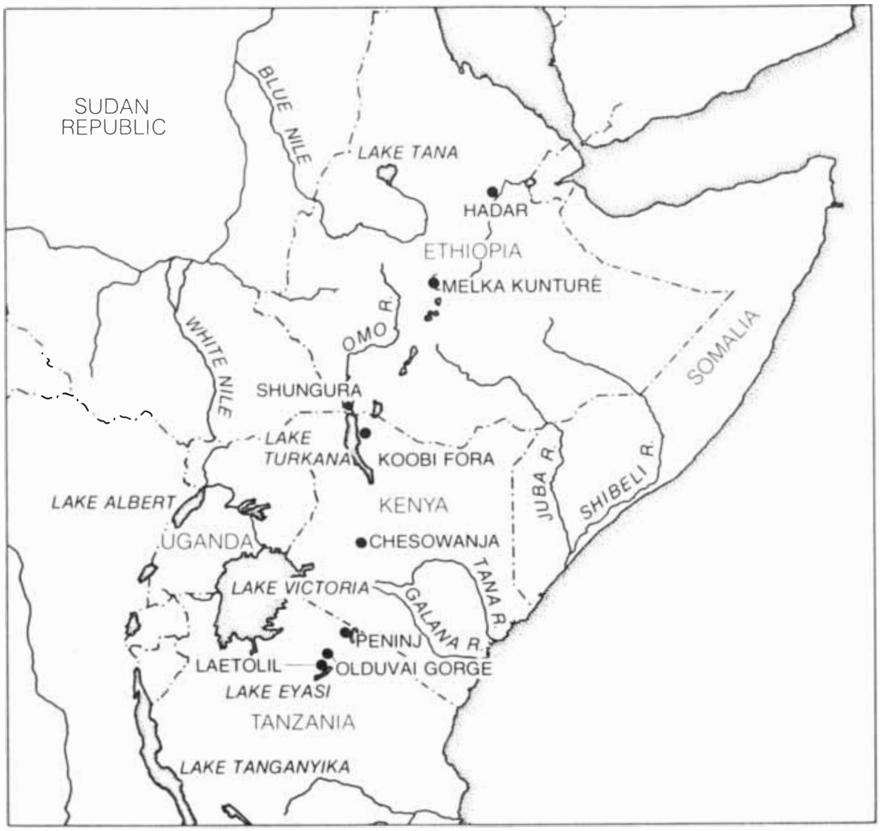
term mating bonds between a male and one or more females. The bonds we call "marriage" involve reciprocal economic ties, joint responsibility for aspects of child-rearing and restrictions on sexual access. Another such social contrast is evident in the distinctively human propensity to categorize fellow members of a group according to kinship and metaphors of kinship. Human beings regulate many social relations, mating included, according to complex rules involving kinship categories. Perhaps family ties of a kind exist among apes, but explicit categories and rules do not. These differences are emphasized by the virtual absence from observed ape behavior of those distinctively human activities that are categorized somewhat vaguely as "symbolic" and "ritual."

Listing the contrasts between human and nonhuman subsistence strategies is inevitably an exercise in oversimplification. As has been shown by contemporary field studies of various great apes and of human beings who, like the San (formerly miscalled Bushmen) of the Kalahari Desert, still support themselves without farming, there is a far greater degree of similarity between the two subsistence strategies than had previously been recognized. For example, with regard to the behavioral repertories involving meat-eating and tool-using the differences between ape and man are differences of degree rather than of kind. Some scholars have even used the data to deny the existence of any fundamental differences between the human strategies and the nonhuman ones.

It is my view that significant differences remain. Let me cite what seem to me to be the two most important. First, whereas humans may feed as they forage just as apes do, apes do not regularly postpone food-consumption until they have returned to a home base, as human beings do. Second, human beings actively share some of the food they acquire. Apes do not, even though chimpanzees of the Gombe National Park in Tanzania have been observed to tolerate scrounging when meat is available.

From Hominid to Human

Two complementary puzzles face anyone who undertakes to examine the question of human origins. The first relates to evolutionary divergence. When did the primate stock ancestral to the living apes diverge from the stock ancestral to man? What were the circumstances of the divergence? Over what geographical range did it take place? It is not yet established beyond doubt whether the divergence occurred a mere five to six million years ago, as Vincent M. Sarich of the University of California at Berkeley and others argue on biochemical grounds, or 15 to 20 million years ago, as many paleontologists believe on the grounds of fossil evidence.



PROMINENT SITES in East Africa include (from north to south) Hadar, Melka Kunturé and Shungura in Ethiopia, the Koobi Fora district to the east of Lake Turkana in Kenya, Chesowanja in Kenya and Peninj, Olduvai Gorge and Laetolil in Tanzania. Dates for clusters of stone tools, some associated with animal bones, uncovered at these sites range from one million years ago (Olduvai Upper Bed II) to 2.5 million (Hadar upper beds). Some sites may be even older.

YEARS BEFORE PRESENT (MILLIONS)	OLDUVAI	KOOBI FORA	OMO VALLEY	OTHER
1.0				
	UPPER BED II			
1.2				PENINJ
	MIDDLE BED II			MELKA KUNTURÉ
1.4		KARARI SITES		CHESOWANJA
	LOWER BED II			
1.6				
	BED I	KBS, HAS		
1.8				
2.0				
			SHUNGURA MEMBER F	
2.2				
			SHUNGURA MEMBER E	
2.4				
		?		HADAR UPPER BEDS
2.6				
2.8				
3.0				
3.2				HADAR LOWER BEDS
3.4				
				LAETOLIL
3.6				

RELATIVE ANTIQUITY of selected sites in East Africa is indicated in this table. Olduvai Gorge beds I and II range from 1.8 to 1.0 million years in age. The Shungura sites in the Omo Valley are more than two million years old. Two Koobi Fora locales, the hippopotamus/artifact site (HAS) and the Kay Behrensmeyer site (KBS), are at least 1.6 million years old. Initial geological studies of the Koobi Fora sites suggested that they might be 2.5 million years old (*colored line*). Only hominid fossils have been found in the lower beds at Hadar and at Laetolil.

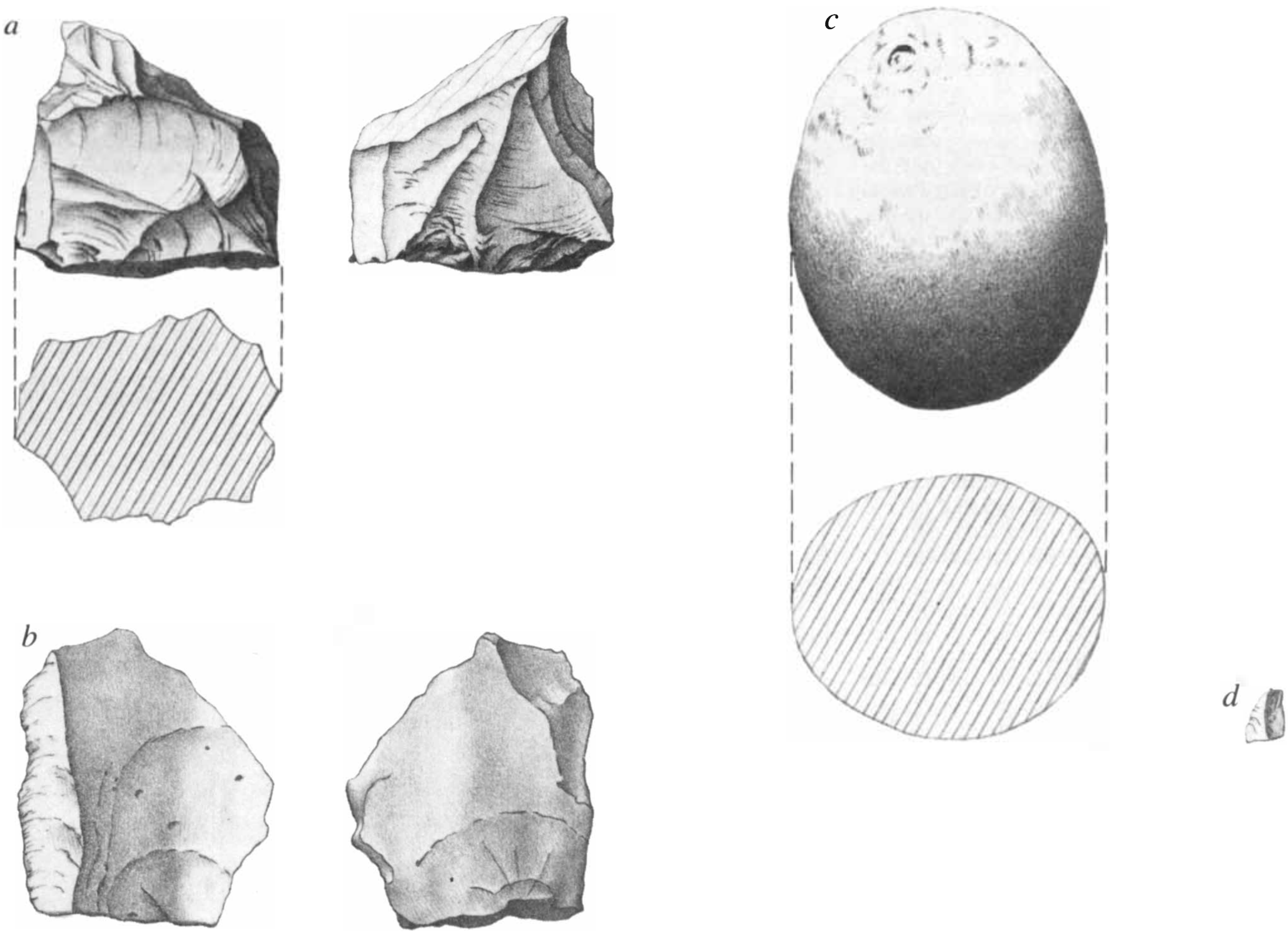
At least one fact is clear. The divergence took place long before the period when the oldest archaeological remains thus far discovered first appear. Archaeology, at least for the present, can make no contribution toward solving the puzzle of the split between ancestral ape and ancestral man.

As for the second puzzle, fossil evidence from East Africa shows that the divergence, regardless of when it took place, had given rise two to three million years ago to populations of smallish two-legged hominids. The puzzle is how to identify the patterns of natural selection that transformed these protohumans into humans. Archaeology has a major contribution to make in elucidating the second puzzle. Excavation of these protohuman sites has revealed evidence suggesting that two million years ago some elements that now distinguish man from apes were already part of a novel adaptive strategy. The indications are that a particularly important part of that strategy was food-sharing.

The archaeological research that has inspired the formulation of new hypotheses concerning human evolution began nearly 20 years ago when Mary Leakey and her husband Louis discovered the fossil skull he named "Zinjanthropus" at Olduvai Gorge in Tanzania. The excavations the Leakeys undertook at the site showed not only that stone tools were present in the same strata that held this fossil and other hominid fossils but also that the discarded artifacts were associated with numerous broken-up animal bones. The Leakeys termed these concentrations of tools and bones "living sites." The work has continued at Olduvai under Mary Leakey's direction, and in 1971 a major monograph was published that has made the Olduvai results available for comparative studies.

Other important opportunities for archaeological research of this kind have come to light in the Gregory Rift Valley, at places such as the Koobi Fora (formerly East Rudolf) region of northern Kenya, at Shungara in the Omo Valley of southwestern Ethiopia and in the Hadar region of eastern Ethiopia. Current estimates of the age of these sites cover a span of time from about 3.2 million years ago to about 1.2 million.

Since 1970 I have been co-leader with Richard Leakey (the son of Mary and Louis Leakey) of a team working at Koobi Fora, a district that includes the northeastern shore of Lake Turkana (the former Lake Rudolf). Our research on the geology, paleontology and paleoanthropology of the district involves the collaboration of colleagues from the National Museum of Kenya and from many other parts of the world. Work began in 1968 and has had the help and encouragement of the Government of Kenya, the National Science Foundation and the National Geographic Society. Our investigations have yielded archaeological evidence that corroborates and complements the earlier evidence from Olduvai Gorge. The combined data make it possible to see just how helpful archaeology can be in answering



KOOBI FORA ARTIFACTS include four from the HAS assemblage (*left*) and four from the KBS assemblage (*right*). All are shown actual size; the stone is basalt. The HAS core (*a*) shows what is left of a piece of stone after a number of flakes have been struck from it by percussion. The jagged edges produced by flake removal give the core potential usefulness as a tool. The flakes were detached from the core by blows with a hammerstone like the one shown here (*c*). The sharp edges of the flakes, such as the example illustrated (*b*), allow their use as cutting tools. The tiny flake (*d*) is probably an accidental product of the percussion process; the presence of many stone splinters such

questions concerning human evolution.

At Koobi Fora, as at all the other East African sites, deposits of layered sediments, which accumulated long ago in the basins of Rift Valley lakes, are now being eroded by desert rainstorms and transient streams. As the sedimentary beds erode, a sample of the ancient artifacts and fossil bones they contain is exposed at the surface. For a while the exposed material lies on the ground. Eventually, however, the fossil bones are destroyed by weathering or a storm washes away stone and bone alike.

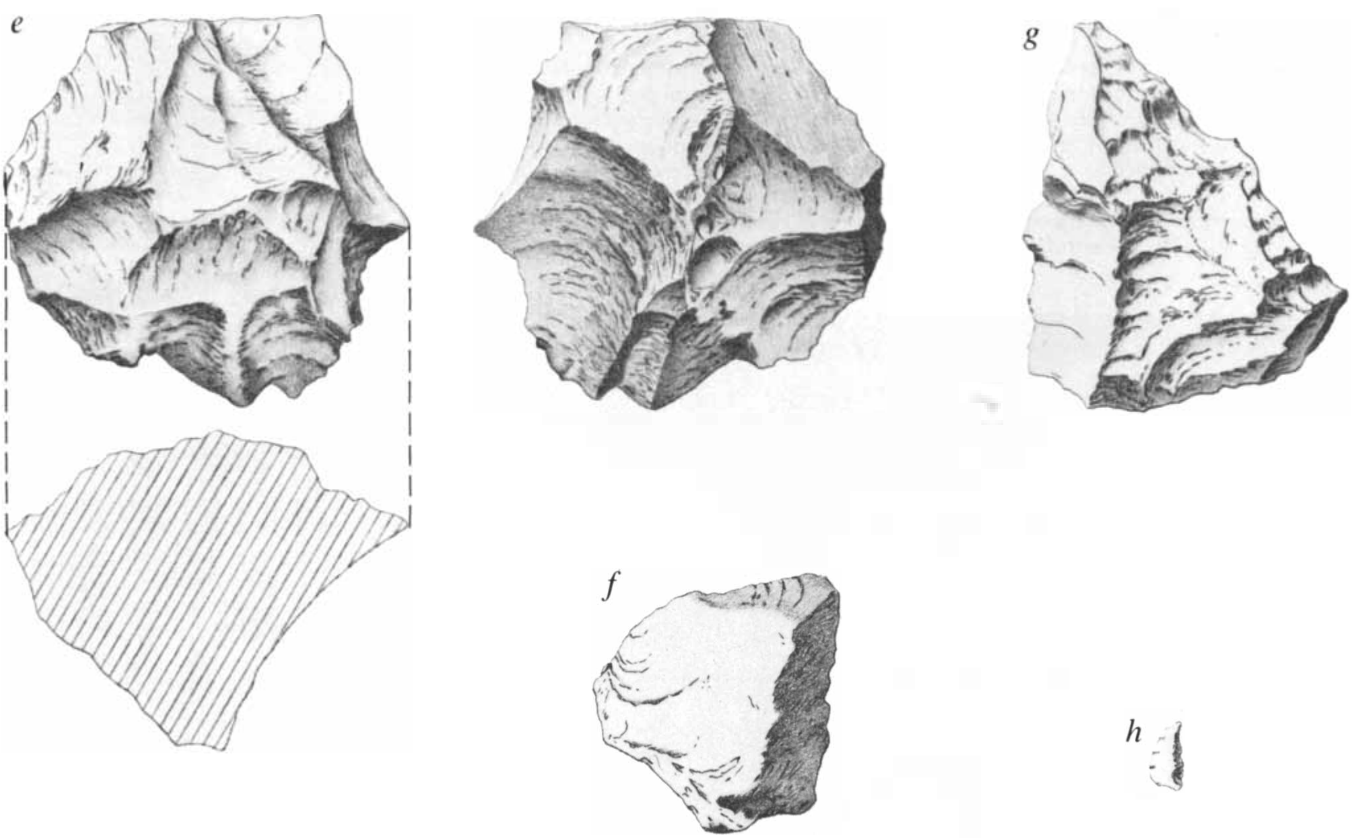
All field reconnaissance in East Africa progresses along essentially similar lines. The field teams search through eroded terrain looking for exposed fossils and artifacts. In places where concentrations of fossil bone or promising archaeological indications appear on the surface the next step is excavation. The digging is done in part to uncover further specimens that are still in place in the layers of sediments and in part to gather exact information about the original stratigraphic location of the surface material. Most important of all, excavation allows the investigators to plot in detail the relative locations of the material that is unearthed. For example, if there are associations among bones and between bones and stones, excavation will reveal these characteristics of the site.

The Types of Sites

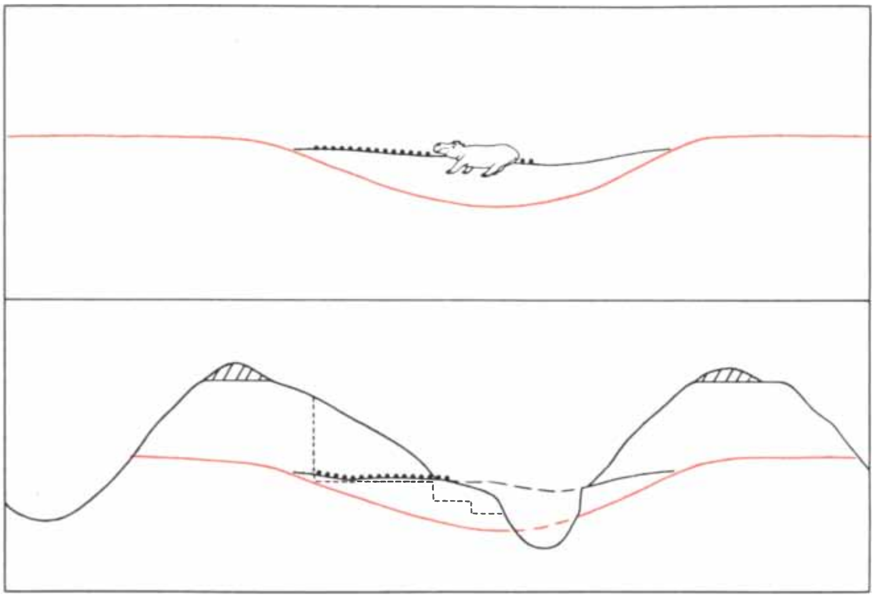
The archaeological traces of protohuman life uncovered in this way may exhibit several different configurations. In some ancient layers we have found scatterings of sharp-edged broken stones even though there are no other stones in the sediments. The broken stones come in a range of forms but all are of the kind produced by deliberate percussion, so that we can classify them as undoubted artifacts. Such scatterings of artifacts are often found without bone being present in significant amounts. These I propose to designate sites of Type A.

In some instances a layer of sediment may include both artifacts and animal bones. Such bone-and-artifact occurrences fall into two categories. The first consists of artifacts associated with bones that represent the carcass of a single large animal; these sites are designated Type B. The second consists of artifacts associated with bones representing the remains of several different animal species; these sites are designated Type C.

The discovery of sites with these varied configurations in the sediments at Koobi Fora and Olduvai provides evidence that when the sediments containing them were being deposited some 2.5 to 1.5 million years ago, there was at least one kind of hominid in East Africa that habitually carried objects such as stones from one place to another and made sharp-edged tools by deliberately fracturing the stones it carried with it. How does this archaeological evidence match up with the hominid fossil record? The fossil evidence indicates that



as this one in the HAS tool clusters indicates that the stone tools were made on the spot. At the same time the absence of local unworked stone as potential raw material for tools suggests that the cores were carried to the site by the toolmakers. The artifacts from the second assemblage also include a core (*e*) that has had many flakes removed by percussion and another small splinter of stone (*h*). The edges of the two flakes (*f, g*) are sharp enough to cut meat, hide, sinew or wood. As at the hippopotamus/artifact site, the absence of local raw material for stone tools at the Kay Behrensmeyer site suggests that suitable lumps of lava must have been transported there by the toolmakers.



KOOBI FORA LANDSCAPE in the vicinity of the hippopotamus/artifact site consisted of a level floodplain near the margin of a lake (*top section*). Protohuman foragers apparently found the carcass of a hippopotamus lying in a stream-bed hollow and made tools on the spot in order to butcher the carcass. Their actions left a scatter of stone tools among the bones and on the ground nearby. The floodplain was buried under layers of silt and ash and was subsequently eroded (*bottom section*), exposing some bones and tools. Their discovery led to excavation.

two and perhaps three species of bipedal hominids inhabited the area at this time, so that the question arises: Can the species responsible for the archaeological evidence be identified?

For the moment the best working hypothesis seems to be that those hominids that were directly ancestral to modern man were making the stone tools. These are the fossil forms, of early Pleistocene age, classified by most paleontologists as an early species of the genus *Homo*. The question of whether or not contemporaneous hominid species of the genus *Australopithecus* also made tools must be set aside as a challenge to the ingenuity of future investigators. Here I shall simply discuss what we can discover about the activities of early toolmaking hominids without attempting to identify their taxonomic position (or positions).

HAMMERSTONE unearthed at the hippopotamus/artifact site is a six-centimeter basalt pebble; it is shown here being lifted from its position on the ancient ground surface adjacent to the hippopotamus bones. Worn smooth by water action before it caught the eye of a toolmaker some 1.7 million years ago, the pebble is battered at both ends as a result of use as a hammer.

Reading the Evidence

As examples of the archaeological evidence indicative of early hominid patterns of subsistence and behavior, consider our findings at two Koobi Fora excavations. The first is a locality catalogued as the hippopotamus/artifact site (HAS) because of the presence of fossilized hippopotamus bones and stone tools.

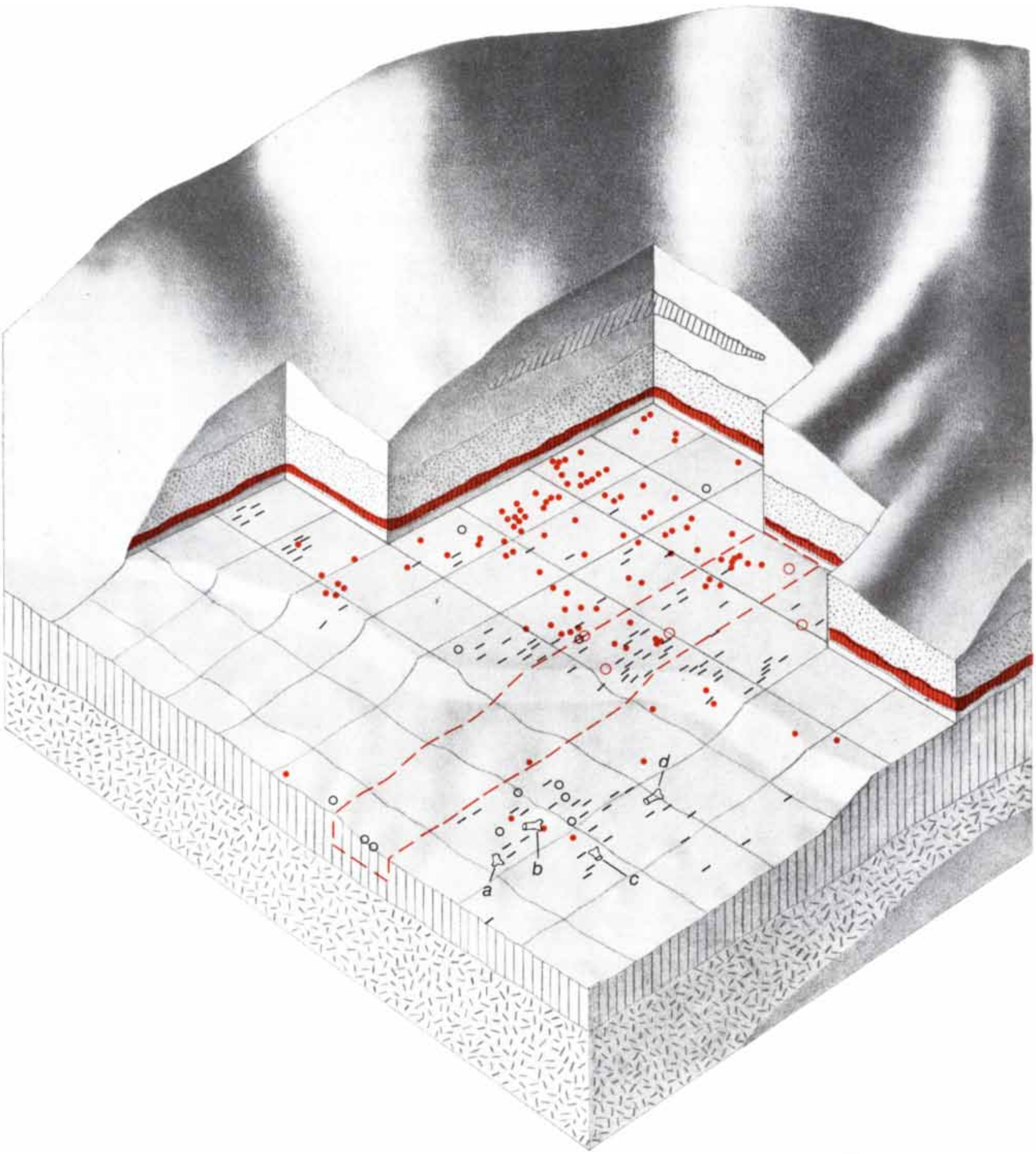
The site is 15 miles east of Lake Turkana. There in 1969 Richard Leakey discovered an erosion gully cutting into an ancient layer of volcanic ash known as the KBS tuff. (KBS stands for Kay Behrensmeyer site; she, the geologist-paleoecologist of our Koobi Fora research team, first identified the ash layer at a nearby outcrop.) The ash layer is the uppermost part of a sedimentary deposit known to geologists as the Lower Member of the Koobi Fora Formation; here the ash had filled in one of the many dry channels of an ancient delta. Leakey found many bones of a single hippopotamus carcass weathering out of the eroded ash surface, and stone artifacts lay among the bones.

J. W. K. Harris, J. Onyango-Abuje and I supervised an excavation that cut into an outcrop where the adjacent delta sediments had not yet been disturbed by erosion. Our digging revealed that the hippopotamus carcass had originally lain in a depression or puddle within an ancient delta channel. Among the hippopotamus bones and in the adjacent stream bank we recovered 119 chipped stones; most of them were small sharp flakes that, when they are held between the thumb and the fingers, make effective cutting implements. We also recovered chunks of stone with scars showing that flakes had been struck from them by percussion. In Paleolithic tool classification these larger stones fall into the category of core tool or chopper. In addition our digging exposed a rounded river pebble that was battered at both ends; evidently it had been used as a hammer to strike flakes from the stone cores.

The sediments where we found these artifacts contain no stones larger than a pea. Thus it seems clear that the makers of the tools had carried the stones here from somewhere else. The association between the patch of artifacts and the hippopotamus bones further suggests that toolmakers came to the site carrying stones and hammered off the small sharp-edged flakes on the spot in order to cut meat from the hippopotamus carcass. We have no way of telling at present whether the toolmakers themselves killed the animal or only came on it dead. Given the low level of stone technology in evidence, I am inclined to suspect scavenging rather than hunting.

The HAS deposit was formed at least 1.6 million years ago. The archaeological evidence demonstrates that the behavior of some hominids at that time differed from the behavior of modern great apes in that these protohumans not only made cutting tools but also ate meat from the carcasses of large animals. The hippopotamus/artifact site thus provides corroboration for evidence of similar behavior just as long ago obtained from Mary Leakey's excavations at Olduvai Gorge.

This finding does not answer all our questions. Were these protohumans roaming the landscape, foraging and hunting, in the way that a troop of baboons does today? Were they instead hunting like a pride of lions? Or did some other behavioral pattern prevail? Excavation of another bone-and-arti-



FINDINGS at the hippopotamus/artifact site are shown schematically in this block diagram; squares are one meter to a side. In the foreground are the objects that had been exposed by weathering: hippopotamus limb bones (*a–d*) and teeth (*small open circles*), many fragments of bone (*short dashes*) and a few stone artifacts (*colored dots*). Trenching (*dashed line, color*) and hillside excavation over a wide area exposed an ancient soil surface (*color*) overlying a deposit of silty tuff. Lying on the ancient surface were stone cores (*open circles, color*) from which sharp-edged flakes had been struck, more than 100 other stone artifacts and more than 60 additional fragments of teeth and bones. The scatter of tools and broken bones suggests the hypothesis that the toolmakers fed on meat from the hippopotamus.

fact association, only a kilometer away from the hippopotamus/artifact site, has allowed us to carry our inquiries further.

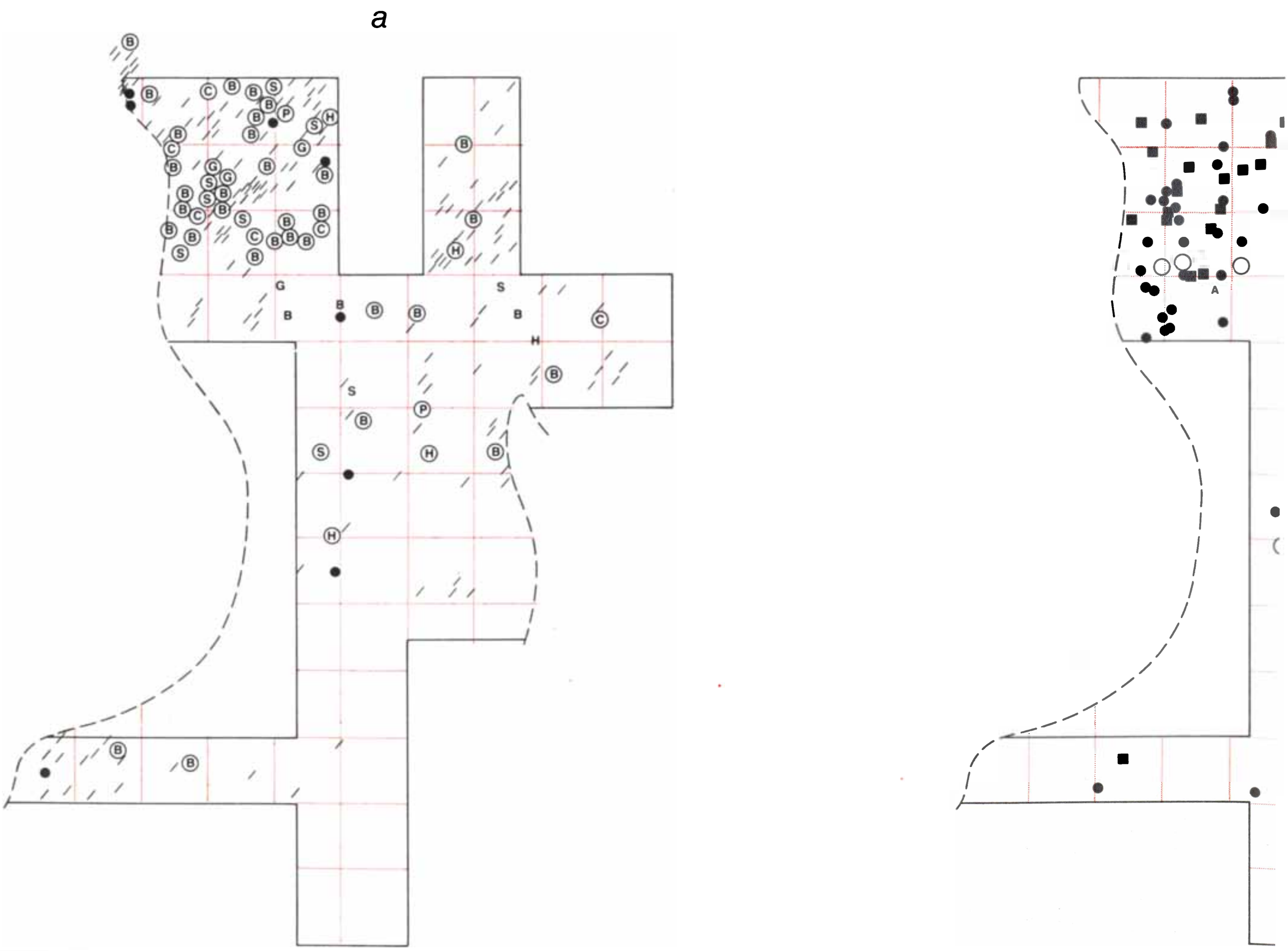
The second site had been located by Behrensmeyer in 1969. Erosion was uncovering artifacts, together with pieces of broken-up bone, at another outcrop of the same volcanic ash layer that contained the HAS artifacts and bones. With the assistance of John Barthelme of the University of California at Berkeley and others I began to excavate the site. The work soon revealed a scatter of several hundred stone tools in an area 16 meters in diameter. They rested on an ancient ground surface that had been covered by layers of sand and silt. The concentration of artifacts exactly coincided with a scatter of fragmented bones. Enough of them, teeth in particular, were identifiable to demonstrate that parts of the remains of several animal species were present. John M. Harris of the Louis Leakey Memorial Institute in Nairobi recognized, among other species, hippopotamus, giraffe, pig, porcupine and such bovids as waterbuck, gazelle and what may be either hartebeest or wildebeest. It was this site that was designated KBS. The site obviously represented the second category of bone-and-artifact associations: tools in association with the remains of many different animal species.

Geological evidence collected by A. K. Behrensmeyer of Yale University and others shows that the KBS deposit had accumulated on the sandy bed of a stream that formed part of a small delta. At the time when the toolmakers used the stream bed, water had largely ceased to flow. Such a site was probably favored as a focus of hominid activity for a number of reasons. First, as every beachgoer knows, sand is comfortable to sit and lie on. Second, by scooping a hole of no great depth in the sand of a stream bed one can usually find water. Third, the growth of trees and bushes in the sun-parched floodplains of East Africa is often densest along watercourses, so that shade and plant foods are available in these locations. It may also be that the protohuman toolmakers who left their discards here took shelter from predators by climbing trees and also spent their nights protected in this way.

Much of this is speculative, of course, but we have positive evidence that the objects at the KBS site did accumulate in the shade. The sandy silts that came to cover the discarded implements and fractured bones were deposited so gently that chips of stone small enough to be blown away by the wind were not disturbed. In the same silts are the impressions of many tree leaves. The species of tree has not yet been formally identified, but Jan Gilette of the Kenya National Herbarium notes that the impressions closely resemble the leaves of African wild fig trees.

Carrying Stones and Meat

As at the hippopotamus/artifact site, we have established the fact that stones



BONES AND STONE TOOLS were also found in abundance at the Kay Behrensmeyer site. As the plot of bone distribution (*a*) shows, the animal remains represent many different species. These are identified by capital letters; if the find was a tooth the letter is circled. Most are small to medium-sized bovids, such as gazelle, waterbuck and hartebeest (*B*). The remains of crocodile (*C*), giraffe (*G*), hippopotamus (*H*), porcupine (*P*) and extinct species of pig (*S*) were also present. Dots and dashes locate unidentified teeth and bone fragments respectively.

larger than the size of a pea do not occur naturally closer to the Kay Behrensmeyer site than a distance of three kilometers. Thus we know that the stones we found at the site must have been carried at least that far. With the help of Frank Fitch and Ron Watkins of the University of London we are searching for the specific sources.

It does not seem likely that all the animals of the different species represented among the KBS bones could have been killed in a short interval of time at this one place. Both considerations encourage the advancement of a tentative hypothesis: Like the stones, the bones were carried in, presumably while there was still meat on them.

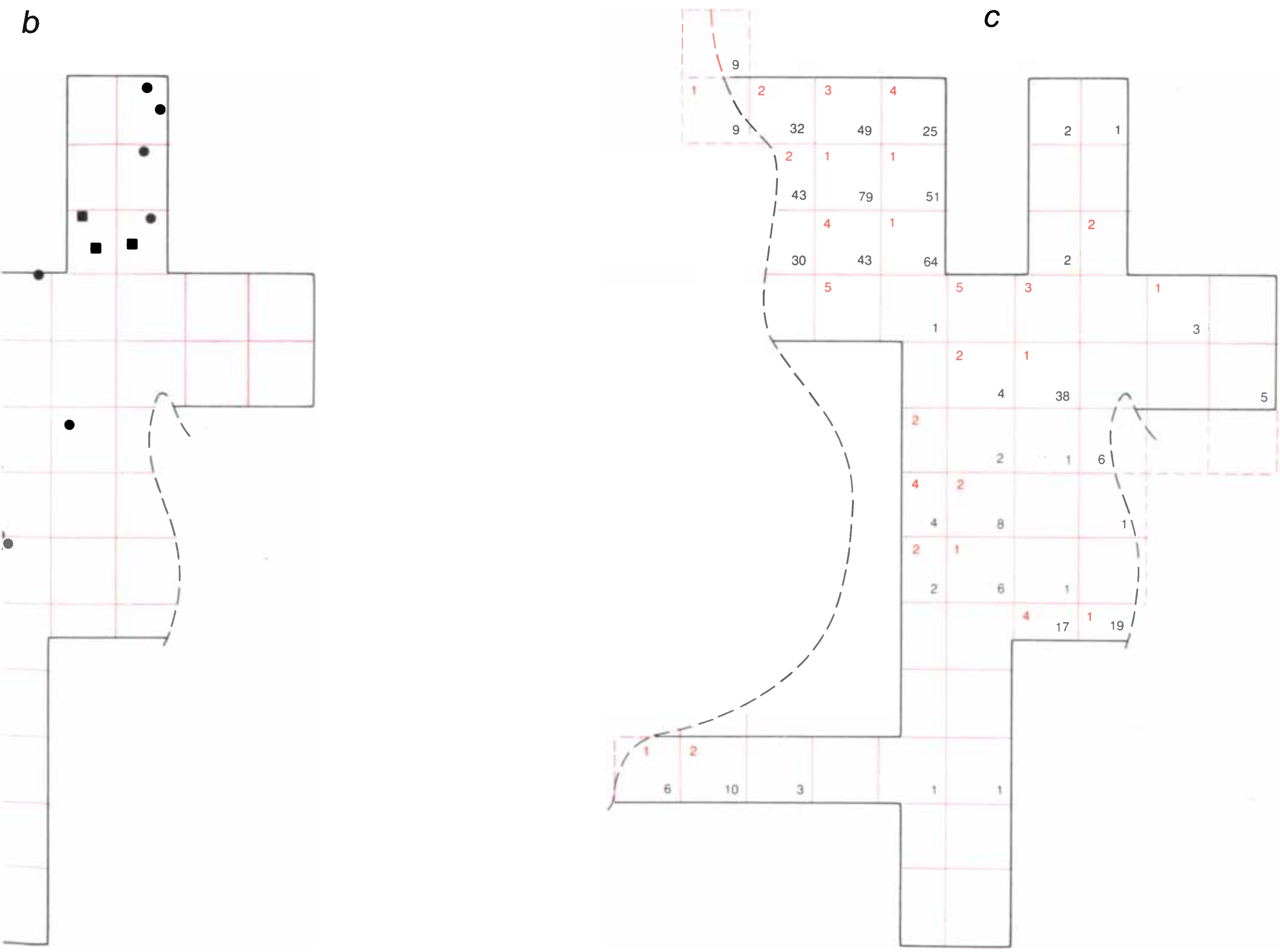
If this hypothesis can be accepted, the Kay Behrensmeyer site provides very early evidence for the transport of food as a protohuman attribute. Today the carrying of food strikes us as being commonplace, but as Sherwood Washburn of the University of California at Berkeley observed some years ago such an action would strike a living ape as being novel and peculiar behavior indeed. In short, if the hypothesis can be accepted, it suggests that by the time the KBS deposit was laid down various fundamental shifts had begun to take place in hominid social and ecological arrangements.

It should be noted that other early sites in this category are known in East Africa, so that the Kay Behrensmeyer site is by no means unique. A number of such sites have been excavated at Olduvai Gorge and reported by Mary Leakey. Of these the best preserved is the "Zinjanthropus" site of Olduvai Bed I, which is about 1.7 million years old. Here too a dense patch of discarded artifacts coincides with a concentration of broken-up bones.

There is an even larger number of Type A sites (where concentrations of artifacts are found but bones are virtually or entirely absent). Some are at Koobi Fora; others are in the Omo Valley, where Harry V. Merrick of Yale University and Jean Chavaillon of the French National Center for Scientific Research (CNRS) have recently uncovered sites of this kind in members E and F of the Shungura Formation. The Omo sites represent the oldest securely dated artifact concentrations so far reported anywhere in the world; the tools were deposited some two million years ago.

One of the Olduvai sites in this category seems to have been a "factory": a quarry where chert, an excellent tool material, was readily available for flaking. The other tool concentrations, with very few associated bones or none at all, may conceivably be interpreted as foci of hominid activity where for one reason or another large quantities of meat were not carried in. Until it is possible to distinguish between sites where bone was never present and sites where the bones have simply vanished because of such factors as decay, however, these deposits will remain difficult to interpret in terms of subsistence ecology.

What, in summary, do these East Af-



The plot of artifact distribution (*b*) shows that three of four stone cores (*open circles*), most waste stone (*squares*) and flakes and fragments of flakes (*dots*) were found in 12 adjacent squares. Also found here was an unworked stone (*A*) that, like the cores, must have been carried to the site from a distance. Plotting of all tools and bones unearthed at the site was not attempted. Numbers in grid squares (*c*) show how many flakes and bits of waste stone (*color*) and fragments of bone (*black*) were recorded without exact plotting in each square.

rican archaeological studies teach us about the evolution of human behavior? For one thing they provide unambiguous evidence that two million years ago some hominids in this part of Africa were carrying things around, for example stones. The same hominids were also making simple but effective cutting tools of stone and were at times active in the vicinity of large animal carcasses, presumably in order to get meat. The studies strongly suggest that the hominids carried animal bones (and meat) around and concentrated this portable food supply at certain places.

Model Strategies

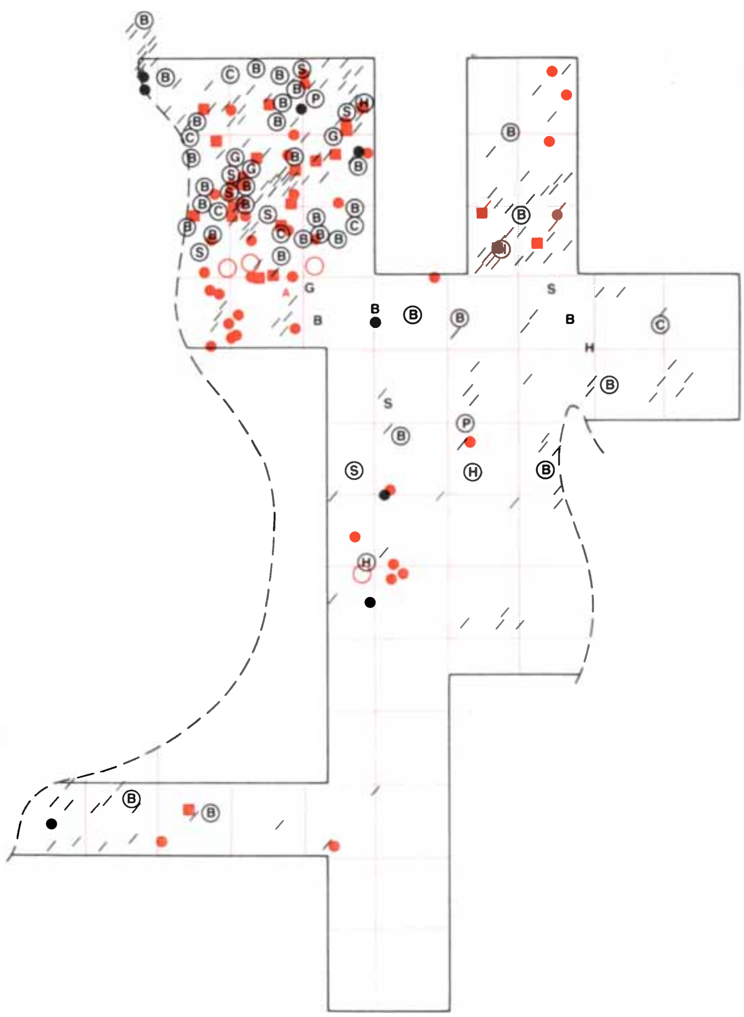
These archaeological facts and indications allow the construction of a theoretical model that shows how at least some aspects of early hominid social existence may have been organized. Critical to the validity of the model is the inference that the various clusters of remains we have uncovered reflect social and economic nodes in the lives of the toolmakers who left behind these ancient patches of litter. Because of the evidence suggestive of the transport of food to certain focal points, the first question that the model must confront is why early hominid social groups departed from the norm among living subhuman primates, whose social groups feed as they range. To put it another way, what ecological and evolutionary advantages are there in postponing some food consumption and transporting the food?

Several possible answers to this question have been advanced. For example, Adrienne Zihlman and Nancy Tanner of the University of California at Santa Cruz suggest that when the protohumans acquired edible plants out on the open grasslands, away from the shelter of trees, it would have been advantageous for them to seize the plant products quickly and withdraw to places sheltered from menacing predators. Others have proposed that when the early hominids foraged, they left their young behind at "nest" or "den" sites (in the manner of birds, wild dogs and hyenas) and returned to these locales at intervals, bringing food with them to help feed and wean the young.

If we look to the recorded data concerning primitive human societies, a third possibility arises. Among extant and recently extinct primitive human societies the transport of food is associated with a division of labor. The society is divided by age and sex into classes that characteristically make different contributions to the total food supply. One significant result of such a division is an increase in the variety of foodstuffs consumed by the group. To generalize on the basis of many different ethnographic reports, the adult females of the society contribute the majority of the "gathered" foods; such foods are mainly plant products but may include shellfish, amphibians and small reptiles, eggs, insects and the like. The adult males usually, although not invariably, contribute most of the "hunted" foodstuffs: the flesh of mammals, fishes, birds and so forth. Characteristically the males and females range in separate groups and each sex eventually brings back to a home base at least the surplus of its foraging.

Could this simple mechanism, a division of the subsistence effort, have initiated food-carrying by early hominids? One cannot dismiss out of hand the models that suggest safety from competitors or the feeding of nesting young as the initiating mechanisms for food-carrying. Nevertheless, neither model seems to me as plausible as one that has division of labor as the primary initiating mechanism. Even if no other argument favored the model, we know for a fact that somewhere along the line in the evolution of human behavior two patterns became established: food-sharing and a division of labor. If we include both patterns in our model of early hominid society, we will at least be parsimonious.

Other arguments can be advanced in favor of an early development of a division of labor. For example, the East African evidence shows that the protohuman toolmakers consumed meat from a far greater range of species and sizes of animals than are eaten by such living primates as the chimpanzee and the baboon. Among recent human hunter-gatherers the existence of a division of labor seems clearly related to the females being encumbered with children, a handicap that bars them from hunting or scavenging, activities that require speed afoot or long-range mobility. For the protohumans too the incorporation of meat in the diet in significant quantities may well have been a key factor in the development not only of a division



CLUSTERED MIXTURE of artifacts and animal bones at the Kay Behrensmeyer site is evident when the stone (*color*) and bone (*black*) plots are superposed. Combinations of this kind are sometimes produced by stream action, but such is not likely to be the case here, as is attested by the preservation of leaf impressions and other readily washed-away debris such as fine splinters of stone. It appears instead that the protohumans who made and discarded their tools here were also responsible for the bone accumulation because they met here to share their food.

At last. A little cartridge camera that does what the big cameras do.

The one and only Minolta 110 Zoom SLR.

Those little 110 cartridge-loading cameras give you lots of convenience, but not much versatility.

Big single lens reflex cameras give you lots of versatility, but may be more than you want to carry around.

Here's a new kind of camera that gives you the best of both worlds: the little Minolta 110 Zoom SLR.

How little is it?

Only 5.3 inches wide by 4.3 inches deep by 2.1 inches high. And it weighs only 15.1 ounces. You can carry it anywhere.

How easy is it to load?

Use drop-in film cartridges for color slides or color or black-and-white snapshots. You just slip in the cartridge. The camera sets the film speed automatically, including the new super-fast ASA 400.

It's a reflex camera.

A single lens reflex camera means that you look through the lens that's taking the picture. So you see beforehand what your slides or prints will be like.

Focusing is through the lens, also. It's quick and easy because of a microprism circle that breaks up the image until it's perfectly clear...and in focus.

And when you look through the viewfinder, you'll also see glowing lights which tell you if your batteries are O.K. and warn of over or under exposure.

Zoom lens.

This handy feature lets you zoom in for a bigger image or zoom back to get more into the scene. And it's built into the camera, not an extra-cost accessory.

Close-up lens.

For even closer shots, there's a built-in "macro" lens that lets you take pictures as close to your subject as 11.3 inches.

Automatic exposure, too.

You just choose a lens opening and the 110 Zoom SLR selects the precise shutter speed you need, automatically, anywhere from an action-stopping 1/1000th of a second to a full 10 seconds. And there's an "override" control for unusual lighting situations.

Automatic flash.

Slip the optional Minolta Auto Electroflash 25 onto the 110 Zoom SLR and you've got an automatic electronic flash camera that you can hold in the palm of your hand. You'll never need flashbulbs again.

The 110 Zoom SLR even comes with its own rubber lens shade to keep stray light from interfering with your pictures. And there's a safety lock to keep you from taking pictures accidentally.

If you'd like to know more about this revolutionary little camera, see your photo dealer or write Minolta Corporation, 101 Williams Drive, Ramsey, N.J. 07446.

MINOLTA

110 Zoom SLR. There's never been a camera like it.

of labor but also of the organization of movements around a home base and the transport and sharing of food.

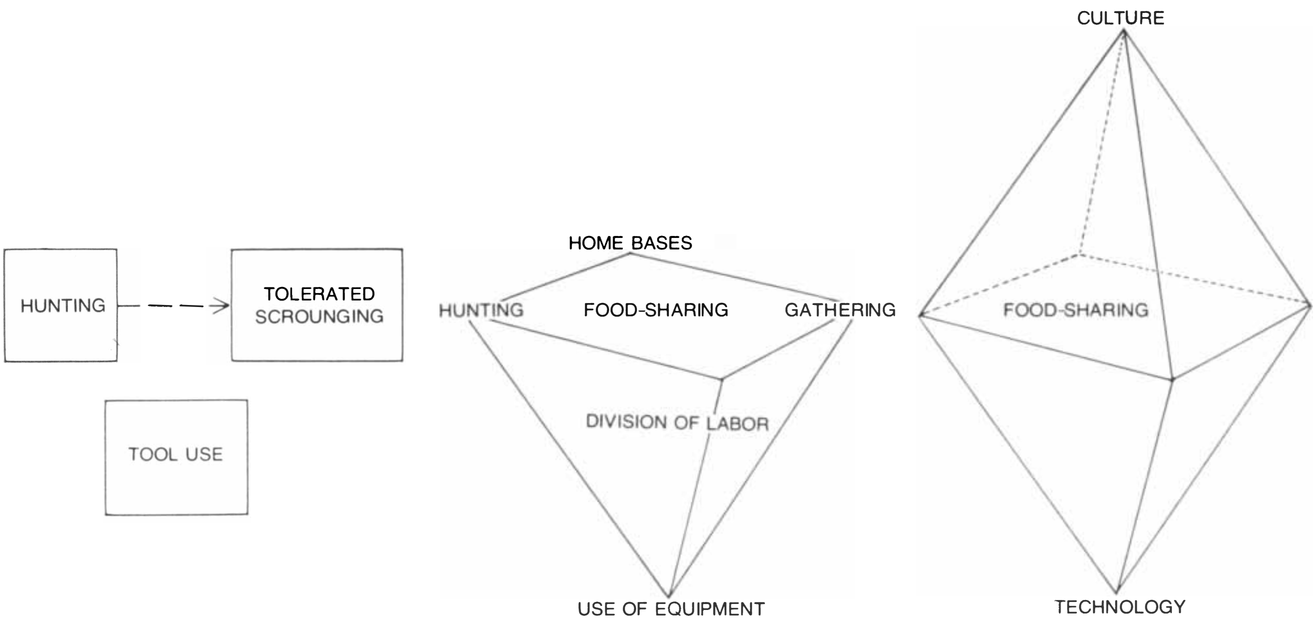
The model I propose for testing visualizes food-sharing as the behavior central to a novel complex of adaptations that included as critical components hunting and/or scavenging, gathering and carrying. Speaking metaphorically, food-sharing provides the model with a kind of central platform. The adaptive system I visualize, however, could only have functioned through the use of tools and other equipment. For example, without the aid of a carrying device primates such as ourselves or our ancestors could not have transported from the field to the home base a sufficient amount of plant food to be worth sharing. An object as uncomplicated as a bark tray would have served the purpose, but some such item of equipment would have been mandatory. In fact, Richard Borshay Lee of the University of Toronto has suggested that a carrying device was the basic invention that made human evolution possible.

What about stone tools? Our ancestors, like ourselves, could probably break up the body of a small animal, as chimpanzees do, with nothing but their hands and teeth. It is hard to visualize them or us, however, eating the meat of an elephant, a hippopotamus or some other large mammal without the aid of a cutting implement. As the archaeological evidence demonstrates abundantly, the protohumans of East Africa not only knew how to produce such stone flakes by percussion but also found them so useful that they carried the raw materials needed to make the implements with them from place to place. Thus whereas the existence of a carrying device required by the model remains hypothetical as far as archaeological evidence is concerned, the fact that tools were used and carried about is amply attested to.

In this connection it should be stressed that the archaeological evidence is also silent with regard to protohuman consumption of plant foods. Both the morphology and the patterns of wear observable on hominid teeth suggest such a plant component in the diet, and so does the weight of comparative data on subsistence patterns among living nonhuman primates and among nonfarming human societies. Nevertheless, if positive evidence is to be found, we shall have to sharpen our ingenuity, perhaps by turning to organic geochemical analyses. It is clear that as long as we do not correct for the imbalance created by the durability of bone as compared with that of plant residues, studies of human evolution will tend to have a male bias!

As far as the model is concerned the key question is not whether collectable foods—fruits, nuts, tubers, greens and even insects—were eaten. It is whether these protohumans carried such foods about. Lacking any evidence for the consumption of plant foods, I shall fall back on the argument that the system I visualize would have worked best if the mobile hunter-scavenger contribution of meat to the social group was balanced by the gatherer-carrier collection of high-grade plant foods. What is certain is that at some time during the past several million years just such a division of labor came to be a standard kind of behavior among the ancestors of modern man.

A final cautionary word about the model: The reader may have noted that I have been careful about the use of the words "hunter" and "hunting." This is because we cannot judge how much of the meat taken by the protohumans of East Africa came from opportunistic scavenging and how much was obtained by hunting. It is reasonable to assume that the carcasses of animals killed by carnivores and those of animals that had otherwise died or been disabled would always have provided active scavengers a certain amount of meat. For the present it seems less reasonable to assume that protohumans, armed primitively if at all, would be particularly effective hunters. Attempts are now under way, notably by Elizabeth Vrba of South Africa, to distinguish between assemblages of bones attributable to scavenging and assemblages attributable to hunting, but no findings from East Africa are yet available. For the present I am inclined to accept the verdict of J. Desmond



BEHAVIOR PATTERNS that differ in degree of organization are contrasted in these diagrams. Living great apes, exemplified here by the chimpanzee, exhibit behavior patterns that became important in human evolution but the patterns (*left*) exist largely as isolated elements. Hunting occurs on a small scale but leads only to "tolerated scrounging" rather than active food-sharing; similarly, tools are used but tool use is not integrated with hunting or scrounging. The author's model (*center*) integrates these three behavior patterns and others into a coherent structure. Food-sharing is seen as a central structural element, incorporating the provision of both animal and plant foods, the organization of a home base and a division of labor. Supporting the integrated structure is a necessary infrastructure of tool and equipment manufacture; for example, without devices for carrying foodstuffs there could not be a division of labor and organized food-sharing. In modern human societies (*right*) the food-sharing structure has undergone socioeconomic elaboration. Its infrastructure now incorporates all of technology, and a matching superstructure has arisen to incorporate other elements of what is collectively called culture.

Clark of the University of California at Berkeley and Lewis R. Binford of the University of New Mexico. In their view the earliest meat-eaters might have obtained the flesh of animals weighing up to 30 kilograms by deliberate hunting, but the flesh of larger animals was probably available only through scavenging.

Tools as Testimony

Of course, the adaptive model I have advanced here reflects only a working hypothesis and not established fact. Nevertheless, there is sufficient evidence in its favor to justify looking further at its possible implications for the course of human evolution. For example, the model clearly implies that early toolmaking hominids displayed certain patterns of behavior that, among the patterns of behavior of all primates, uniquely characterize our own species and set it apart from its closest living relatives, the great apes. Does this mean that the toolmaking hominids of 1.5 to two million years ago were in fact "human"?

I would surmise that it does not, and I have been at pains to characterize these East African pioneers as protohumans. In summarizing the contrasts between living men and living apes I put high on the list language and the cultural phenomena that are dependent on it. We have no direct means of learning whether or not any of these early hominids had language. It is my suspicion, however, that the principal evolutionary change in the hominid line leading to full humanity over the past two million years has been the great expansion of language and communication abilities, together with the cognitive and cultural capabilities integrally related to language. What is the evidence in support of this surmise?

One humble indicator of expanding mental capacities is the series of changes that appears in the most durable material record available to us: the stone tools. The earlier tools from the period under consideration here seem to me to show a simple and opportunistic range of forms that reflect no more than an uncomplicated empirical grasp of one skill: how to fracture stone by percussion in such a way as to obtain fragments with sharp edges. At that stage of toolmaking the maker imposed a minimum of culturally dictated forms on his artifacts. Stone tools as simple as these perform perfectly well the basic functions that support progress in the direction of becoming human, for example the shaping of a digging stick, a spear and a bark tray, or the butchering of an animal carcass.

The fact is that exactly such simple stone tools have been made and used ever since their first invention, right down to the present day. Archaeology also shows, however, that over the past several hundred thousand years some assemblages of stone tools began to reflect a greater cultural complexity on the part of their makers. The complexity is first shown in the imposition of more arbitrary tool forms; these changes were followed by increases in the number of such forms. There is a marked contrast between the pure opportunism apparent in the shapes of the earliest stone tools and the orderly array of forms that appear later in the Old Stone Age when each form is represented by numerous standardized examples in each assemblage of tools. The contrast strongly suggests that the first toolmakers lacked the highly developed mental and cultural abilities of more recent humans.

The evidence of the hominid fossils and the evidence of the artifacts together suggest that these early artisans were nonhuman hominids. I imagine that if we had a time machine and could visit a place such as the Kay Behrensmeyer site at the time of its original occupation, we would find hominids that were living in social groups much like those of other higher primates. The differences would be apparent only after prolonged observation. Perhaps at the start of each day we would observe a group splitting up as some of its members went off in one direction and some in another. All these subgroups would very probably feed intermittently as they moved about and encountered ubiquitous low-grade plant foods such as berries, but we might well observe that some of the higher-grade materials—large tubers or the haunch of a scavenged carcass—were being reserved for group consumption when the foraging parties reconvened at their starting point.

To the observer in the time machine behavior of this kind, taken in context with the early hominids' practice of making tools and equipment, would seem familiarly "human." If, as I suppose, the hominids under observation communicated only as chimpanzees do or perhaps by means of very rudimentary protolinguistic signals, then the observer might feel he was witnessing the activities of some kind of fascinating bipedal ape. When one is relying on archaeology to reconstruct protohuman life, one must strongly resist the temptation to project too much of ourselves into the past. As Jane B. Lancaster of the University of Oklahoma has pointed out, the hominid life systems of two million years ago have no living counterparts.

Social Advances

My model of early hominid adaptation can do more than indicate that the first toolmakers were culturally protohuman. It can also help to explain the dynamics of certain significant advances in the long course of mankind's development. For example, one can imagine that a hominid social organization in-

volving some division of labor and a degree of food-sharing might well have been able to function even if it had communicative abilities little more advanced than those of living chimpanzees. In such a simple subsistence system, however, any group with members that were able not only to exchange food but also to exchange information would have gained a critical selective advantage over all the rest. Such a group's gatherers could report on scavenging or hunting opportunities they had observed, and its hunters could tell the gatherers about any plant foods they had encountered.

By the same token the fine adjustment of social relations, always a matter of importance among primates, becomes doubly important in a social system that involves food exchange. Language serves in modern human societies not only for the exchange of information but also as an instrument for social adjustment and even for the exchange of misinformation.

Food-sharing and the kinds of behavior associated with it probably played an important part in the development of systems of reciprocal social obligations that characterize all human societies we know about. Anthropological research shows that each human being in a group is ordinarily linked to many other members of the group by ties that are both social and economic. The French anthropologist Marcel Mauss, in a classic essay, "The Gift," published in 1925, showed that social ties are usually reciprocal in the sense that whereas benefits from a relationship may initially pass in only one direction, there is an expectation of a future return of help in time of need. The formation and management of such ties calls for an ability to calculate complex chains of contingencies that reach far into the future. After food-sharing had become a part of protohuman behavior the need for such an ability to plan and calculate must have provided an important part of the biological basis for the evolution of the human intellect.

The model may also help explain the development of human marriage arrangements. It assumes that in early protohuman populations the males and females divided subistence labor between them so that each sex was preferentially tapping a different kind of food resource and then sharing within a social group some of what had been obtained. In such circumstances a mating system that involved at least one male in "family" food procurement on behalf of each child-rearing female in the group would have a clear selective advantage over, for example, the chimpanzees' pattern of opportunistic relations between the sexes.

I have emphasized food-sharing as a principle that is central to an understanding of human evolution over the

past two million years or so. I have also set forth archaeological evidence that food-sharing was an established kind of behavior among early protohumans. The notion is far from novel; it is implicit in many philosophical speculations and in many writings on paleoanthropology. What is novel is that I have undertaken to make the hypothesis explicit so that it can be tested and revised.

Accounting for Evolution

Thus the food-sharing hypothesis now joins other hypotheses that have been put forward to account for the course of human evolution. Each of these hypotheses tends to maintain that one or another innovation in protohuman behavior was the critical driving force of change. For example, the argument has been advanced that tools were the "prime movers." Here the underlying implication is that in each successive generation the more capable individuals made better tools and thereby gained advantages that favored the transmission of their genes through natural selection; it is supposed that these greater capabilities would later be applied in aspects of life other than technology. Another hypothesis regards hunting as being the driving force. Here the argument is that hunting requires intelligence, cunning, skilled neuromuscular coordination and, in the case of group hunting, cooperation. Among other suggested prime movers are such practices as carrying and gathering.

If we compare the food-sharing explanation with these alternative explanations we see that in fact food-sharing incorporates many aspects of each of the others. It will also be seen that in the food-sharing model the isolated elements are treated as being integral parts of a complex, flexible system. The model itself is probably an oversimplified version of what actually happened, but it seems sufficiently realistic to be worthy of testing through further archaeological and paleontological research.

Lastly, the food-sharing model can be seen to have interconnections with the physical implications of fossil hominid anatomy. For example, a prerequisite of food-sharing is the ability to carry things. This ability in turn is greatly facilitated by a habitual two-legged posture. As Gordon W. Hewes of the University of Colorado has pointed out, an important part of the initial evolutionary divergence of hominids from their primate relatives may have been the propensity and the ability to carry things about. To me it seems equally plausible that the physical selection pressures that promoted an increase in the size of the protohuman brain, thereby surely enhancing the hominid capacity for communication, are a consequence of the shift from individual foraging to food-sharing some two million years ago.

The Birth of Massive Stars

The birth of stars many times hotter and more massive than the sun may be triggered by shock waves traveling through large cool clouds of interstellar gas and dust

by Michael Zeilik

Stars in our galaxy range in mass from a hundredth the mass of the sun to about a hundred solar masses. The least massive stars have lifetimes as long as the age of the universe: tens of billions of years, after which they slowly fade to become cold cinders. The massive stars have lifetimes of only a few millions to tens of millions of years, after which they catastrophically explode. During their short life span the thermonuclear furnace deep within them manufactures elements as heavy as carbon and iron; at their death the awesome violence of the supernova explosion forges elements heavier than iron and blasts as much as 90 percent of the star's material into interstellar space. Out of this recycled material new stars and planets will be born: stars such as the sun and planets such as the earth. Moreover, life arose on our planet because massive stars lived and died; without supernova explosions the carbon that is the key to life as we know it would not be distributed throughout interstellar space.

Since the death throes of massive stars are so crucial to cosmic evolution, such stars raise critical questions for astronomy. How are massive stars born? Where are they born? How do conditions become ripe for their birth? Within the past few years new radio and infrared telescopes have made it possible for many astronomers to probe the birthplaces of massive stars. Recent observations seem to indicate that the birth of such stars is sequentially triggered by shock waves traveling through gigantic clouds of interstellar gas and dust.

Massive stars are bluish-white giant stars of the spectral types O and B. The surface temperature of these stars is much hotter than that of the sun: between 16,000 and 45,000 degrees Kelvin. (The surface temperature of the sun is 6,000 degrees K.) Such stars are also between 800 and 500,000 times more luminous than the sun. They range in mass from about six to 60 solar masses. A crucial fact about Type O and Type B stars is that most of them are located in the spiral arms of the galaxy.

Where and how do Type O and Type B stars form? To answer that question one first needs a general picture of how one might expect massive stars to be born. One simple physical concept lies behind the formation of stars: gravitational instability. The concept is not new; Newton first perceived it late in the 17th century.

Imagine a uniform, static cloud of gas in space. Imagine then that the gas is somehow disturbed so that one small spherical region becomes a little denser than the gas around it so that the small region's gravitational field becomes slightly stronger. It now attracts more matter to it and its gravity increases further, causing it to begin to contract. As it contracts its density increases, which increases its gravity even more, so that it picks up even more matter and contracts even further. The process continues until the small region of gas finally forms a gravitationally bound object.

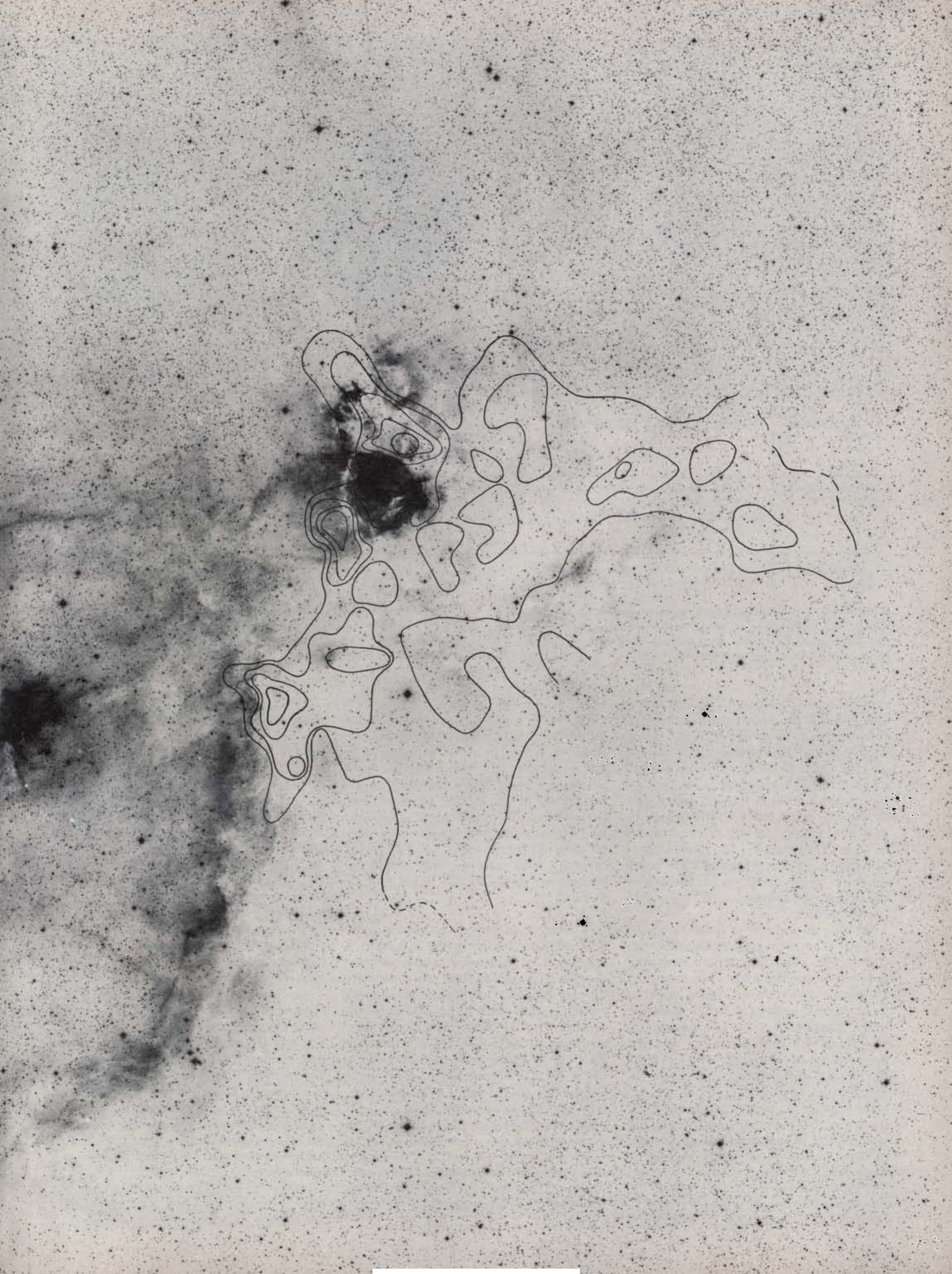
This simple picture suffers from a serious complication that was first pointed out by James Jeans at the beginning of the century: the role of the internal pressure of the gas. The gas's internal pressure arises from collisions between the particles in it, and is directly proportional to the gas's temperature and density. The effect of the internal pressure is to push the sphere of gas apart from within. A cloud of gas contracting under the force of its own gravity will heat up because the gravitational energy of the infalling matter is transformed into heat. Thus a contracting region of gas has a higher internal pressure than its surroundings both from its higher density and from its higher temperature. In order for the inward pull of gravity to win out and to form the gas into a gravitationally bound object such as a star, the outward push of the internal pressure must be overcome. Basically gravity will win when the disturbed region is large enough to have enough mass to contract in spite of its internal pressure.

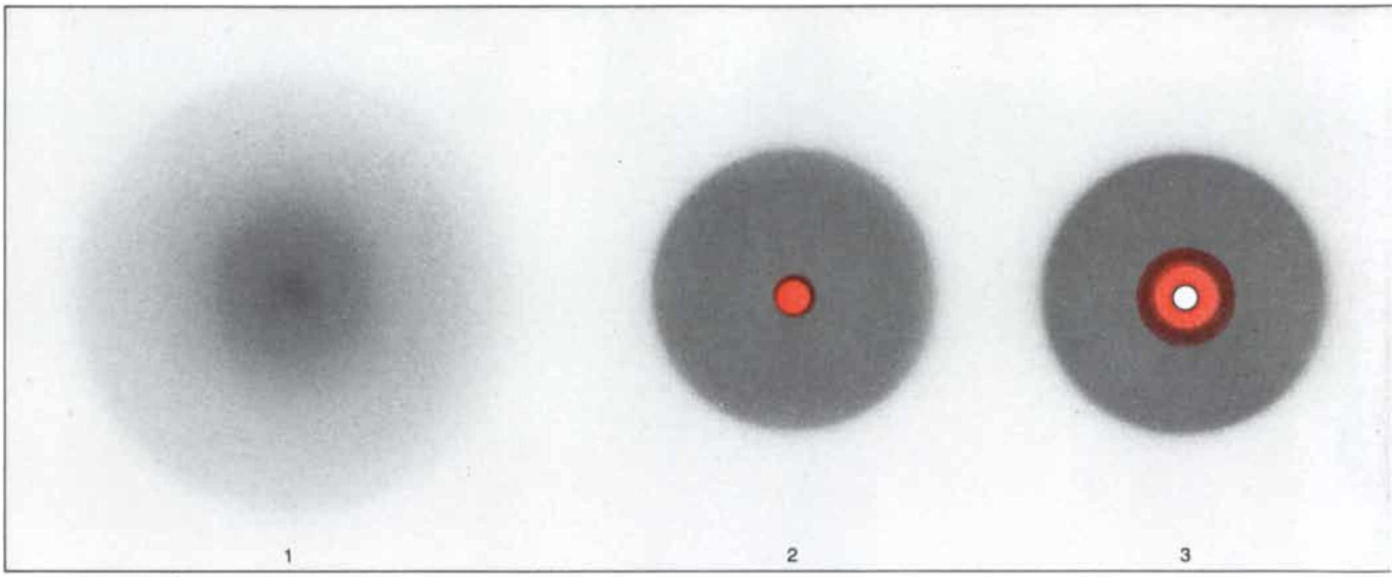
Let us assume for the moment that these conditions can be met in a cloud of molecules in interstellar space, and let us look at the cloud both before and after the gravitational instability has been set up. How will the gravitationally unstable portion of the cloud evolve into a star?

Theoretical work indicates that the forming star will progress through several evolutionary stages, and that each stage is marked by different observational signposts that can be identified with different kinds of telescopes. Let me outline one possible sequence of events for the birth of a single massive star to show what some of the signposts are that announce the different stages of its formation.

Molecular clouds from which massive stars are born consist of grains of dust and the molecules of gases such as molecular hydrogen (H_2), the hydroxyl radical (OH), hydrogen cyanide (HCN) and carbon monoxide (CO). Most of the

GLOWING NEBULAS IN CASSIOPEIA, designated IC 1795 and IC 1805, are prime birthplaces of hot, bluish-white giant stars of spectral types O and B. In this negative print made with the 48-inch Schmidt telescope on Palomar Mountain the nebulas appear black: IC 1795 is the dense portion of the cloud near the north (*center*) of the photograph and IC 1805 is the dense knot to the southeast (*middle left*). The glowing gas of the nebulas is largely composed of atoms of hydrogen that have been ionized, or dissociated, by the ultraviolet radiation of the stars buried deep within them. Ionized hydrogen is designated H II, and the glowing nebulas are known as H II regions. The massive stars themselves are obscured from view at visible wavelengths by the gas and dust. The photograph shows that IC 1795 and IC 1805 are actually part of one large complex of gas and dust. They are also sites of radio emission, and their radio-wavelength components have been respectively designated W3 and W4. The contour lines, which map the presence of carbon monoxide (CO) show that to the west (*right*) of the nebulas there is a large invisible molecular cloud. Two concentrations of contour lines just above and below IC 1795 are sites of particularly strong radio emission; the one above is W3 A and the one below is W3 OH. They are believed to be regions that are collapsing to form new massive stars.





EVOLUTIONARY STAGE	PRESTELLAR CLOUD COLLAPSES	VERY YOUNG COOL STAR SHINES FROM GRAVITATIONAL CONTRACTION	STAR BEGINS THERMONUCLEAR FUSION IN ITS CORE
OBSERVATIONAL SIGNPOSTS	COOL DENSE MOLECULAR CLOUD; SOURCES OF RADIO EMISSION FROM MOLECULES SUCH AS CARBON MONOXIDE (CO)	COMPACT FAR-INFRARED SOURCE ASSOCIATED WITH MOLECULAR CLOUD RADIO EMISSION FROM HYDROXYL (OH) AND WATER (H_2O)	COMPACT NEAR-INFRARED SOURCE AND RADIO SOURCE IN MOLECULAR CLOUD
DURATION (YEARS)	300,000	25,000	25,000

BIRTH OF A MASSIVE STAR happens quickly in a sequence of stages, each of which has a distinct observational signpost by which it can be detected. In the first stage a shock wave causes a portion of a large, cool cloud of gas and dust (*medium gray*) to become gravitationally unstable, and that region (*dark gray*) begins to collapse (*1*). As it collapses it becomes hotter and denser (*2*) until it begins to glow red (*color dot*) from the heat liberated by the gravitational collapse. Soon thermonuclear fusion begins in the star's core (*3*), and the star's sur-

molecules are hydrogen. The dust in the cloud is mixed in with the gas, and on the average there is about one grain of dust for every 10^{12} atoms of gas. A typical molecular cloud may have a diameter of tens of parsecs (one parsec is 3.26 light-years), a mass 100,000 times that of the sun, a density of 1,000 molecules per cubic centimeter and a temperature of 15 or 20 degrees K. The observational signpost for this prestellar stage is the emission by the cloud's many molecules of radio waves at wavelengths measured in millimeters.

Picture a portion of such a cloud somehow becoming gravitationally unstable. The unstable portion contracts, gets denser and heats up. The core of the fragment grows denser faster than its outer layers, collapses faster than the outer layers and heats up quickly; in 10,000 or 100,000 years the internal temperature reaches eight or 10 million degrees K.: the ignition temperature for thermonuclear reactions. When the nuclear fires ignite, they generate so much heat that the internal pressure in the core of the fragment increases enough to halt the gravitational collapse. A massive star is born.

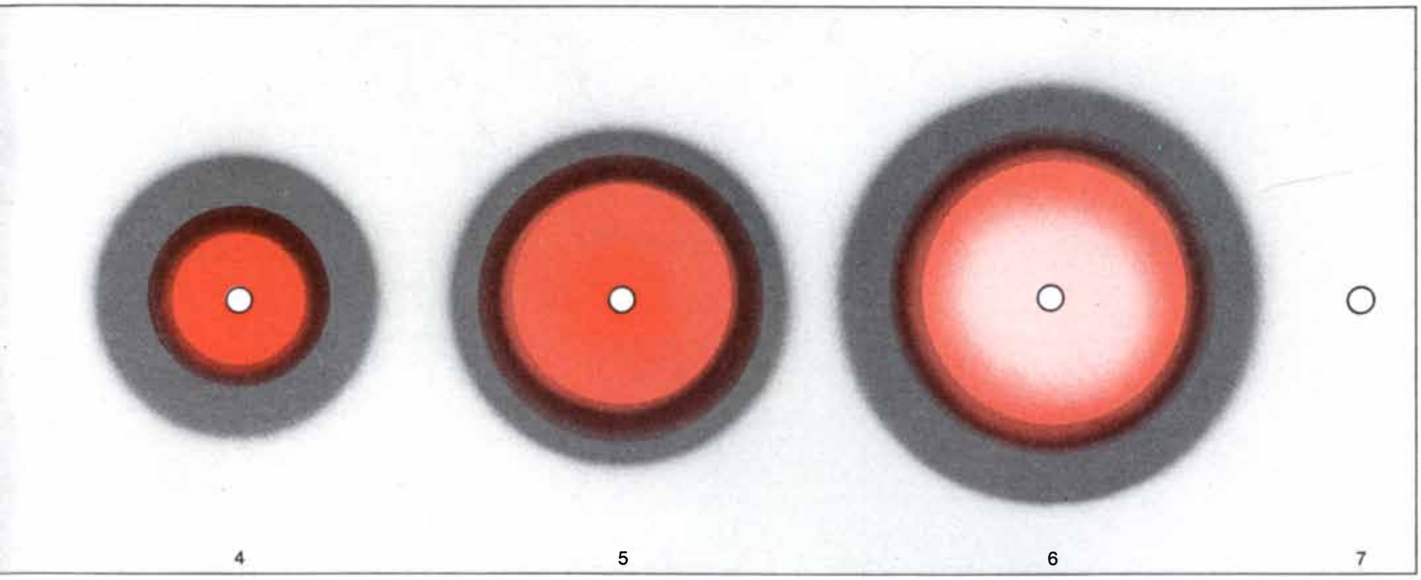
Theoretical work indicates that in the beginning the newly formed massive star has a low surface temperature but a high luminosity. As the star evolves its temperature rapidly increases but its luminosity remains roughly the same. For example, a star 15 times as massive as the sun begins life with a surface temperature of about 4,000 degrees K. and a luminosity about 10,000 times greater than that of the sun. In a mere 140,000 years or so its surface temperature rises to some 32,000 degrees K., but its luminosity will increase only about 60 percent to some 16,000 times that of the sun.

How can an infant massive star be detected? Right after it is born it cannot be seen at visible wavelengths because it is still hidden from view by the surrounding dust. Its evolution from having a cool surface to having a hot one affects the enveloping gas and dust, however, in ways that can be observed. The key physical process is that as a star's surface temperature goes up it emits more of its radiation in higher-energy, shorter-wavelength form. For example, as the surface temperature of the 15-solar-mass star rises from 4,000 to 32,000 degrees, the peak of its output shifts from a wavelength of 7,000 angstroms in the red region of the visible spectrum to 910 angstroms in the ultraviolet.

What does this shift do to the gas and the dust around the young star? At first the star emits mostly low-energy red light. The dust readily absorbs the light, blocking out a direct view of the star at visible wavelengths. Because the dust grains absorb the star's energy, however, they heat up. Since the energy they absorb is low, they reach a temperature of only about 30 or 50 degrees K. At those temperatures the dust grains emit most of their radiation in the far-infrared region of the spectrum: radiation with a wavelength of 100 micrometers or longer. Therefore the first sign that a newborn star is present in a molecular cloud is a compact source of far-infrared emission.

At this early stage small eddies or clumps of gas within the cloud may temporarily form and reach a suitable density and temperature for water and hydroxyl molecules to emit radio waves. The sites of water and hydroxyl emission are very small, about the size of the solar system, and probably exist for less than 10,000 years. This emission can readily be observed with radio telescopes and marks the birthplace of a new star.

As the infant star evolves and its surface temperature rises the dust surrounding it must absorb the higher-ener-



STAR IS IN EARLY NORMAL LIFE	YOUNG H II REGION BEGINS TO EXPAND	OLD VERY-EXPANDED H II REGION BEGINS BLENDING INTO INTERSTELLAR MEDIUM	NAKED TYPE O OR B STAR
INFRARED AND RADIO SOURCE IN MOLECULAR CLOUD	WEAK INFRARED EMISSION; DIFFUSE CENTIMETER-WAVELENGTH RADIO EMISSION; BRILLIANT NEBULA VISIBLE; TYPE O OR B STAR JUST VISIBLE	NO INFRARED EMISSION; VERY DIFFUSE CENTIMETER-WAVELENGTH RADIO EMISSION; FAINT NEBULA VISIBLE; TYPE O OR B STAR PLAINLY VISIBLE	REMAINS A SINGLE TYPE O OR B STAR WITH NO SURROUNDING H II REGION
30,000	500,000	2,000,000	6,000,000

face temperature begins to rise until the star becomes white or blue (*white dot*). The energy from the star begins to heat up the surrounding gas; in a few tens of thousands of years the young star is emitting so much ultraviolet radiation that it begins to ionize the gas in the cloud around it (*4*) to form an H II region (*light color*). The heated gas in the H II region begins expanding outward into the cool molecular cloud (*5*), until it sweeps much of the cloud away (*6*) into the interstellar medium (*light gray*) leaving the naked massive star (*7*).

gy radiation. This radiation raises the temperature of the dust to between 100 and 150 degrees K. At that temperature the dust in a small region around the star radiates most of its energy in the near-infrared region of the spectrum: radiation having a wavelength of between 10 and 30 micrometers. The dust farther out is heated less and mostly emits far-infrared radiation. Thus the higher-surface-temperature stage in the evolution of a massive star is heralded by the presence of a compact source of near-infrared emission surrounded by a more extended source of far-infrared emission.

Eventually the star begins to radiate copious amounts of ultraviolet radiation. The ultraviolet photons work on the gas in the molecular cloud in two ways. First, the radiation dissociates molecules of hydrogen into atoms. Second, as the star begins to emit photons with a wavelength shorter than 912 angstroms, the ultraviolet radiation begins to ionize the atoms of hydrogen, dissociating them into their constituent protons and electrons. Ionized hydrogen is designated H II, and so the zone of ionized hydrogen that forms immediately around the star is known as an H II region. In the H II region free electrons zip past protons (the hydrogen nuclei) and generate radio waves with a wavelength of about a centimeter. The centimeter radio waves easily penetrate the dust cloaking the star, so at this stage in the star's evolution its presence is marked by a compact region of centimeter-wavelength radio emission. Meanwhile the dust mixed in with the ionized hydrogen in the H II region continues to emit near-infrared radiation.

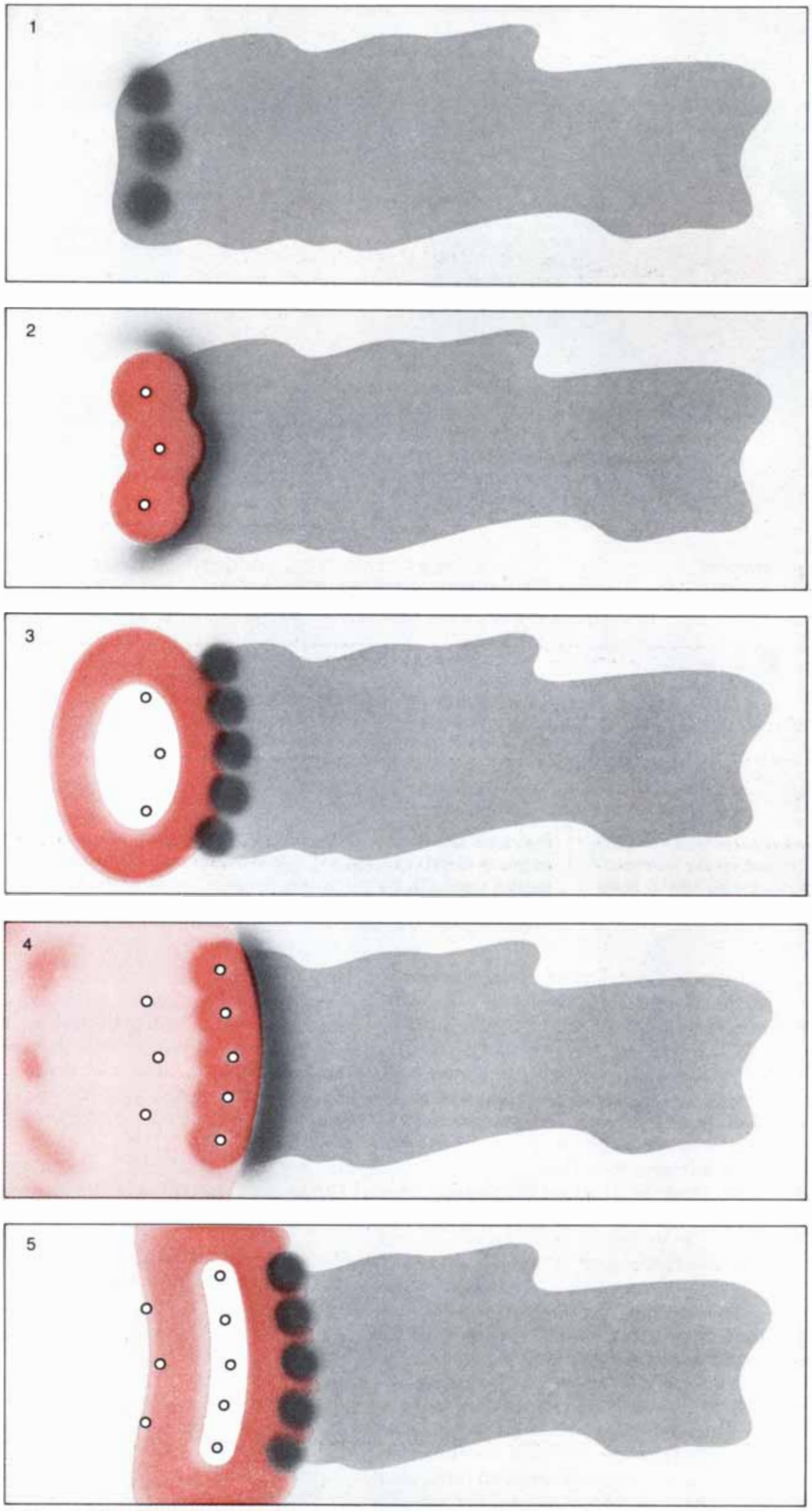
Around the star the ultraviolet photons absorbed by the gas heat the H II region to some 10,000 degrees K., so that the H II region is much hotter than the gas surrounding it. The hot H II region expands, pushing into the cooler molecular cloud and thinning out the dust and gas surrounding the star. After a few tens of thousands of years enough dust will have been swept away for it to no longer obscure the H II region. At that stage the H II region continues to emit radio waves and infrared radiation, although less intensely, and the ionized hydrogen becomes visible. The most familiar H II region at this stage visible from the solar system is the Great Nebula in Orion.

As time passes the H II region will continue to expand, all the while pushing out the molecular cloud. The strength of the radio and infrared emission will drop as the ionized gas becomes less dense. At that stage one observes a large, spread-out H II region weakly emitting radio waves and no longer emitting any infrared radiation.

At last the H II region balloons into the interstellar medium, blowing away the molecular cloud. The massive star is left naked in space and its birth process is complete. At that stage the telltale placental material has been discarded. The entire process takes only a few million years.

In this theoretical outline of the birth of a massive star I have ignored two critical points. The first is the observational facts that many of the Type O and Type B stars in the galaxy are found in loose groupings of a few hundred stars known as OB associations, and that many H II regions have within them not a single star but a small cluster of Type O and Type B stars. Such observations imply that massive stars frequently form not in isolation but in groups. The second point I have ignored is a theoretical puzzle: What process promotes the gravitational instability necessary for the formation of massive stars?

Recently Bruce G. Elmegreen and Charles J. Lada of the Center for Astrophysics of the Harvard College Observatory and the Smithsonian Astrophysical Observatory have proposed a theo-



MASSIVE STARS FORM IN GROUPS known as OB associations, not in isolation. A shock wave induces a gravitational instability at one edge of a molecular cloud (*medium gray*). The shock wave typically causes a layer of gas to begin to collapse and form into stars (*1*). Their expanding H II regions (*color*) create shock waves of their own that travel through the cloud, sweeping up matter behind them (*2*). Eventually the density of that new layer of material is great enough to become gravitationally unstable, fragment and begin collapsing into a second generation of stars (*3*). These new stars eventually also create H II regions that expand (*4*), forming shock waves of their own that plow into the interstellar cloud (*5*), which induce a third generation of stars to form (*right*); meanwhile older generations have continued to evolve to maturity (*left*). In such a cloud one can see evolutionary sequence of star formation across cloud.

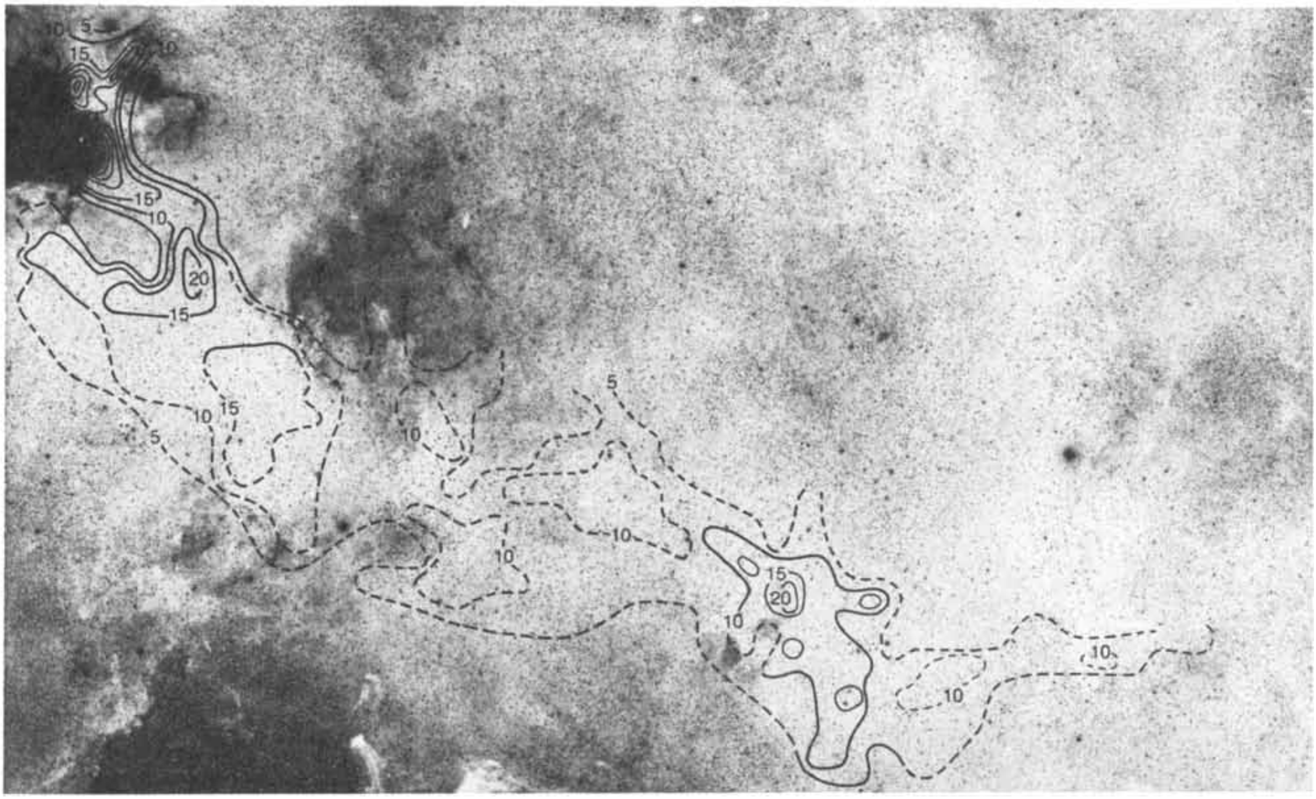
retical model for the formation of massive stars that relates the two points. Their model relies on an observation made by Adriaan Blaauw of the University of Leiden some years ago. Blaauw observed that OB associations are composed of separate subgroups of stars; each subgroup consists of between four and 20 stars and ranges in extent between two parsecs and 200 parsecs. The subgroups of a single OB association appear to be lined up in an evolutionary sequence: the oldest, most spread-out subgroup is found at one edge of the association and the youngest, most compact subgroup is found at the other edge. This sequential lineup of subgroups indicates that the primordial cloud out of which the OB association formed underwent successive bursts of star formation. In each burst a subgroup of stars was born.

How does it happen that the subgroups of an OB association are arranged in sequence of age? Elmegreen and Lada emphasize the fact that the H II region that forms around newly born massive stars expands and pushes into the surrounding molecular cloud at a speed of between five and 10 kilometers per second, fast enough to create a shock wave. As the shock wave plows into the molecular cloud, it picks up material in a layer behind it. After a few million years the material in the layer behind the shock front reaches a density high enough for it to become gravitationally unstable. It breaks into fragments, the fragments collapse and the formation of a new subgroup of massive stars begins. The new stars rise in temperature, ionize the gas around them and create their own H II region, which expands to create another shock wave that expands into the molecular cloud.

In a few million years the layer of material built up behind the new shock wave forms another subgroup of stars, and so the process repeats itself across the molecular cloud. In a chain reaction of self-destruction the molecular cloud, fragment by fragment and burst by burst, gives rise to subgroups of massive stars in a sequential chain, and their birth disrupts and dissipates the cloud. Cosmic evolution demands the cloud's death for the lives of the new stars.

Are there any observations that support the model that massive stars form in shock-driven sequences from giant molecular clouds? In my opinion there are. Indeed, it was recent radio and infrared observations that prompted Elmegreen and Lada to take a new look at the formation of OB associations.

I shall concentrate on the observational evidence in two H II regions: M17 (the Omega Nebula in Sagittarius) and W3 (Westerhout 3, the third object in an early catalogue of radio sources compiled in the 1950's by Gart Westerhout of the University of Maryland). In both



HUGE MOLECULAR CLOUD associated with M17 is shown in this negative print of a region in Sagittarius, made with the 48-inch Schmidt telescope on Palomar Mountain. M17 itself, and the current site of the birth of massive stars, is the black region in the far top left corner. The contour lines that wind across the photograph down and to the right trace the outline of the molecular cloud according to the emission detected from carbon monoxide, mapped by Bruce G. Elmegreen, Charles J. Lada and Dale F. Dickinson with the millimeter-wave radio telescope at the McDonald Observatory in Texas. The numbers on each contour line indicate the temperature of the carbon monoxide in degrees Kelvin. (Broken contour lines indicate regions where the mapping was less complete.) There are fewer stars visible in the region of the cloud because the cloud's dust blocks light of some background stars. This particular cloud is 2,500 parsecs away and 170 parsecs long. (A parsec is 3.26 light-years.) The illustrations on the next three pages show the site of star birth in greater detail.

regions the signposts of star formation line up in a clear pattern across our line of sight.

M17 is a bright, beautiful H II region easily visible in a small telescope. It is called the Omega Nebula because in a small telescope it has the shape of the Greek capital letter omega (Ω). Wide-angle photographs made with large telescopes reveal that to the east the brightest part of the nebula fades into a faint, lacy structure. In contrast, to the west the brightest part of the nebula seems to end abruptly.

With a radio telescope M17 is observed to emit strong centimeter-wavelength radiation. The amount of radio emission is what would be expected from a subgroup of about 10 massive stars. Curiously the peak of the radiation does not coincide with the brightest visible part of M17 but lies a bit to the west of it—at the sharp edge of the visible nebula. Therefore the position of the peak of the centimeter-wavelength radio emission indicates that some Type O and Type B stars must be hidden from view by dust.

With telescopes operating at millimeter wavelengths Lada and his colleagues have discovered two dense fragments of a molecular cloud just to the west of the brightest part of the H II region. In fact, the H II region appears to be pushing into those clouds. If the model of sequential star formation applies, then massive stars should be forming in the zone between the H II region and the southwestern molecular cloud.

Is that the case? M. Beetz of the Max Planck Institute for Astronomy in Heidelberg and his colleagues have photographed this region of M17 with special plates sensitive to infrared radiation. They found a cluster of at least six Type O and Type B stars in the interface zone between the H II region and the southwestern molecular cloud. Moreover, that zone is exactly the place where the centimeter-wavelength radio emission is the strongest. There the Type O and Type B stars must be, hidden by the dust from our view but providing the ultraviolet photons that ionize the gas and generate the radio emission.

There are a few other signposts of evidence for recent star formation. Douglas E. Kleinmann of the Center for Astrophysics and Edward L. Wright of the Massachusetts Institute of Technology have found an intense starlike source of 20-micrometer emission at the center of the southwestern molecular cloud, close to two locations of strong emission from water molecules. Both of these sources of emission are next to the part of M17 that is emitting the centimeter radio waves. Elmegreen and Lada have also discovered a gigantic molecular cloud extending far to the southwest of M17. The two molecular clouds right next to M17 appear to be two small fragments of the larger cloud, which has an estimated size of 22 by 86 parsecs and a mass of as much as a million times that of the sun. Recently Elmegreen, Lada and Dale F. Dickinson have extended these observations and have found that the giant molecular cloud extends a total of 44 by 170 parsecs.

An evolutionary lineup stretches from the eastern side of the nebula (the oldest side) to the western side (the youngest) across our line of sight. The observed sequence is the following: first, a visible H II region in which Type O and Type B stars are embedded; second, a source of centimeter-wavelength radio emission and infrared radiation that surrounds a small cluster of Type O and Type B stars; third, two dense fragments of a molecular cloud, one of which seems to be forming stars; fourth, a huge strung-out molecular cloud consisting of many fragments. This observed sequence of phenomena fits the model of sequential star formation well. We are seeing a shock wave generated by an expanding H II region ram into a molecular cloud and promote the formation of massive stars. To the east of the shock wave lie pieces of the molecular cloud that formed stars millions of years ago; to the west of the shock wave lies more of the molecular cloud awaiting eventual self-destruction.

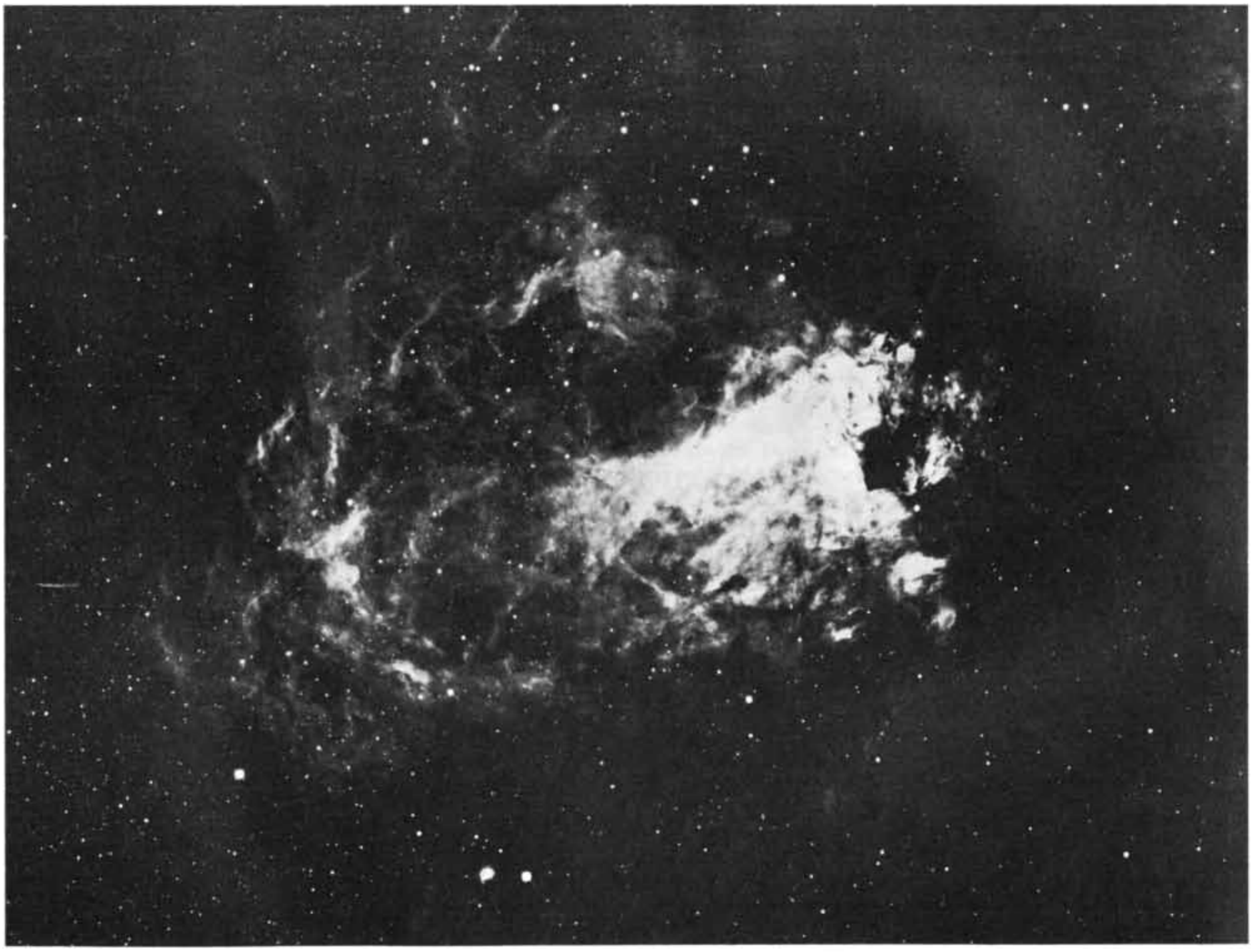
M17 is not the only place in the sky where one can view the sequence. In the constellation Cassiopeia lie two large H II regions designated IC 1795 and IC 1805. Both emit centimeter radio waves, and their radio counterparts are respectively denoted W3 and W4. In visible-wavelength photographs W4, which lies to the east of W3, is the more extended and diffuse of the two H II regions. This observation indicates that W4 must be the older H II region. Near the center of W4 is a cluster of at least 20 Type O and Type B stars; radio observations of W4 indicate that few if any Type O and Type B stars are still hidden by dust.

W3 appears to be more compact than W4, so that it must be the younger of the two regions. Radio observations of W3 show that it is located at the western edge of the visible H II region of IC 1795. Moreover, W3 actually consists of a number of compact radio components. The largest of them has been designated W3 A. One of the smallest and weakest components coincides with a source of radiation from hydroxyl molecules, which is designated W3 OH.

C. G. Wynn-Williams, Eric E. Becklin and Gerry Neugebauer of the California Institute of Technology have examined the W3 region with the five-meter (200-inch) Hale telescope on Palomar Mountain. They have found that all the radio components are emitting near-infrared radiation and that the radiation is particularly strong from W3 A and W3 OH. To complement those observations Kleinmann, Wright, Giovanni G. Fazio, Robert W. Noyes, Frank Low and I observed the W3 region at far-infrared wavelengths with the one-meter balloon-borne telescope of the Center for Astrophysics and the University of Arizona. We found that W3 A and W3 OH were also the strongest sources of far-infrared radiation.

What do all those observations mean? Speaking generally, it appears that W3 is a region actively forming massive stars, and that some of the stars seem to have been formed very recently. Each of the specific radio objects and infrared objects is probably a massive star or stars at a different evolutionary stage. From the evidence W3 OH is the youngest source in W3; it is probably a massive star that has just been born but has not yet settled down to normal life. All its radio and infrared radiation is being emitted from a volume only a few hundred times larger than the solar system.

Stepping back, as it were, from this narrow focus on W3 one sees that W3 is part of a complex that includes the radio source W4 and the two H II regions IC 1795 and IC 1805. Just to the west of the complex, according to observations by



BIRTHPLACE OF MASSIVE STARS in M17 is shown in more detail in a black-and-white photograph (*left*) and an accompanying map (*right*). To the far left the older portion of the visible H II region is expanding into the general interstellar medium. To the right the H II region seems to end abruptly. Radio contours on the map, however, indicate that the borderline is actually the beginning of the large molecular cloud shown on the preceding page. A shock wave from the H II region pushes into the cloud. Again the numbers on the contour lines indicate temperature of carbon monoxide in degrees K. Two fragments of the cloud are warmer than the rest of the gas and dust.

Lada, Elmegreen and William DeCampli of the Center for Astrophysics and Patrick Thaddeus and Hong-Ih Cong the National Aeronautics and Space Administration's Institute for Space Studies, lies a molecular cloud some 50 parsecs in extent. This cloud exhibits three hot spots of emission from molecules of carbon monoxide. The northern one lies close to W3 A. The middle one is associated with W3 OH. The southern one is a source of both far-infrared radiation and centimeter-wavelength radio emission. The nature of this source is not clear; it may be younger than W3 OH.

Like M17, the W3 / W4 complex is a case study of the sequential formation of massive stars from a molecular cloud: star birth triggered by shock waves. From east to west across the complex one sees the oldest areas blending into the youngest ones. First, farthest to the east, is IC 1805 (associated with W4), an aged H II region that will soon dissipate into interstellar space, leaving behind a loose cluster of Type O and Type B stars. Second is the younger H II region IC 1795, which is expanding into the large molecular cloud to the west. Third, at the interface of IC 1795 and the molecular cloud, is a string of compact H II regions and compact infrared sources. Fourth and farthest to the west is the molecular cloud itself, which will die in a few tens of millions of years in giving birth to more stars.

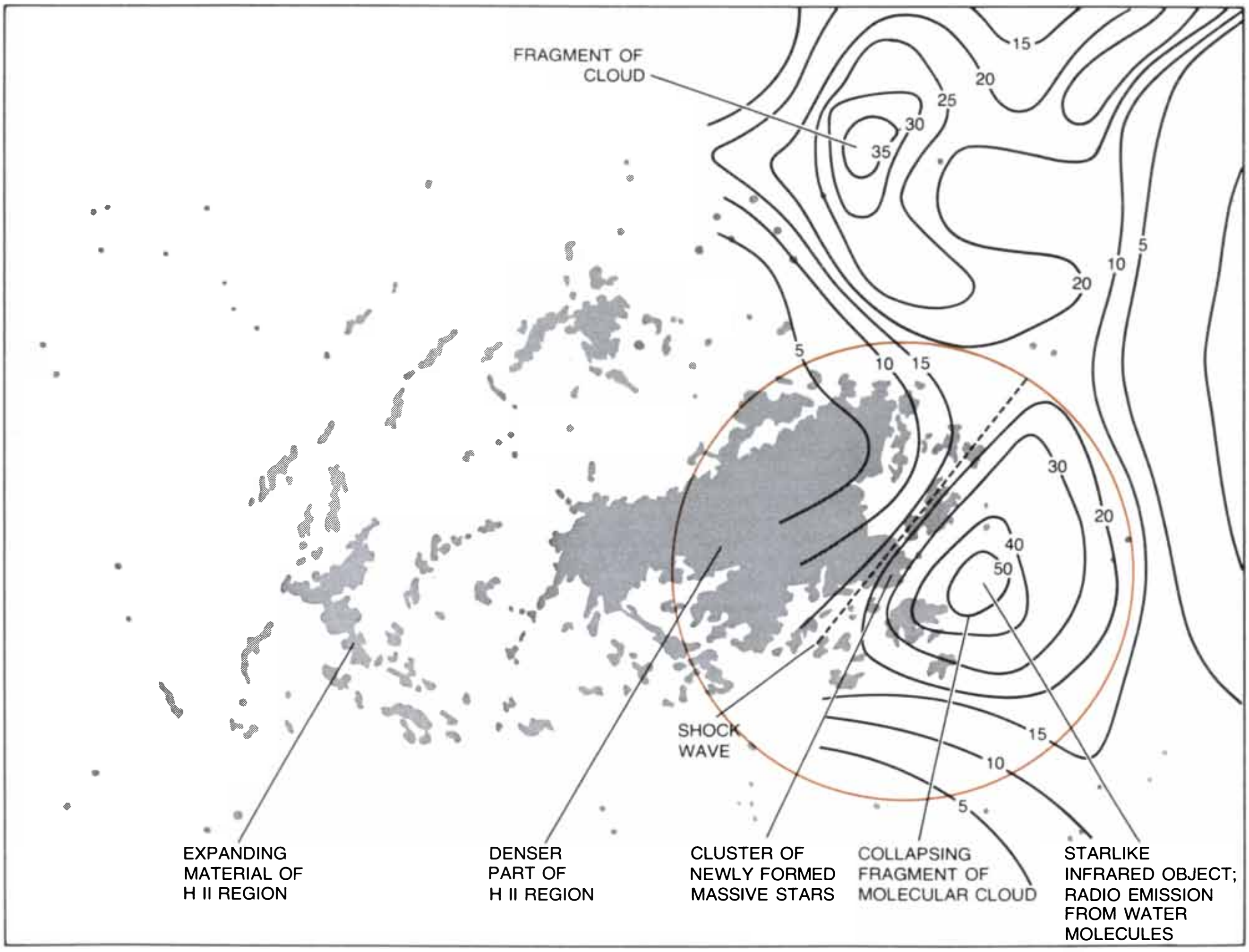
I have presented two of the best examples of evidence for the shock-driven formation of massive stars from molecular clouds. One should not think that all the observations made at millimeter wavelengths and at infrared wavelengths fit into the picture. Those that do fit, however, convince me that the model is a good one. It is a simple concept, and that simplicity is appealing.

Once the formation of massive stars begins the stars consume the stuff of molecular clouds and dissipate the material that does not end up as stars. But what starts the formation of the stars in the first place? What process introduces the initial gravitational instability? The observational evidence to date implies that most molecular clouds by themselves are not gravitationally unstable. Some outside force must disturb them. What are the possibilities?

One is that a molecular cloud is disturbed by a shock wave from the explosion of a supernova. When a massive star dies, it blasts material at high speed into the interstellar medium and creates a shock wave that as it continues to propagate can last for about a million years. If such shock waves drive molecular clouds into an initial gravitational instability, then the death of a massive star of one generation would be directly connected to the birth of massive stars of the next generation.

A second possibility is that two molecular clouds might collide. Lyman Spitzer of Princeton University has calculated, however, that the frequency of such collisions should be low, less than one collision in the galaxy every 10 million years.

A third possibility is that a molecular



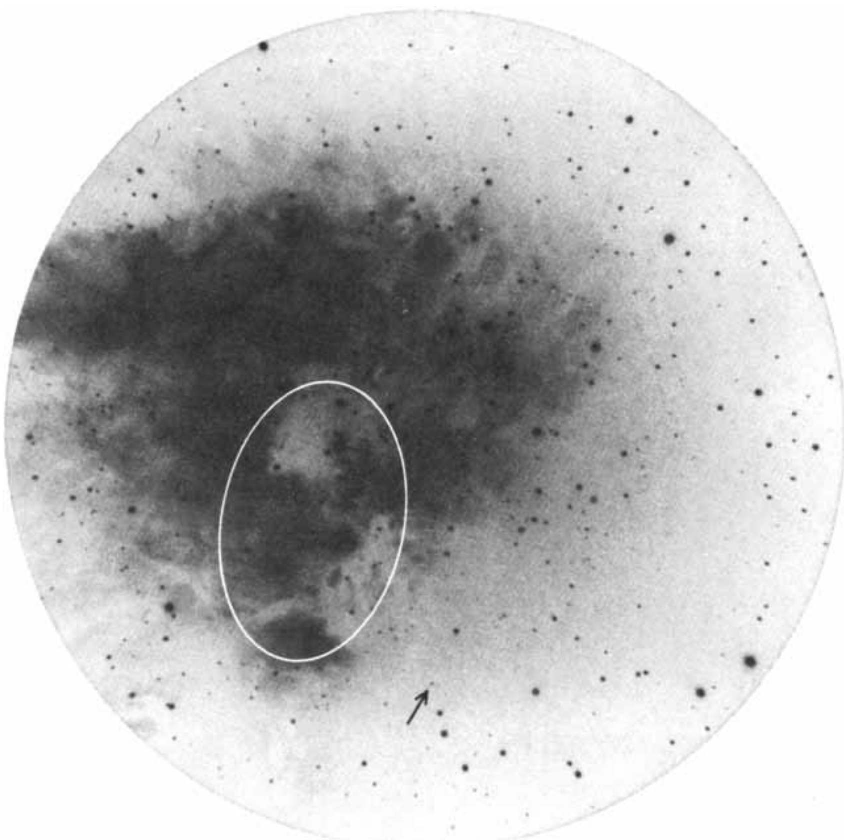
Moreover, the southern one has emission from molecules of water and hydroxyl, and in the middle of it is an intense starlike infrared source, detected by Douglas E. Kleinmann of the Center for Astrophysics and Edward L. Wright of the Massachusetts Institute of Technology. It is likely that the southern fragment of the molecular cloud is a region where gas is collapsing to form new stars. Infrared observations between the H II region and the collapsing portion of the molecular cloud indicate that deep inside the gas and dust is a cluster of very young massive stars. The color circle outlines the area with the stars, which are shown in the infrared photograph on the next page.

cloud might be disturbed by a shock wave from spiral density waves in the galaxy. To account for the spiral structure of our galaxy some theoreticians propose that two spiral density waves, which in effect are sound waves, travel through the disk in which the spiral arms of the galaxy lie. As the density waves plow through the interstellar medium they pile up material and create shock waves that trigger the formation of stars. At the present time none of these three possibilities has a distinct advantage.

I have described possible processes for the birth of massive stars. What about the birth of smaller stars such as the sun, stars that are more common in our galaxy? I believe that such stars also form by gravitational instability but that the instability is promoted in a different way. The exact process is still unclear. Perhaps stars such as the sun develop from the "globules" of cool material that can be seen in some H II regions and also by themselves in interstellar space. Radio observations suggest that at least some of those globules are contracting [see "Bok Globules," by Robert L. Dickman; SCIENTIFIC AMERICAN, June, 1977]. The globules may be remnants of a dissipated large molecular cloud or may be small interstellar clouds that never gathered enough material to grow into the large molecular clouds in which massive stars are born. How stars such as the sun are formed remains one of the most vexing problems of contemporary astrophysics.

Finally there is the most fundamental question of all: How do giant molecular clouds form in the first place? We cannot fully explain how stars are born until we know how their birthplaces are created.

The shock-driven model for the formation of massive stars from giant molecular clouds I have presented here makes a good case history of how technological advances can promote discoveries and theories. Thirteen years ago Blaauw suggested that massive stars form in sequences of subgroups. He could not, however, peer into the celestial birthplaces to see stars in the making because he had no instruments for observing at millimeter wavelengths or in the infrared. Today we have tools that have enlarged our vision and have helped to sharpen our perception of the grand scheme of cosmic evolution.



NEWBORN STARS deep in the heart of M17 are revealed in this photograph made at the infrared wavelength of 9,200 angstroms by M. Beetz and his colleagues, working with the 123-centimeter telescope of the Centro Astronomico Hispano–Aleman Almeria on the Calar Alto in Spain. The objects within the portion of nebulosity enclosed by the ellipse appear to be massive stars that may be only a few tens of thousands of years old. Southwest of the cluster is starlike infrared source discovered by Kleinmann and Wright (*arrow*) in southern fragment of molecular cloud. Photograph is circular because telescope has a very small circular field of view.

Cutlass Salon takes on Audi Fox and VW Dasher.

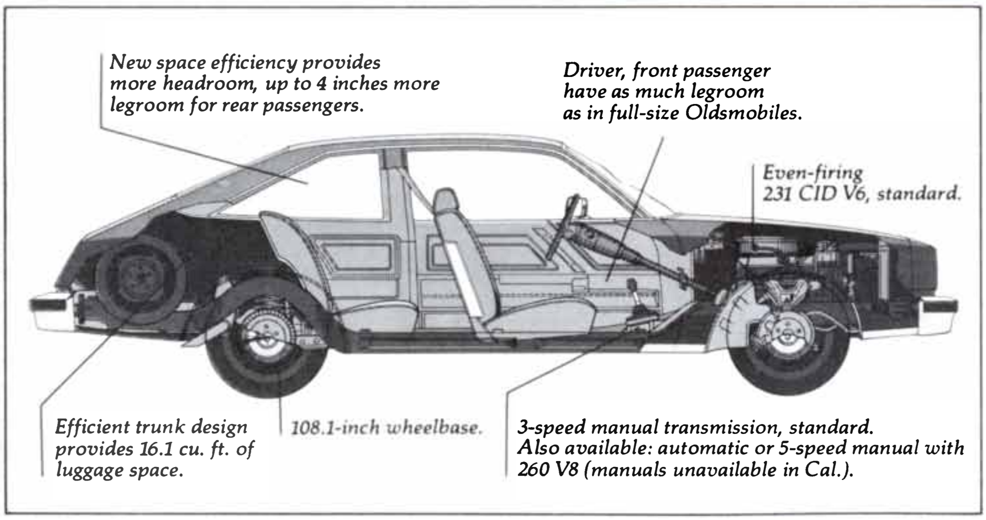
Noticed what's happened to the price of imports today? It stands to reason that if a foreign car is going to be priced higher than an Oldsmobile Cutlass Salon, you might expect more size and room for your money. Yet, Salon gives you more headroom, legroom and shoulder room than these imports. Plus the roomy comfort, trunk space and fuel economy you need. And, Oldsmobile engineering.

	Price	Economy		Room and Comfort						Trunk
		Power Train	EPA Mileage† Hwy.-City Combined	Front Head-Room† (in.)	Front Leg-Room† (in.)	Front Shoulder Room† (in.)	Rear Head-Room† (in.)	Rear Leg-Room† (in.)	Rear Shoulder Room† (in.)	Luggage Capacity† (cu. ft.)
Cutlass Salon Coupe	$4715*	231 CID 6 cyl., automatic*	27-19-22	37.9	42.8	56.8	38.2	35.1	55.7	16.1
Audi Fox 849	$6170*	97 CID 4 cyl., automatic*	29-20-23	37.6	40.9	53.9	37.3	31.7	53.3	10.7
VW Dasher 3243/83	$6024*	97 CID 4 cyl., automatic*	29-20-23	37.6	40.9	53.9	37.0	30.5	53.3	18.4

*Manufacturers' suggested retail prices for models shown equipped with automatic transmission, including dealer prep; specific features will vary, car to car. Cal. MSRP are: Salon, $4790; Fox, $6280; Dasher, $6109. Taxes, license, destination charges and other available equipment additional. Salons are equipped with GM-built engines from various divisions. See your dealer for details.

†Source: 1978 EPA data. Mileage figures are from the EPA Federal Buyer's Guide and are estimates: your mileage will vary with how and where you drive, your car's condition and equipment. Cal. EPA highway-city-combined estimates are: Salon, 23-16-18; Fox, 30-22-25; Dasher, 30-22-25.

CUTLASS SALON



Oldsmobile

Can we build one for you?

There's a lot of News in Olds today.

Kimberlite Pipes

These remarkable fossil volcanoes rise from a great depth. They are the ultimate source of diamonds and also of rocks that may be specimens of materials from the earth's mantle

by Keith G. Cox

Living on the surface of the earth, geologists have little direct knowledge of the planet's interior. Of the three broad layers that make up the earth's structure—the crust, the mantle and the core—only the crust is accessible, and even in its thickest regions the crust represents only about 1 percent of the earth's radius. Certain physical characteristics of the deeper layers, such as their average density and the speed with which they transmit earthquake waves, can be deduced from the surface. For studies of chemical composition, however, there is no adequate substitute for a specimen of mantle material.

An extraordinary source of such specimens is the rare rock type called kimberlite. Kimberlite formations generally take the form of small vertical shafts, called pipes, which are demonstrably of volcanic origin. The pipes have been studied extensively, in large part because they are of economic importance: they are the ultimate source of natural diamonds. For the geologist, however, kimberlite pipes supply gems of a different kind: rocks brought up from a great depth. Some of these rocks may be samples of material characteristic of that found in the upper portions of the earth's mantle.

Until about 100 years ago the only known deposits of diamonds were in river gravels. In 1870, however, alluvial diamond deposits in southern Africa were traced to their source, the kimberlite pipes of Jagersfontein and Dutoitspan. The pipes were near a town that is now the South African city of Kimberley, and the characteristic rock type in which the diamonds are found was named for the town.

Several other pipes have since been discovered at Kimberley, and isolated pipes and small groups of pipes are scattered in other parts of southern Africa. There is a group of 17 pipes in Lesotho, the small country surrounded entirely by South Africa, and others are known in Botswana, Namibia, Angola and, to the north, in Tanzania. Elsewhere in the world the only comparable concentration of kimberlite deposits is in the Yakutsk Republic in Siberia; the kimberlites were discovered there only in 1954. In North America there is a concentration of pipes along the border between Colorado and Wyoming, and others are known in Montana and in the Canadian Arctic. Few of the American kimberlite pipes are large or economically important, although a solitary pipe at Murfreesboro, Ark., was briefly worked as a diamond mine.

Compared with the commoner remnants of volcanic activity on the earth's surface, kimberlite pipes are quite small features. The largest have diameters at the surface of less than two kilometers, and many pipes of economic importance are only a few hundred meters in diameter. The pipes generally have the form of a cylinder or a narrow cone that tapers slightly with increasing depth. In the vicinity of the pipes kimberlite can also be found in associated formations called dykes, which are vertical slabs formed by the intrusion of molten material into fissures in the surrounding rocks. The pipes probably erupted at the surface when they were formed and were then marked by an open crater and a small cone of ejected material. In almost all cases, however, subsequent erosion has removed the surface features and the uppermost strata of both the kimberlite and the surrounding rocks. The pipes now available for study are exposed at deeper erosion levels.

Diamonds are released from the kimberlite by erosion, and they generally settle in stream beds. Subsequent geological changes may bury and consolidate these alluvial deposits, but the diamonds, being extremely durable, remain unaltered. Most of the known kimberlite pipes were emplaced in the Cretaceous period, some 70 million to 130 million years ago. Diamonds are found in alluvial deposits of several geological ages, however, indicating that there were also pipes in earlier periods. For example, there are extensive alluvial diamond fields in Brazil that are not associated with any known kimberlites. Presumably the older pipes are now hidden by younger, overlying formations. Curiously, one of the largest pipes in South Africa, the Premier, is dated at more than 1,150 million years and hence is much older than the characteristic Cretaceous pipes of the region.

Kimberlite is a highly variable rock type. Most kimberlite exposed at the surface, called "yellow ground" by miners and prospectors, is severely weathered. At deeper levels there is a material that is better preserved called "blue ground," but only in recent years have samples of the native kimberlite become readily available. Fresh kimberlite is a hard, dark gray or blue rock whose structure gives unmistakable evidence of an igneous origin. The kimberlite was extruded into its present position as a molten liquid; it was then cooled by contact with the volcanic conduit and finally solidified.

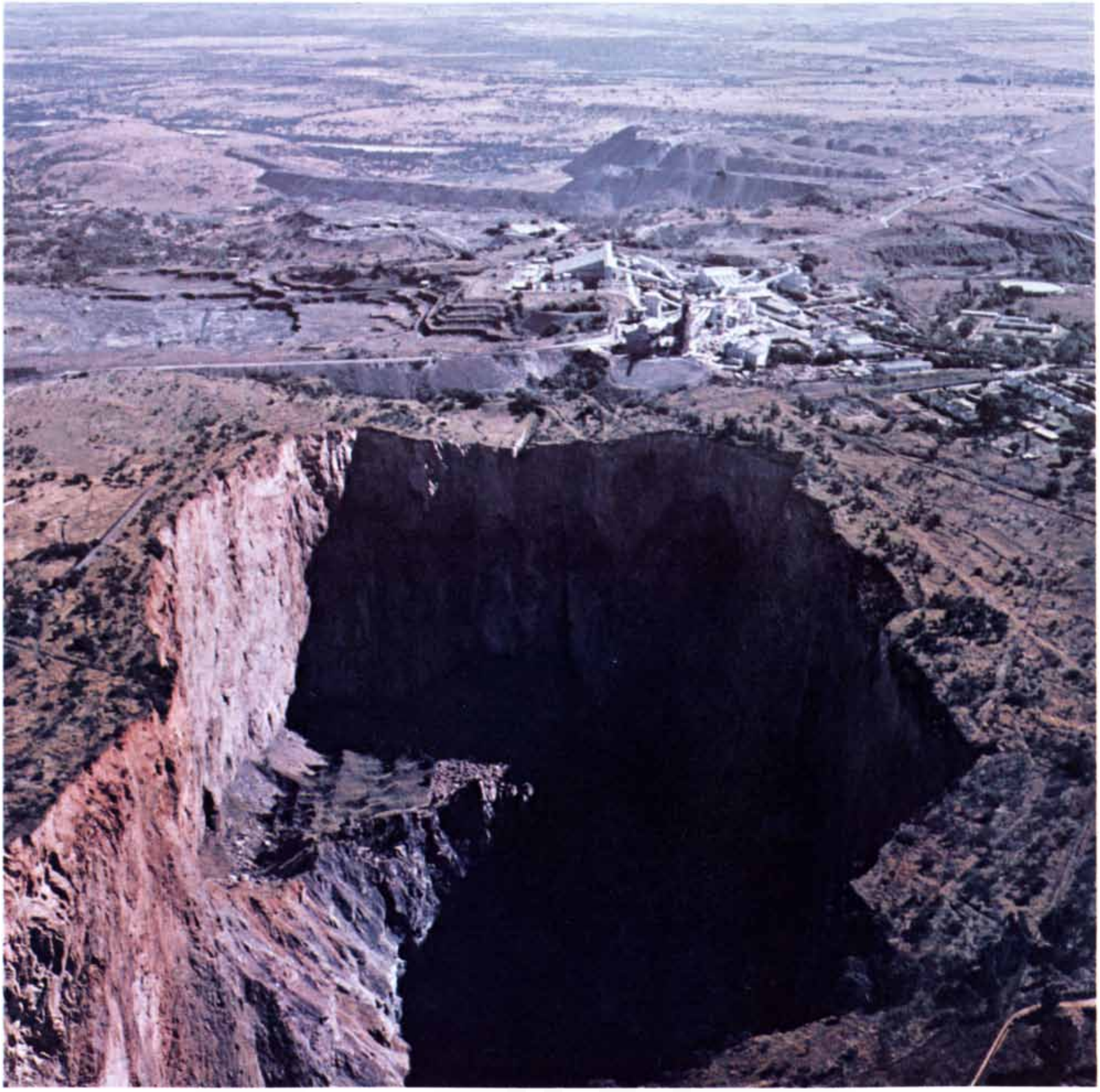
The major constituents of kimberlite are silicates, that is, compounds of silicon and oxygen with metal ions. In general, minerals cannot be defined as simple chemical compounds because their composition is not determined by a fixed ratio of atoms. Often two or more compounds are present and are said to be in solid solution with one another. As in a liquid solution, the component substances can be mixed in any ratio over a wide range. One important constituent of kimberlite is the mineral called olivine, which is a solid solution of magnesium silicate (Mg_2SiO_4) and iron silicate (Fe_2SiO_4). Another silicate present is phlogopite, a kind of mica rich in potassium and magnesium, and there are also various silicate minerals that are classified as serpentines. The serpentines are formed by the hydration of olivine, or in other words by chemically adding water to it. Kimberlite also contains the mineral calcite, which is not a silicate but consists of more or less pure calcium carbonate ($CaCO_3$).

Of the materials found in kimberlite pipes kimberlite itself may be less interesting than some of the foreign bodies that appear as inclusions within the kimberlite matrix. Among these inclusions, of course, are diamonds, and it is to their presence that we owe much of our knowledge of these remarkable volcanoes. Without the economic incentive for exploration and mining that diamonds provide, few samples of kimberlitic rocks would be available for study.

Another kind of inclusion in kimberlite, and one that is far commoner than diamond, consists of rocks torn loose from the walls of the volcanic pipe. As the molten kimberlite approaches the surface and the pressure on it is thereby reduced, dissolved gases (mainly water vapor and carbon dioxide) are forced out of solution. As a result the volcano erupts explosively. The sides of the conduit are abraded, an effect that is intensified near the surface, so that the pipe grows wider near the mouth. Some of the wall rock is probably ejected during the eruption, but much of it is shattered, ground up and incorporated into the kimberlite. These inclusions are called xenoliths (from the Greek for foreign rocks).

Because the xenoliths derive from rock strata that can be observed in the terrain surrounding the pipe, their depth of origin can be determined. Many of them fell into the pipe and are found hundreds of meters below the equivalent rocks in the pipe walls. Xenoliths found in blind offshoots of the pipe (that is, in tubes that did not reach the surface) indicate that some fragments fell down and then were carried upward again.

Perhaps the greatest scientific interest in kimberlites derives from a third kind of inclusion: the rocks called ultramafic nodules. Like diamonds, the ultramafic nodules are rare. Also like diamonds, they are thought to come up from a great depth, perhaps as much as 250 kil-



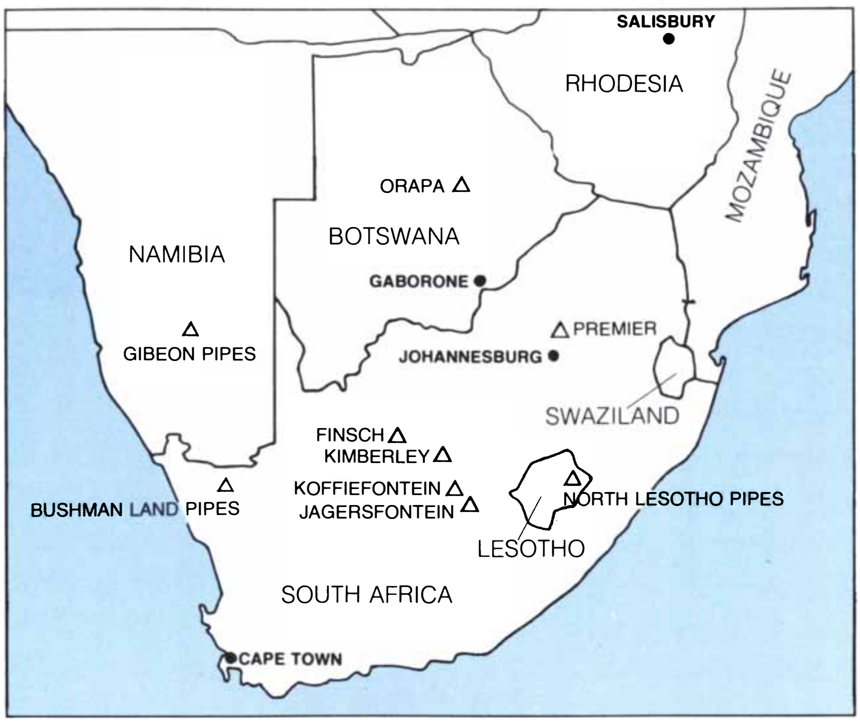
KIMBERLITE PIPE in South Africa has been extensively excavated by open-cast mining for diamonds. The kimberlite deposit itself has been removed to a depth of a few hundred meters, so that the form of the original deposit is revealed by the size and shape of the cavity. Compared with commoner types of volcano, kimberlite pipes are quite small geological features; they generally range from a few hundred meters to two kilometers in diameter. In the vertical dimension, however, these narrow shafts extend through the earth's crust and into the upper mantle and they may have carried rocks from that region to the surface. Pipe shown is the Premier, near Johannesburg.

ometers below the surface. They have a characteristic rounded form, like beach stones, caused by abrasion in the pipe.

An important question is why ultramafic nodules appear in kimberlite pipes when they are almost never seen in other volcanoes, some of which may also originate deep within the mantle. One hypothesis is that the kimberlites rise through the earth much faster than the magmas of other volcanoes. If the fluid rock behaves like an ordinary liquid, such as water, then a rapid ascent would be required to transport the largest of the nodules. Such liquids are said to have zero shear strength: one part of the fluid can move freely past another part. Hence a solid body immersed in the fluid can be carried upward only if the force of drag created by the moving fluid exceeds the weight of the nodule. Calculations based on the weight of the largest nodules suggest that they may have been carried to the surface in a period of hours or at most a few days.

An alternative to the rapid-ascent hypothesis has recently emerged. R. S. J. Sparks, H. Pinkerton and R. MacDonald of the University of Manchester have proposed that under some circumstances molten silicates may not behave like ordinary fluids; they may have a shear strength greater than zero. The ultramafic nodules might then be transported not by the force of drag but as fixed inclusions in the fluid matrix. A peculiar property of such fluids is that they retain their shear strength only when they are flowing at low speed, and so the alternative to a rapid ascent is an exceptionally slow one.

ULTRAMAFIC NODULES are rare inclusions in kimberlite pipes that seem to have been transported in the solid state from the upper mantle. Several small, broken nodules are embedded in the kimberlite matrix of the rock at left; the rounded boulder in the center is a single large nodule. The smooth, rounded forms characteristic of the nodules are produced by abrasion during transport. The term "ultramafic" describes rocks made up mainly of magnesium and iron silicates. The commonest class of nodules, which includes the ones shown here, is composed of the rock type called peridotite, which is thought to be a major constituent of the mantle.



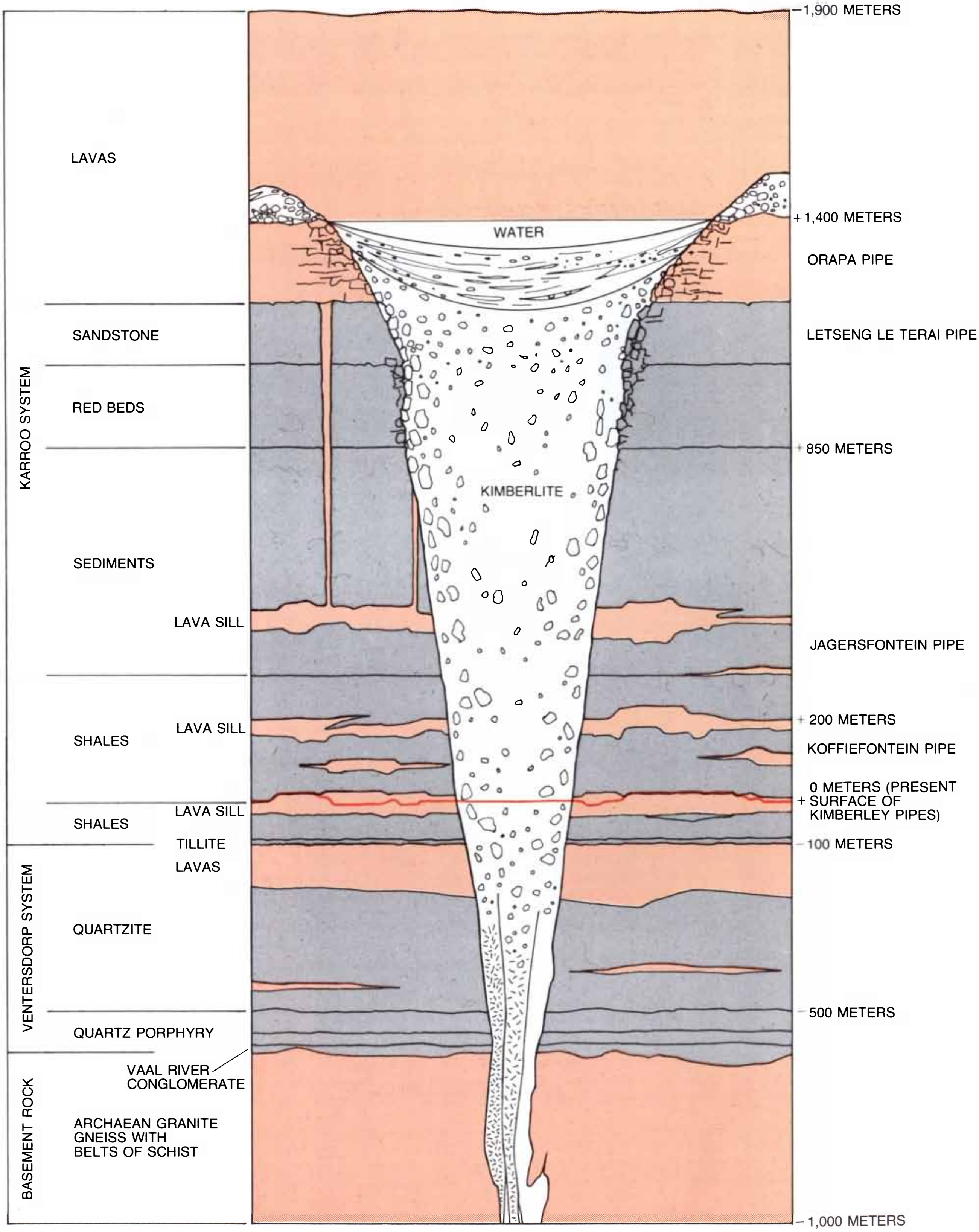
ABUNDANCE OF KIMBERLITE PIPES is greater in southern Africa than in any other region of the world. At most of the locations marked there is not just one pipe but a cluster of them; the North Lesotho group, for example, includes 17 pipes. The only concentration of pipes comparable to the one in southern Africa is in Siberia. Perhaps significantly, both southern Africa and Siberia were sites of extensive volcanic activity before the pipes were formed.

Physical conditions inside the earth and certain physical properties of materials found there can be deduced from surface measurements. For example, since the mass and the angular momentum of the earth are known, the distribution of density can be calculated. From that information the internal pressure can be determined as a function of depth. It is assumed that the pressure is hydrostatic, or in other words that it is exerted equally in all directions by material that is not itself compressible. It can then be shown that the pressure increases by about one kilobar, or 1,000 times atmospheric pressure at sea level, for every three kilometers of depth. Internal temperatures can be estimated by measuring the flow of heat near the surface. Temperature increases with depth everywhere, but the increase is slower under continents than it is under the sea floor.

The most revealing view of the earth's interior comes from the observation of earthquake waves propagated over long distances. The refraction of these waves at various depths reveals sudden changes in their speed. Those changes may correspond to discontinuities in other properties as well.

The division of the earth into crust, mantle and core is based on such seismic observations. The crust is some 10 kilometers thick under the oceans, and it ranges from 35 to 70 kilometers in thickness under the continents. The mantle extends from the base of the crust to a depth of about 2,900 kilometers, and the core extends to the center of the earth at a depth of 6,370 kilometers. The mantle can be further divided into upper and lower layers at a depth of about 700 kilometers.

Seismic observations have revealed



MODEL OF A KIMBERLITE PIPE is based on the structures of several pipes exposed at various erosion levels. The pipes erupted in the Cretaceous period, some 70 to 130 million years ago, when the strata that are exposed at Kimberley today were 1,400 meters below the surface. The geological structure of the region consists of an alternating sequence of igneous rocks (*color*) and sedimentary rocks (*gray*); both the igneous and sedimentary layers include a variety of rock types. The eruption of the kimberlite probably produced a shallow crater, surrounded by a low cone of ejected material. Abrasion of the wall rocks by the ascending kimberlites, which became more intense near the surface, gave the pipe the form of a narrow cone. Both fragments of the wall rock and the much rarer ultramafic nodules are incorporated into the kimberlite. Several African pipes (*noted at right*) are exposed at higher levels than those near Kimberley, the South African town that is the namesake of the rock type. The model was devised by J. B. Hawthorne of De Beers Consolidated Mines Ltd.

another discontinuity within the upper mantle, which serves as the basis for an independent and complementary interpretation of the earth's structure. The material above this region, including all of the crust and a part of the upper mantle, is the lithosphere. It consists of rigid plates, several of which are in relative motion. The zone below is the asthenosphere; it is distinguished by lower viscosity, and it may be subject to partial melting. The plates of the lithosphere ride on and slide across the comparatively fluid asthenosphere.

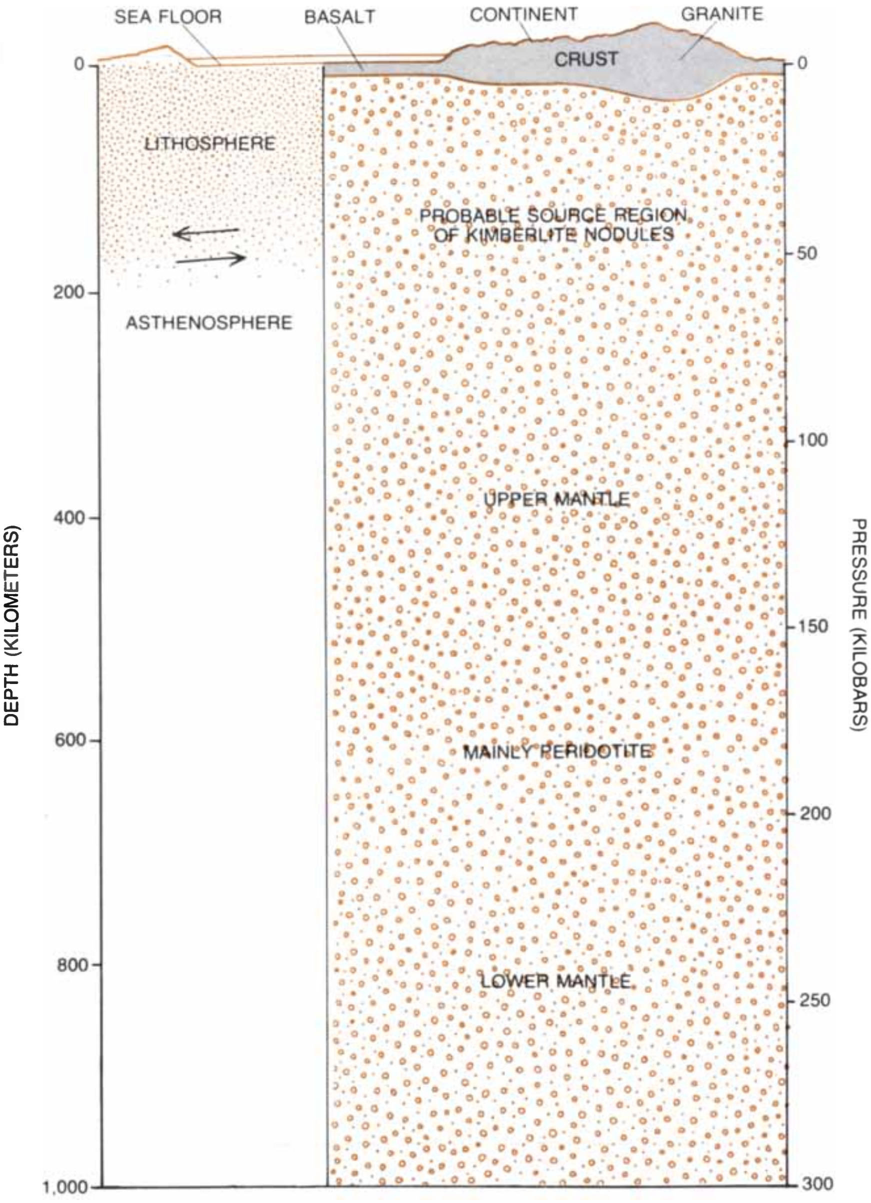
Some of the physical discontinuities deduced from seismic observations surely correspond to boundaries of chemical composition. In the continental crust the dominant constituents are granites and related rock types. They are rich in silicon, aluminum, sodium and potassium, and are comparatively poor in magnesium and iron. The oceanic crust is made up largely of basalt, which has more magnesium and iron than granite, but is poorer in sodium and potassium. Slices of oceanic crust are occasionally thrust upward in collisions between lithospheric plates, and even more occasionally they carry with them a layer of material from the topmost few kilometers of the mantle. The mantle rocks exposed by such upthrust plates are predominantly of the type called peridotite. The peridotites have a composition that is close to being the inverse of the granites in the continental crust: they are very rich in magnesium but are slightly depleted of silicon and are considerably depleted of aluminum, sodium and potassium.

The best available information on the composition of the earth as a whole comes not from terrestrial rocks but from meteorites. The most primitive meteorites, the chondrites, have an unusual dropletlike texture. They are thought to be unaltered specimens of the material from which the solar system condensed. The chondrites consist of a silicate rock similar to peridotite in overall composition, together with metallic iron that contains a substantial amount of nickel in solution. If the average compositions of the chondrites and the earth are similar, then the mantle must be mainly peridotite and the core must be iron and nickel. (The crust makes up such a small part of the earth's bulk that it can be neglected.)

The ultramafic nodules of kimberlite pipes provide important evidence corroborating this hypothesis. Unlike meteorites they are actual samples of terrestrial material, and hence their interpretation does not depend on the adoption of a particular model for the evolution of the earth. Unlike the mantle material associated with upthrust plates they are from deep within the mantle and not from the boundary layer where the mantle meets the crust.

"Mafic" is a term applied to minerals that contain large amounts of the silicates of magnesium and iron; in ultramafic rocks these minerals are the dominant constituents. Most of the ultramafic nodules found in kimberlite are made up of peridotite; peridotite in turn is composed mainly of olivine and another silicate mineral, pyroxene.

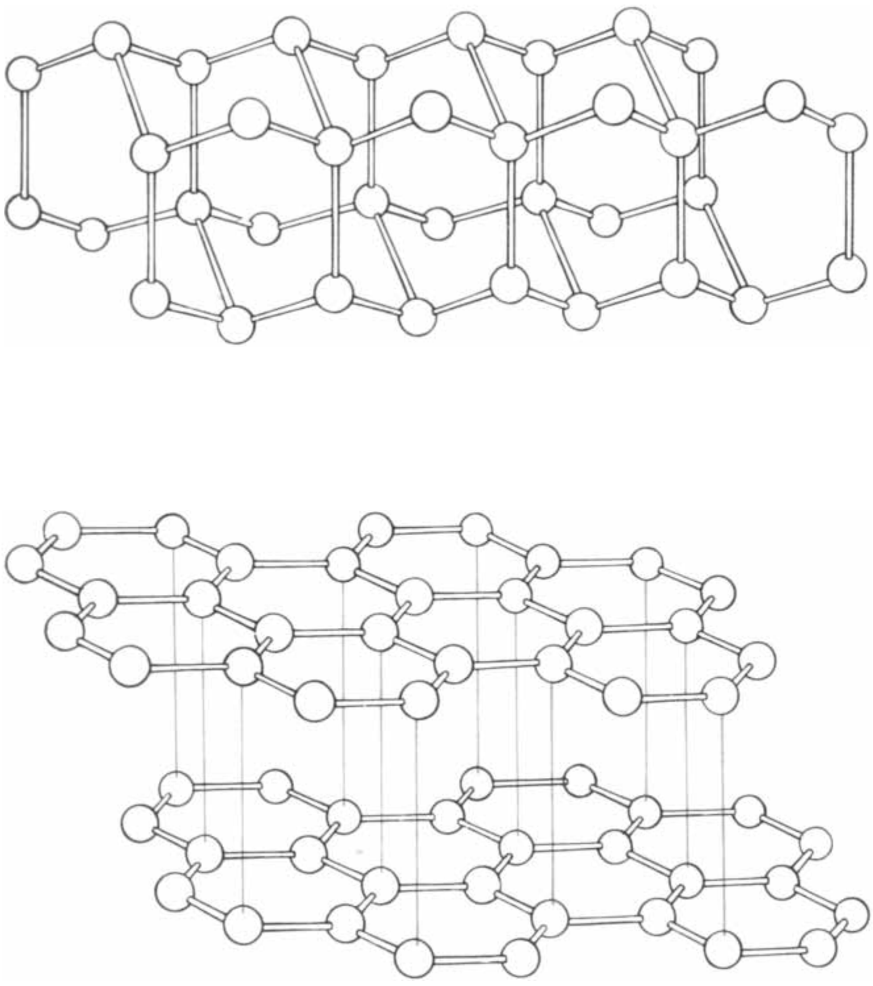
The olivine in peridotite is rich in magnesium, so that in its composition Mg_2SiO_4 dominates over Fe_2SiO_4. The nodules contain two kinds of pyroxene, which can be distinguished not only by their composition but also by the symmetry systems of their crystals. One is an orthopyroxene and forms orthorhombic crystals; the other is a clinopyroxene and forms monoclinic crystals. The orthopyroxene, which is the most abundant mineral after olivine, consists of a magnesium silicate ($MgSiO_3$) with some iron silicate ($FeSiO_3$) in solid solution. The clinopyroxene is chrome diop-



STRUCTURE OF THE EARTH has been determined largely by seismic observations. The boundary between the crust and the mantle, which is at a depth of 10 kilometers under the oceans and from 35 to 70 kilometers under the continents, is marked by a discontinuity in the speed of earthquake waves. It probably also represents a boundary between regions of different chemical composition. The oceanic crust is mainly basalt and the continental crust mainly granite, whereas the mantle is apparently peridotite. Only the upper 1,000 kilometers of the earth's structure is shown; the mantle continues to a depth of 2,900 kilometers and the core extends from that level to the center of the earth at 6,370 kilometers. A seismic discontinuity at a depth of about 700 kilometers divides the mantle into upper and lower regions, but it probably represents only a change in crystal structure, not in chemical composition. A zone of low earthquake-wave velocities is the basis for another system of layers (*left*). Material above this zone is the rigid lithosphere, which slides over the more plastic asthenosphere. Ultramafic nodules in kimberlite are thought to have come from the mantle at depths of from 100 to 300 kilometers.

	NODULE COMPONENT	APPROXIMATE COMPOSITION	CRYSTAL SYSTEM
PERIDOTITE	OLIVINE	Mg_2SiO_4 WITH Fe_2SiO_4	ORTHORHOMBIC
PERIDOTITE, PYROXENE	ORTHOPYROXENE	$MgSiO_3$ WITH $FeSiO_3$	ORTHORHOMBIC
PERIDOTITE, PYROXENE	CLINOPYROXENE (CHROME DIOPSIDE)	$CaMgSi_2O_6$ WITH Fe AND Cr	MONOCLINIC
PERIDOTITE, ECLOGITE	GARNET (PYROPE)	$Mg_3Al_2Si_3O_{12}$ WITH Fe	CUBIC
ECLOGITE	CLINOPYROXENE (OMPHACITE)	$CaMgSi_2O_6$ WITH Fe, Na AND Al	MONOCLINIC

MINERAL CONSTITUENTS of ultramafic nodules are mainly silicates, or compounds of silicon and oxygen with metal ions. In the commoner peridotite nodules the major constituents are olivine and orthopyroxene; most peridotites also contain small quantities of clinopyroxene and garnet, but those components are sometimes lacking. Another nodule type, eclogite, is made up of garnet and a clinopyroxene called omphacite. A distinguishing feature of nodule minerals is that they are much richer in magnesium than rocks typical of the earth's crust.



DIAMOND AND GRAPHITE are both crystalline forms of carbon, but they are stable at different combinations of pressure and temperature. Graphite, the low-pressure form, consists of stacked planes of hexagonal rings. Diamond, which can be formed only at high pressure, has a more symmetrical structure in which each carbon atom is surrounded by four others. The presence of diamond in kimberlite gives a clue to the conditions under which the rock was formed.

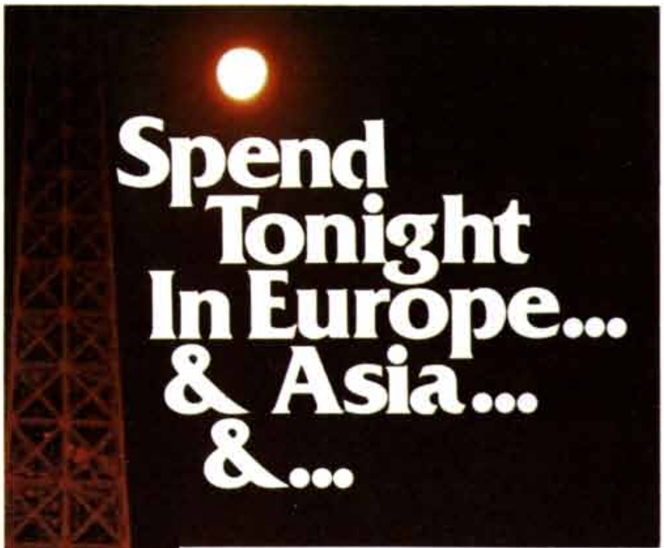
side, a bright green mineral that contains substantial amounts of calcium. Its elemental composition is approximately $CaMgSi_2O_6$, but it also contains some iron and chromium. Many peridotites also have small quantities of pyrope, a red-to-violet variety of garnet. The approximate composition of pyrope is $Mg_3Al_2Si_3O_{12}$, but it too has some iron in solution. Olivine and orthopyroxene are invariably the major constituents of peridotitic nodules. Chrome diopside and pyrope are usually present, although only in small quantities, but in some nodules they are absent entirely.

Although the peridotitic nodules are by far the most abundant type in most kimberlite pipes, other kinds of nodule are known. One of the most interesting is eclogite, which is composed largely of garnet and a calcium-rich pyroxene called omphacite. Eclogites have several intriguing properties, not the least of which is that they occasionally contain diamonds. (Most diamonds, however, are found not in eclogites but as single crystals within the kimberlite matrix.)

The predominance of peridotite in the ultramafic nodules lends strong support to the hypothesis that peridotite is the major constituent of the mantle. As I shall show below, the ultramafic nodules seem to have been brought up from a depth of 100 kilometers or more. Since there are no discontinuities in the physical properties of the mantle down to 700 kilometers, it seems reasonable to conclude that the entire upper mantle consists mainly of peridotite. In fact, it is likely that the lower mantle is also peridotite. The refractive zone at 700 kilometers probably marks not a change in composition but a mineralogical transition where high pressure alters the crystal structure of the peridotitic materials.

The depth at which an igneous rock was formed can sometimes be estimated from knowledge of the minerals it contains. The conditions required to create the mineral, and particularly the temperature and pressure, can be determined in the laboratory. Geophysical measurements then provide an estimate of where in the earth those conditions prevail. In kimberlites diamond provides an upper limit to the depth of formation and various forms of silica give a lower limit.

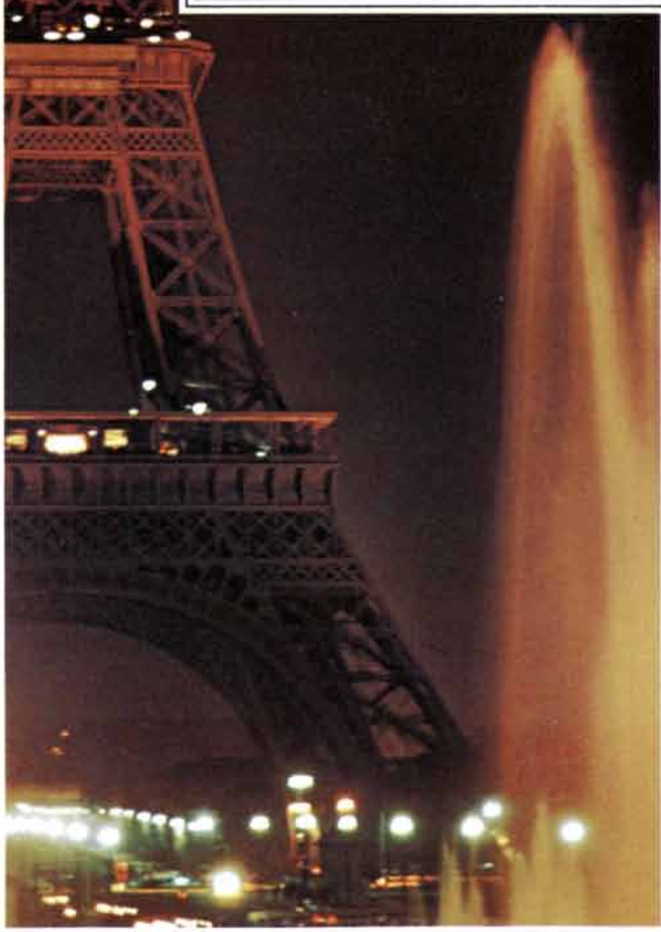
Diamond is a crystalline form of carbon that can be created only at high temperature and pressure. Under milder conditions the favored crystalline form of carbon is graphite. The set of all combinations of temperature and pressure where the two forms are in equilibrium is called the diamond-graphite inversion curve. A specimen of carbon at any point on this curve could assume either crystalline form; increasing the pressure or decreasing the temperature would favor the creation of diamond.



just about any other place on earth.

Wouldn't it be fascinating to visit each of the modern 'Seven Wonders of the World' all in one night from the comfort of your own home? With the new McKay Dymek DR33 Professional Receiver, you can take a trip to all Seven Wonders in one night and do much, much more.

To initiate your visit to the first Wonder, the Suez Canal, just tune in any Middle-Eastern shortwave broadcast. This is accomplished effortlessly with the DR33's solid state, phased locked, digital synthesis tuning system. We call it the 'professional' receiver, but any non-technical



person can easily operate it and tune in all wave frequencies throughout the world, including messages from the second Wonder, the Dneporges Dam in Russia.

As you're sitting in your favorite chair at home, the 'Professional' receiver is rapidly picking up overseas phone calls or even Radio Free Europe Broadcasts. Hear the latest-breaking international news direct from Moscow, Washington and London.

And while you're in England, let the DR33 show you the third Wonder, the Atomic Energy Research Establishment in Harwell. Imagine, in minutes you can enjoy such diverse international transmissions as the dry wit of the British BBC, to the specialized lingo of a speeding convoy of American Ham radio operators near the fourth Wonder, the Alcan Highway in Alaska. And it's all amazingly affordable and possible with McKay's international receiver.

With its American-made quality, the DR33 takes you in seconds from the cultural center, San Francisco and the Golden Gate Bridge - the fifth Wonder - to the Eiffel Tower in Paris, Wonder number six.

This unit, with a handsome oak cabinet and six digit readout resolution, costs much less than other competing professional systems. The McKay Dymek DR33 is your ticket to any message between 50 and 29,700 kHz. It can monitor communications of international aircraft and then in a few simple seconds tune in standard AM stations being transmitted from close by the seventh and final Wonder, the Empire State Building in New York City.

There you have it. All 'Seven Wonders of the World' in less than an hour!

From your favorite, cozy chair at home, you've experienced different cultures and perspectives of people throughout the world. With your DR33, you've done more traveling in one night than most people do in a lifetime.

So spend tonight in Europe. . . & Asia . . . & wherever; spend it with the McKay Dymek DR33 Professional All Wave Receiver.

Order factory Direct. Call **toll free** today! Money Back guarantee. Rent/Own Plan available.
BankAmericard/Visa & Master Charge accepted
Nationwide 800/854-7769
California 800/472-1783

McKay Dymek Company
111 S. College Ave.
P.O. Box 5000
Claremont, CA 91711

McKay Dymek Company
111 S. College Ave., P.O. Box 5000, Claremont, CA 91711

Please send me your full color brochure and specifications on the McKay Dymek Receivers and Accessories.

Name ____________________

Address ____________________

City ________ State ______ Zip ______ Phone ________

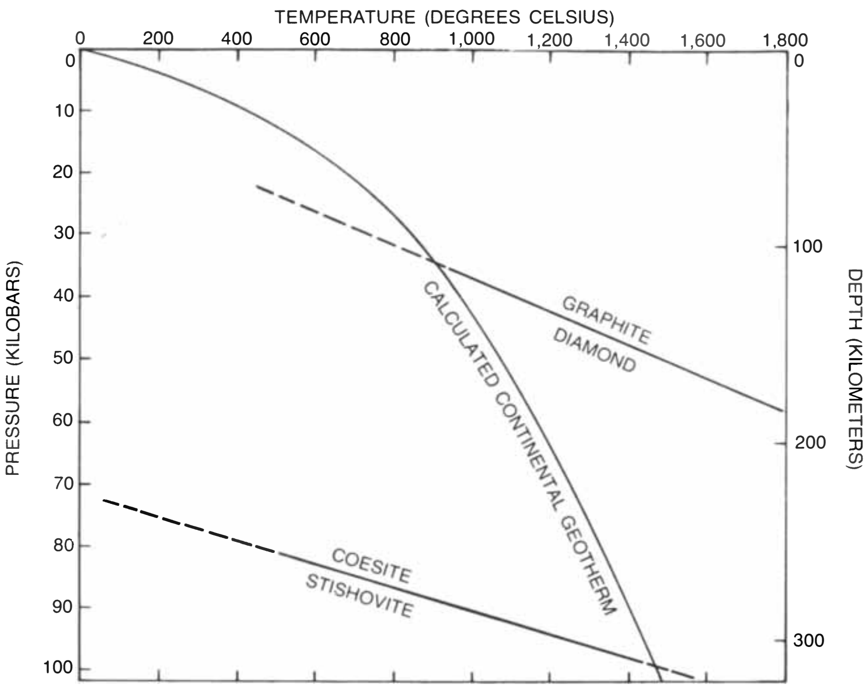

Why give less than Courvoisier?
COURVOISIER
NAPOLEON
COGNAC
COURVOISIER
COGNAC
COURVOISIER
VS
COURVOISIER
COGNAC
VSOP
Fine Champagne
COURVOISIER
The Brandy of Napoleon
COGNAC
COURVOISIER
IMPORTED BY W. A. TAYLOR & CO., MIAMI, FLORIDA, 80 PROOF, SOLE DISTRIB TORS FOR THE U.S.A. ©1978

Temperature and pressure within the earth are highly correlated. The graph of temperature as a function of pressure or depth is called the geotherm; since kimberlites are known only on land, the curve of interest is the continental geotherm. The intersection of the diamond-graphite inversion curve and the continental geotherm gives the minimum likely depth for the formation of diamonds. The curves have been found to meet at a pressure of about 35 kilobars and a temperature of roughly 800 degrees Celsius. That pressure corresponds to a depth of about 105 kilometers. At all shallower levels the pressure is too low or the temperature is too high for diamond to form. The presence of diamond in kimberlite therefore indicates that the mineral formed at a depth greater than about 100 kilometers. That depth is only a minimum, however; since diamond is stable at all pressures below the inversion curve it could be formed at any greater depth.

A maximum depth for kimberlites can be calculated from the inversion curve for two crystalline forms of another substance: silica. At low pressure silica is represented by the familiar mineral quartz. As the pressure is increased its crystal structure changes and it is converted first into coesite and finally into stishovite. Coesite is found as minute inclusions in diamond. Recently Joseph R. Smyth of the Los Alamos Scientific Laboratory and C. J. Hatton of the University of Cape Town have also found relatively large crystals of coesite in an eclogite from a South African kimberlite pipe. Stishovite, on the other hand, has never been observed in kimberlitic material. The stishovite-coesite inversion curve appears to cross the continental geotherm at a pressure of about 100 kilobars, corresponding to a depth of about 300 kilometers. Thus the presence of diamond and the absence of stishovite suggest that kimberlites originate at a depth of between 100 and 300 kilometers.

A further check on these findings can be made by comparing the melting point of peridotitic rocks at various pressures and temperatures with the continental geotherm. Since the ultramafic nodules were brought to the surface by volcanic activity, they must clearly have come from a region subject to episodes of melting. The comparison suggests that the mantle is most likely to melt at a depth of very roughly 100 to 200 kilometers, where the melting curve and the geotherm approach each other most closely. A rather modest increase in temperature at this depth would melt dry peridotite. Alternatively the addition of a comparatively small amount of water would depress the peridotite melting point enough to induce at least partial melting.

The peridotite melting curve and the presence of diamond and coesite in kimberlite suggest a broad range of depths from which all ultramafic nodules might be derived. It would also be useful to estimate the depth of origin for individual nodules. Such estimates can be made by determining the state of mineral systems in the nodule and again calibrating the measurements with laboratory observations.

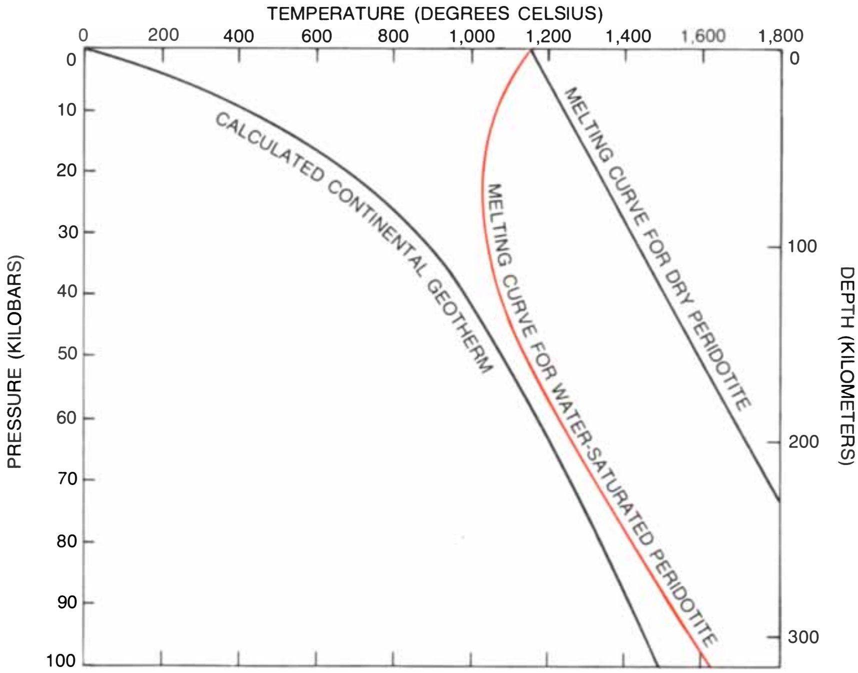


UPPER AND LOWER BOUNDS on depth of kimberlite formation are provided by forms of carbon and silica found in kimberlite pipes. The presence of diamond indicates that kimberlite forms at a depth below the diamond-graphite inversion curve, which marks the zone where these phases are in equilibrium. Similarly, kimberlite contains silica in the form of coesite but not as stishovite, and so it probably comes from a region above the stishovite-coesite inversion curve. Temperatures within the earth can be calculated from measurements of heat flow at the surface; the resulting curve for regions below continents is called the continental geotherm. Kimberlite and the nodules it contains are likely to have formed along the geotherm between those points where it intersects the diamond-graphite and stishovite-coesite inversion curves.

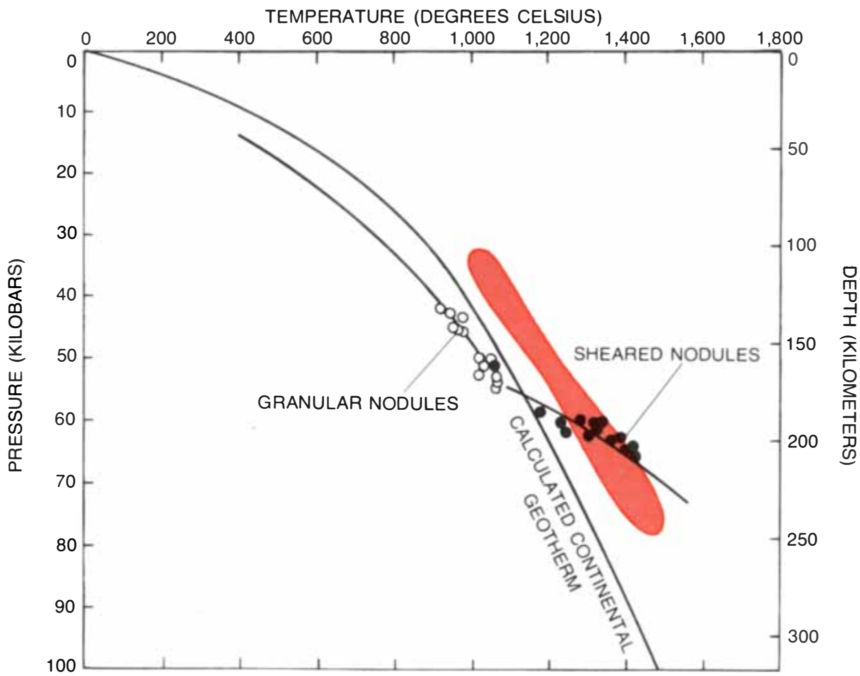
One useful indicator is the extent to which orthopyroxenes and clinopyroxenes are dissolved in each other. In 1966 B. T. C. Davis and Francis R. Boyd of the Carnegie Institution of Washington's Department of Terrestrial Magnetism showed that the mutual solubility of these two phases increases with temperature but is almost independent of pressure. The solubility is thus a potentially useful geothermometer. Pressure can be determined in a similar way by measuring the extent of reaction between coexisting orthopyroxenes and garnets. In this case the degree of reaction is dependent on both pressure and temperature, but the pressure can be deduced if the temperature is estimated first from the orthopyroxene-clinopyroxene solubility. It is important to note that the temperatures and pressures calculated in this way do not necessarily correspond to the depth where the nodules were originally formed. They give only the depth where the mineral systems were last in equilibrium. Prolonged residence at a level above or below the site of formation could "reset" both the thermometer and the pressure gauge.

The first pressure and temperature data for kimberlite nodules were calculated in 1973 by Boyd and P. H. Nixon. They found that nodules from a single pipe often exhibit a wide range of equilibrium pressures and temperatures. The data are not scattered randomly, however; pressure and temperature are correlated in such a way that the values for most nodules fall in a narrow band. The position of the band was found to vary somewhat from pipe to pipe, but on the whole it coincides with the continental geotherm predicted by heat-flow measurements. What this pattern implies is that the ascending kimberlite incorporates an essentially random sampling of the mantle material through which it passes.

One surprising discovery to emerge from Boyd and Nixon's calculations was that the texture of the nodules varies with their apparent depth of origin. In mineralogy texture refers to the spatial relations of mineral grains. In the most common peridotitic nodules the grains are about half a centimeter across; they show no apparent deformation of the crystalline structure and they



MELTING CURVE for peridotite provides further evidence bearing on the possible origin of kimberlite. Since kimberlite is a volcanic product, it must have originated in a region where at least some mantle materials are molten. The melting curve for dry peridotite (*black*) approaches the continental geotherm most closely in the expected depth range of 100 to 300 kilometers. Melting could result from episodic heating of the mantle at that depth, which would produce an upward inflection in the geotherm. Alternatively the addition of water to the mantle rocks (*color*) could depress the melting curve for peridotite enough to cause at least partial melting.



DEPTH OF ORIGIN for ultramafic nodules was estimated by calculating the temperature and pressure at which certain mineral systems in the nodules were last in equilibrium. A series of calculations made in 1973 by Francis R. Boyd and P. H. Nixon of the Carnegie Institution of Washington's Department of Terrestrial Magnetism suggested that the nodules could be divided into two series based on their texture. Granular nodules, the commoner type (*open dots*), fall along a curve near the predicted continental geotherm. Highly deformed specimens called sheared nodules (*solid dots*), however, seem to be derived from deeper levels and also suggest that temperatures there are higher than those calculated from heat-flow measurements. One explanation for this pattern is that the sheared nodules came from the asthenosphere, where the sliding of lithospheric plates could lead to both deformation and heating. Pressures and temperatures for the nodules have been recalculated by Neville L. Carter and J.-C. Mercier of the State University of New York at Stony Brook. Their results (*colored band*) confirm that the sheared nodules come from deeper strata but show no inflection in the geotherm.

give little evidence of having a preferred orientation. In other nodules, however, the structure of some crystals is altered and other crystals are ground up, changes that can be interpreted as the effects of deformation by shearing. The onset of deformation is indicated by olivine crystals, which develop striations, called kink bands, in their crystal lattice and which also become granulated at their margins. With further deformation the rock breaks up into a matrix of finely milled olivine with fragments of orthopyroxene and garnet as insets. Extremes of deformation lead to finely banded rocks in which all the minerals, with the frequent exception of garnet, are much reduced in grain size. In some specimens there is evidence of recrystallization leading to the formation of coarse-grained rocks that differ from ordinary granular peridotites in that the grains have a preferred orientation. The deformed specimens are collectively called sheared nodules.

Several pressure and temperature curves calculated by Boyd and Nixon in 1973 indicated that sheared nodules almost invariably come from a greater depth than granular nodules. Moreover, there appeared to be a sudden upward inflection of the geotherm at the depth of the sheared nodules, that is, the temperature at that depth seemed to be higher than would be expected from heat-flow measurements at the surface. There was widespread speculation that the sheared nodules might represent samples of material from the asthenosphere. The sliding of lithospheric plates over the mantle in this region could certainly lead to extreme deformation of rocks, and the resulting frictional heating might be responsible for the higher temperatures.

This intriguing proposal has since been challenged. The calculation of equilibrium pressures and temperatures has been refined, notably by B. J. Wood and S. Banno of the University of Manchester. Recalculated pressure and temperature curves for kimberlite nodules, such as those made by J.-C. Mercier and Neville L. Carter of the State University of New York at Stony Brook, show no inflection at the depth of the sheared nodules. On the other hand, the observation that sheared nodules generally come from greater depths than granular nodules has been confirmed. The depth estimates are also in general agreement with those made by other techniques. Most equilibrium pressures are in the range from 30 to 80 kilobars, corresponding to depths of roughly 100 to 250 kilometers.

Several workers have proposed alternative explanations for the origin of the sheared nodules, in which the movement of the kimberlite itself is responsible for the deformation. For example, Harry W. Green of the University of California at Davis and Y. Gueguen of

YOU HAVE TO BE FAST TO MAKE THE BIG GREEN TEAM.

At National Car Rental, we've set up our Advance Reservation System with one simple goal—to help you reserve a car fast. How fast?

In the last year we answered over three million calls. And we're answering calls in an average of nine seconds. It's extra hustle like this that makes the Big Green Team so special—all 10,000 of us.

So next time you need to reserve a car just about anywhere in the world, simply call our toll-free number 800-328-4567. Or see your travel consultant. In Minnesota and Canada call collect 612-830-2345.

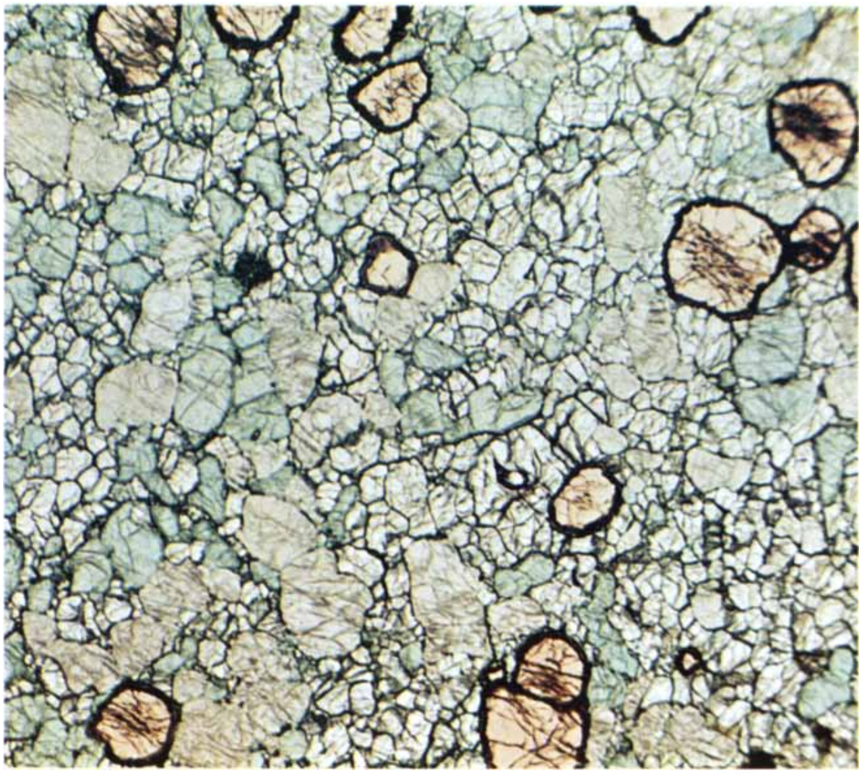
We promise not to keep you waiting.

NATIONAL CAR RENTAL

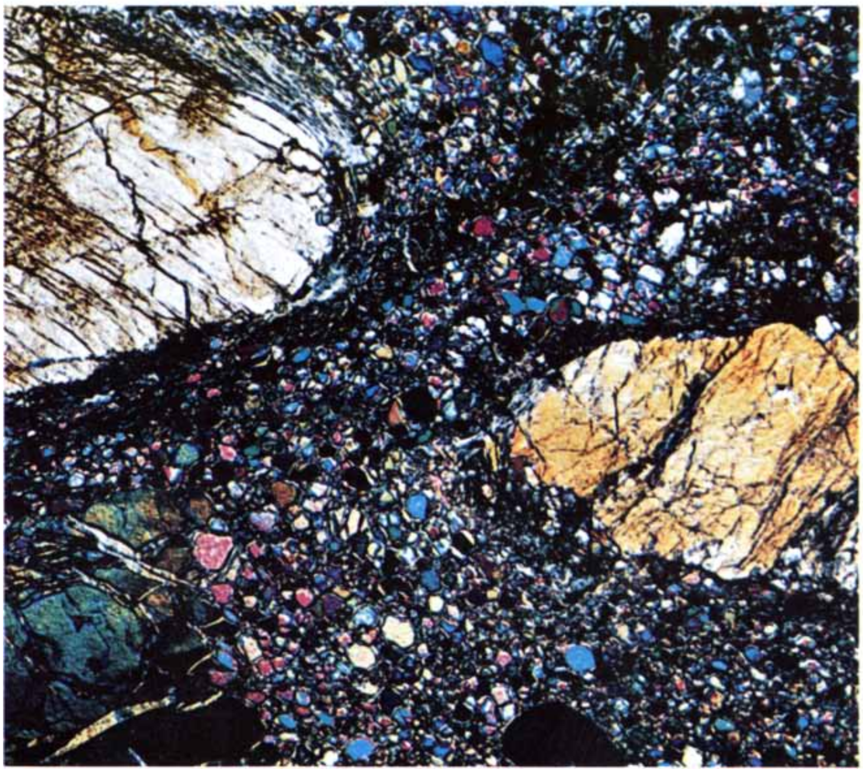
THE BIG GREEN TEAM

We feature GM cars like the Pontiac Grand Prix. We offer S&H Green Stamp Certificates on rentals in all 50 U.S. states.

© 1978, National Car Rental System, Inc. In Canada it's Tilden. In Europe, Africa and the Middle East it's Europcar.



GRANULAR NODULE from the Matsoku pipe in Lesotho reveals an undeformed structure in a micrograph of a polished section. The four major constituents of peridotite are present. The pink (or brown) crystals with dark margins are garnet, green crystals are chrome diopside, the buff-colored crystals are orthopyroxene and the small, colorless crystals are olivine. The micrograph, which was made in natural light, enlarges the specimen to twice its natural size.



SHEARED NODULE is seen in a photomicrograph made between crossed polarizing filters in order to enhance contrast between crystals. The large crystals are orthopyroxene. They are set in a matrix of olivine that has been reduced to a finely milled powder by extreme deformation. The specimen is from the Bultfontein pipe at Kimberley. It is magnified some 20 times.

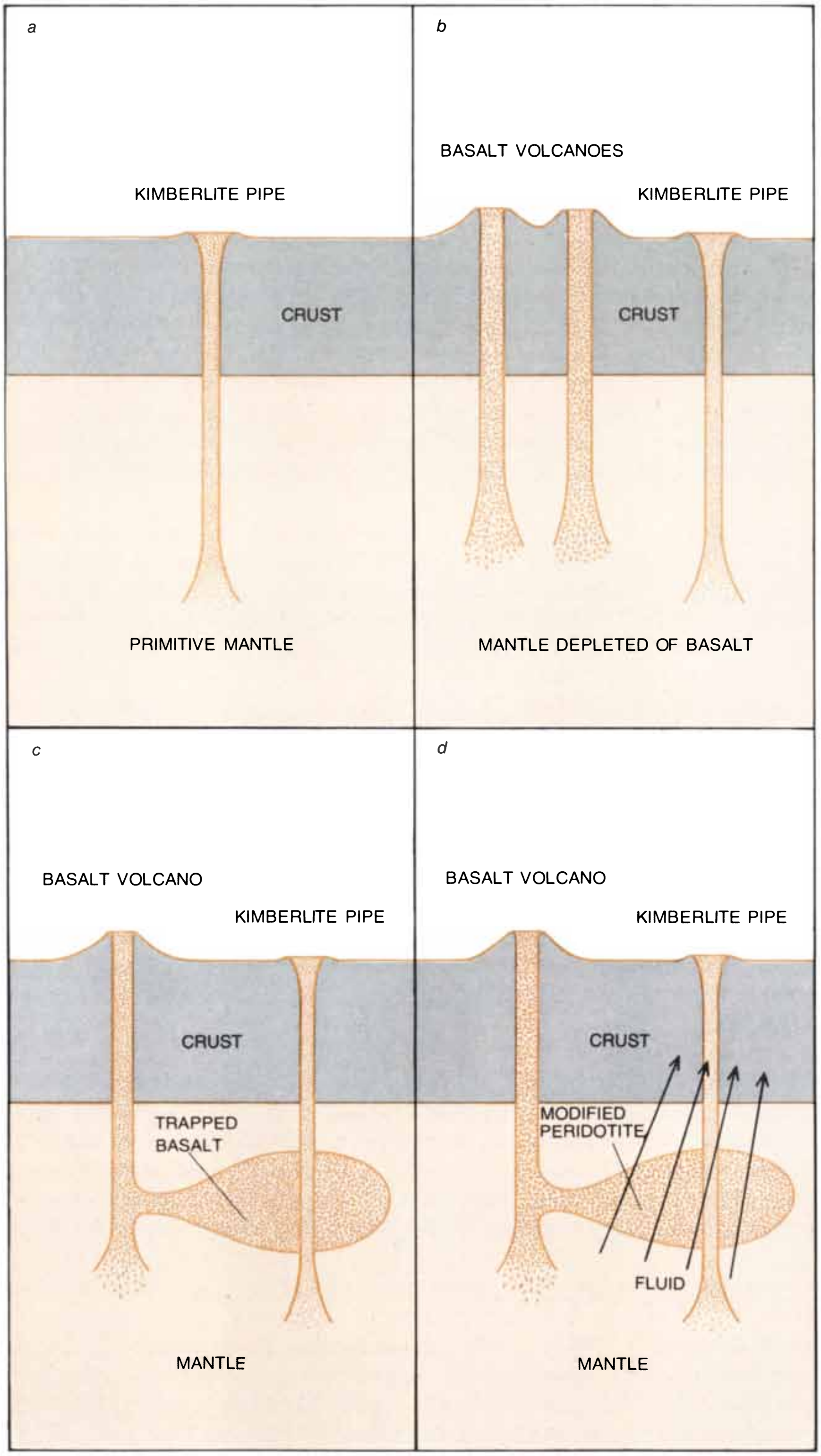
the Laboratory of Structural Geology in Nantes have proposed a model in which an ascending mass of solid peridotite develops a skin of highly deformed rocks. These rocks are then incorporated as sheared nodules into magma formed by melting near the top of the structure. It must be emphasized, however, that the origin and significance of the sheared nodules remains a subject of research and debate. A consensus has not yet emerged.

Although ultramafic nodules are derived from the mantle, it is by no means certain that they are typical of native mantle material. The primitive mantle, formed in the early stages of the earth's history, was presumably much like the material observed in chondritic meteorites, modified only by a melting episode in which the iron-nickel phase was removed to form the core. It seems unlikely that mantle material of this kind might be identified in peridotite nodules found at the surface today. Indeed, evidence from the ratios of isotopes in various igneous rocks suggests that there have been several major episodes of melting in the upper mantle.

One way in which the upper mantle has been altered is by the continuous removal of the more volatile components through volcanic activity. The material represented by the peridotitic nodules may therefore be a refractory residue. When peridotite is heated, one of the first fractions to melt has a composition like that of basalt, and basalt has erupted voluminously in virtually all geological periods. Basalts have more calcium and aluminum than peridotite has, and so the residual mantle should be depleted of these elements; on the other hand, it should be comparatively rich in magnesium. These alterations would be expressed mineralogically by the rareness or absence of garnet and clinopyroxene, as in the commonest kimberlite nodules. Significantly, these nodules appear to have a comparatively shallow source, and the sheared nodules from deeper layers are usually less depleted in basaltic components.

Some kimberlite nodules might also be derived from basaltic liquids that had been trapped in the mantle and crystallized there as cumulates, which are collections of crystalline phases deposited by such liquids as they migrate to the surface. Several nodules with abundant clinopyroxene and garnet have been tentatively identified as cumulates.

The interpretation of kimberlites is complicated by the eventful history of the upper mantle. Even several hundred kilometers under the surface the composition and crystal structure of rocks are altered repeatedly by a variety of chemical and physical processes. For example, fluids containing dissolved salts can penetrate the grain boundaries and mi-



SOURCE MATERIAL of ultramafic nodules is most likely to be some constitutent of the upper mantle, but its exact nature remains in doubt. One possibility (*a*) is that the nodules are specimens of primitive mantle material, essentially unaltered since the early stages of the earth's history. This model is not considered plausible because many processes that could alter the composition of the mantle have long been operating; for example, the extensive eruption of basalt lavas could have left the mantle depleted of certain elements and enriched in others. Another hypothesis (*b*) is that the nodules represent just such a residue of mantle material depleted of basalt. The expected composition of such material is similar to that observed in the commonest peridotite nodules. Other kinds of nodules may be derived from basalt trapped in the mantle or from cumulates precipitated there by basaltic magma migrating to the surface (*c*). Finally, composition and mineral structure of such trapped materials (and indeed of almost all mantle rocks) could be locally modified by the passage of mineral-bearing fluids through them (*d*).

crofractures of solid rock. Chemical reactions with the dissolved ions can completely change the character of the host rock. Phlogopite, the mica found in peridotitic nodules, is frequently formed in this way.

Melting followed by slow cooling and recrystalization has also probably altered the structure of many rocks incorporated in kimberlite nodules. Much of the evidence required for recognition of their source is thereby destroyed. Nevertheless, some events in the history of the nodules can be reconstructed. For example, in many nodules garnet and clinopyroxene tend to be closely associated. In an earlier, high-temperature stage they may have been dissolved in each other or they may both have been dissolved in an orthopyroxene. On cooling, the two components were precipitated, leaving only the spatial association as a clue to the earlier structure.

The origin of eclogite nodules has proved particularly difficult to comprehend. The eclogites are similar to basalt in composition, and they have been widely interpreted as cumulates precipitated from basaltic liquids. Their banded texture is not always consistent with this hypothesis, however. An earlier alternative explanation proposed by Alfred E. Ringwood of the Australian National University suggests that they derived from sections of oceanic crust drawn into the mantle at subduction zones, where converging continents and seafloor meet. If eclogite nodules do originate in this way the presence of diamonds in them is particularly intriguing.

The origin of the kimberlite matrix is perhaps even more obscure than that of the ultramafic nodules. Kimberlites have the high magnesium content typical of mantle liquids, but they are also extremely rich in volatile elements and have higher concentrations of potassium than is normal for such magnesium-rich rocks. Liquids similar to kimberlite might be produced by the fractional crystallization of basaltic magmas, and it may not be a coincidence that both southern Africa and Siberia were sites of widespread basaltic volcanism. Trapped pockets of molten basalt could conceivably have evolved toward a kimberlitic constitution, although in some cases the fluid must have remained trapped for as long as 100 million years. Alternatively fluid kimberlite might be the product of very slight partial melting of the mantle material, perhaps enriched in certain elements by the migration of fluids through it.

The interpretation of kimberlite and the nodules it contains would surely be more secure if their history were less complicated. Even if the story they tell is for now a confusing one, however, they remain among the best available sources of information about the material of the upper mantle.

Insects of the Water Surface

The attractive force between water molecules creates an elastic film at the surface of a calm body of water. All over the world the same four types of insects are found on or near this unusual environment

by Lorus J. Milne and Margery Milne

The surface of a pond, a pool in a stream or even a puddle is a very special environment. Wherever air and calm water meet, a surface film divides the dry world above from the wet world below. Although many species of insects spend at least part of their lives in the water, only a few species of them frequent this surface realm, finding food and company along the fragile interface. These insects are marvelously adapted to exploiting the water surface.

The particular properties of the water surface are created by the attractive forces between water molecules. The molecules within a body of water are attracted equally from all sides, but the attractive force on the molecules near the surface is uneven; hence they are pulled toward the water's center of mass. As a result the surface acts as if it were a stretched elastic membrane, supporting objects with a density greater than that of water and allowing certain insects to stand and move on it.

All over the world the same four types of insects are associated with freshwater surfaces. They are water striders, whirligig beetles, backswimmers and springtails. Each of these insects exists in a different relation to the water surface and to the air and water the surface separates.

The water strider lives on top of the surface film. It is a true bug, that is, a member of the insect order Heteroptera. The water strider is named for its ability to glide or dart across the water surface by rowing with its middle pair of legs; it is also called the pond skater or *wasserläufer* (water runner). There are several genera of water striders, two of which are found all over the world: *Gerris*, which lives on freshwater surfaces, and *Halobates*, which lives on saltwater surfaces.

Most water striders are a little more than five millimeters long. Their dark bodies are slightly flattened and tapered at the front and rear. When the strider stands on the surface film, it is supported by its long, slender middle and hind legs. A shorter pair of forelegs, almost hidden under the strider's head, can be raised from the film to manipulate prey insects. When the strider skates across the water surface, its hind legs are held outstretched and its front ones are held close together on the water to provide support.

The parts of the water strider that come in contact with the surface film are not wettable. A substance is said to be wettable if water molecules adhere to it, that is, if the molecules are more attracted to the substance than to each other. The surface film, created by the attractive force between water molecules, is destroyed when a wettable substance comes in contact with it. Usually only the strider's feet and the part of its hind legs below the knee joint come in contact with the water. These areas are protected by stiff brushes of water-repellent hairs. The rest of the strider's body is covered with water-repellent scales. Even the pair of sharp claws characteristic of the insect foot are well situated, back from the tip of the strider's leg and away from the surface film they could puncture.

It seems clear that water striders can distinguish a suitable water surface from an unsuitable one. A few raindrops send them rushing to the shore, and they also leave the water at the approach of winter and freezing temperatures. On land the striders usually conceal themselves under fallen leaves until the water is once again calm. Rain and ice ruin the surface film, destroying the medium in which the strider detects and obtains food and interacts with other striders.

The water strider obtains information about its environment from tiny ripples in the surface film. Rodney K. Murphey of the State University of New York at Albany discovered that the strider learns of a disturbance in its vicinity through sensors in its four longer legs. Studying the large water strider *Gerris remigis*, he found that the strider depends on vibration sensors located in the flexible membrane between the segments of its feet. The sequence in which these sensors detect an oncoming ripple tells the insect almost exactly how much to turn to face danger or a potential meal. The strider turns toward a disturbance by moving its rowing legs in opposite directions, one forward and one back, just as a human rower would push a pair of oars to turn a boat. The hind leg on the side toward which the strider is turning acts as a pivot, swinging forward slightly as the other hind leg and the front legs provide support.

The strider uses its eyes as well as its vibration sensors to evaluate other situations. Studies of small striders in West Germany show that the insects can see anything as large as a human face from 10 feet away if it appears against a contrasting background. Lanna Cheng of the Scripps Institution of Oceanography reports that when oceanographers tow nets behind ships on dark nights they collect many more seagoing striders than they do during the day or when the moon shines brightly. She offers statistical evidence to show that when there is enough light, a strider can see the net coming and dodge it.

Some combination of visual and vibratory information makes the strider decide whether to approach a disturbance in its vicinity or to flee. If the strider decides to approach, it turns and rows straight toward the disturbance, pausing for about a fifth of a second after each stroke of its rowing legs. During this short pause the strider can, if necessary, detect oncoming ripples and correct its course.

We recently saw an example of the speed and sensitivity of *Gerris remigis* during a small drama close to a New England pond. A sparrow hopping across a grassy area beside the pond disturbed the guardian ants of a subterranean ant nest. We knew the nest was there and understood why the bird suddenly began to scratch itself in different places; the ants were biting and stinging it. The sparrow hopped a short distance away, shook itself vigorously and then flew into the air over the pond. An ant that had held on too long dropped into the water a few feet from the shore. It

WATER STRIDER stands or glides on top of the elastic surface film created by the attractive force between water molecules. The strider's tarsi, or feet, and hind legs below the knee joint, usually the only parts of the insect in contact with the water surface, are covered with water-repellent hairs so that they do not destroy the surface film. In this photograph of two water striders mating what appear to be highlighted bubbles are actually depressions made in the surface film by the water striders' hind legs and middle tarsi.

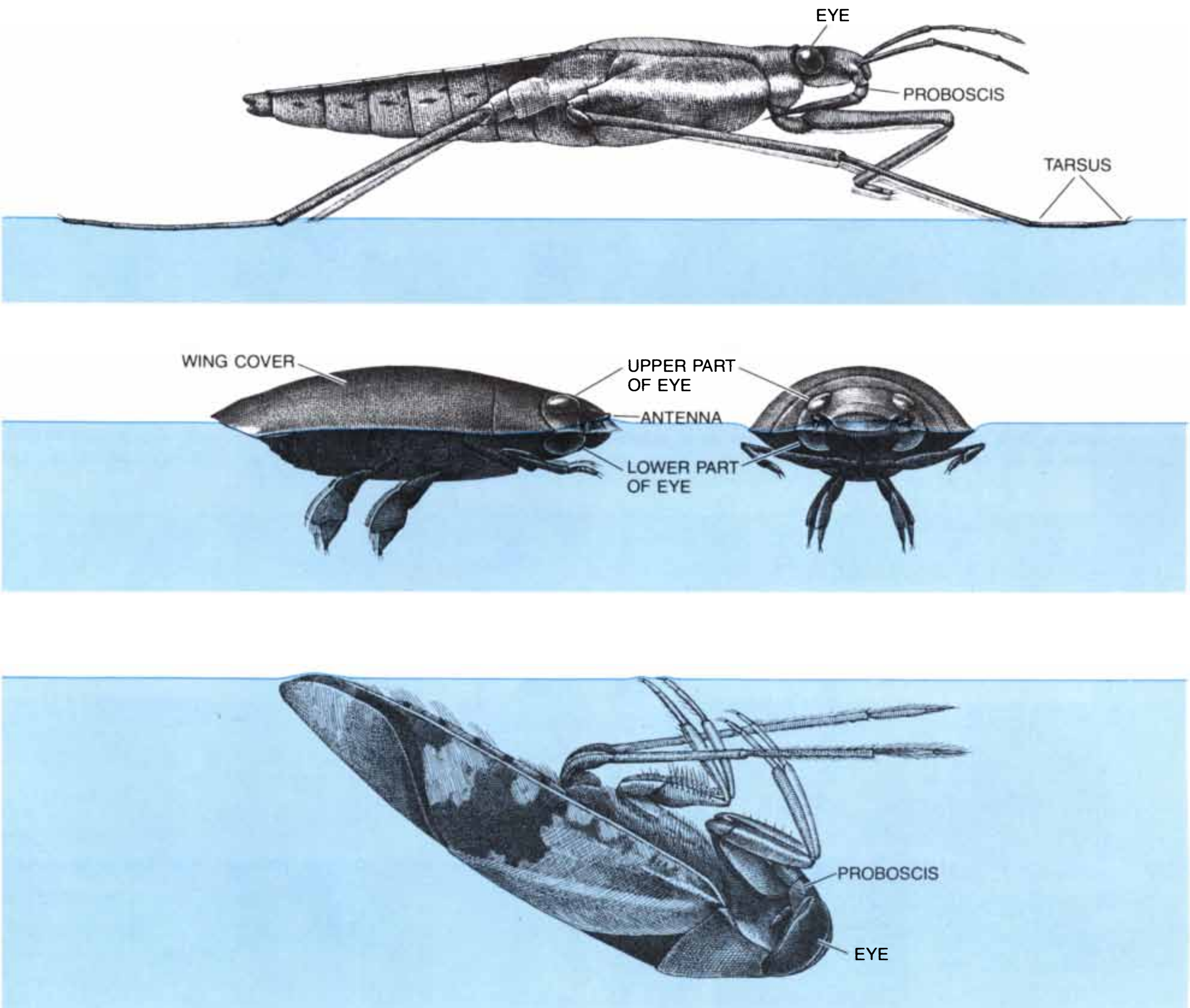
BACKSWIMMER lives in the water below the surface film. It lies belly up because its main air supply is stored on its ventral surface. The insect propels itself through the water by stroking its long, oarlike hind legs. The backswimmer shown in this photograph is resting the sensitive tips of its middle legs against the surface film. Both the backswimmer and the water strider obtain information from tiny ripples in the film. The abdominal tip of the backswimmer's body, which touches the surface film, is reflected in the top of the film.

could not swim, and its struggles set up vibrations in the surface film, promptly alerting two water striders five or six inches away. They turned quickly and came rowing toward the ant at top speed. The strider that arrived first drove its sucking mouthparts neatly into the ant and went off carrying its prize impaled. The mouthparts of the strider pumped an anesthetic, digestive saliva into the body of the ant. In less than three minutes the nutritious contents of the ant liquefied, allowing the strider to suck up its meal. The strider dropped the empty shell of the ant and left it floating on the surface film.

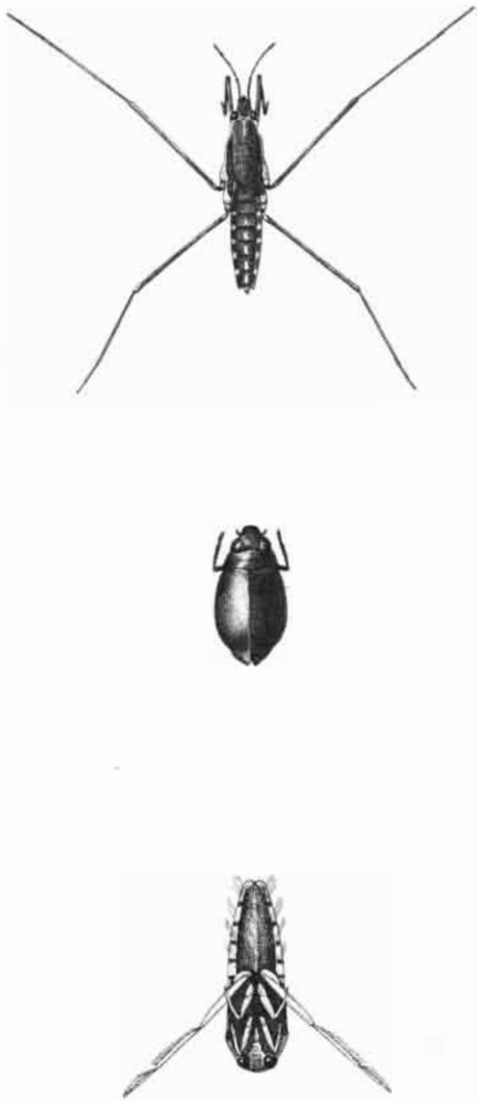
If the prey had been larger, the two water striders might have shared it. It is not rare for several striders to share a large insect such as the winged queen of the large carpenter ant. The combined saliva of the striders probably enables them to subdue such large prey. Most of the time, however, the striders feed independently on smaller insects. Even so they often skate quite close to one another in large groups. Sometimes at midday hundreds of striders congregate in the shade of a large tree overhanging their pool or pond. If one strider bumps into another, the two usually separate immediately. These meetings are probably errors, perhaps the result of the misinterpretation of vibrations in the surface film.

The striders in these groups appear to be moving aimlessly across the water. Actually their movements are quite purposeful, although the purpose is often hidden from an observer above the surface of the water. When some smaller insect that lives below the surface approaches it to get air, the water strider can stab through the surface film, spear the insect, suck it dry and cast it aside without ever bringing it into view.

The surface film also serves as the water striders' instrument of communica-



RELATIVE SIZES of three insects and their different positions with respect to the water surface are shown in this illustration. The water strider (*top*) stands on top of the surface film on its long, slender hind and middle legs or glides across the film by rowing with its middle legs. Its shorter forelegs can be held together on the surface to provide support when the strider is moving or they can be raised from the surface to manipulate food. The strider feeds on smaller insects found on or near the surface film; spearing its prey with its sharp proboscis, the strider pumps an anesthetic and digestive saliva into the body of the insect so that in a few minutes the body contents liquefy and the strider can suck out its meal. The whirligig beetle (*middle*) lives half in the water and half out. The insect can crawl on a solid surface, dive and fly, but most often it plows a twisted path through the surface film by paddling its middle and hind legs. As it zigzags quickly through the water surface it pushes up a series of ripples in the surface film. The whirligig's short antennae are held at the top of the surface film to detect any ripples that are reflected back from an obstacle ahead so that the insect can change its course to avoid a branch floating in the water, to feed on a small insect caught in the surface film and so on. The whirligig also obtains information through its two divided compound eyes: the lower part of the eyes records events below the surface and the upper part records events on and above the surface. When the backswimmer (*bottom*) lies resting against the underside of the water surface, the air that it carries tion; the striders make their own ripples in the surface film in patterns that affect the behavior of other striders. Our own photographs of groups of striders on streams in New England often show a single strider surrounded by concentric circles of ripples: a vigorous message it has just sent out. R. Stimson Wilcox of Purdue University has been able to interpret signals of this kind. Working with Australian striders of the genus *Rhagadotarsus* in his laboratory, Wilcox found that striders reacted characteristically to different signals he generated with an electrically driven float dipping into the strider's aquarium water. The



in its long abdominal grooves pushes it up against the surface, creating bulges where its legs and abdominal tip come in contact with the surface film. The backswimmer can move away from the surface film either by allowing itself to sink farther below the film or by swimming. It will swim to the bottom of a shallow pond in search of insect prey. The backswimmer is a fierce predator, stabbing its prey from below with its sharp beak and sucking out a meal. Views of the three insects from above the surface are shown at the right.

insects responded even in total darkness to the various signals for courting, threatening another strider and so on.

One of the major functions of this type of communication is in the different stages of courtship and mating. Wilcox found that a male strider apprises females of his desire to mate when he has found a suitable site for mating. The site is a fixed or floating object in the water, such as a water plant or a piece of wood floating on the surface. The male grasps the object with his forelegs or stands near it and begins to shake the surface film by moving his rowing legs up and down. Each call begins at a high frequency (23 to 29 waves per second), stabilizes at a lower frequency (18 to 20 waves per second) and ends with one or two low-frequency waves (10 to 17 waves per second). The male sends out seven to 15 of the waves. Wilcox could find no individuality in these courtship calls; the ripples generated by a single male varied as much as the ripples generated by different males.

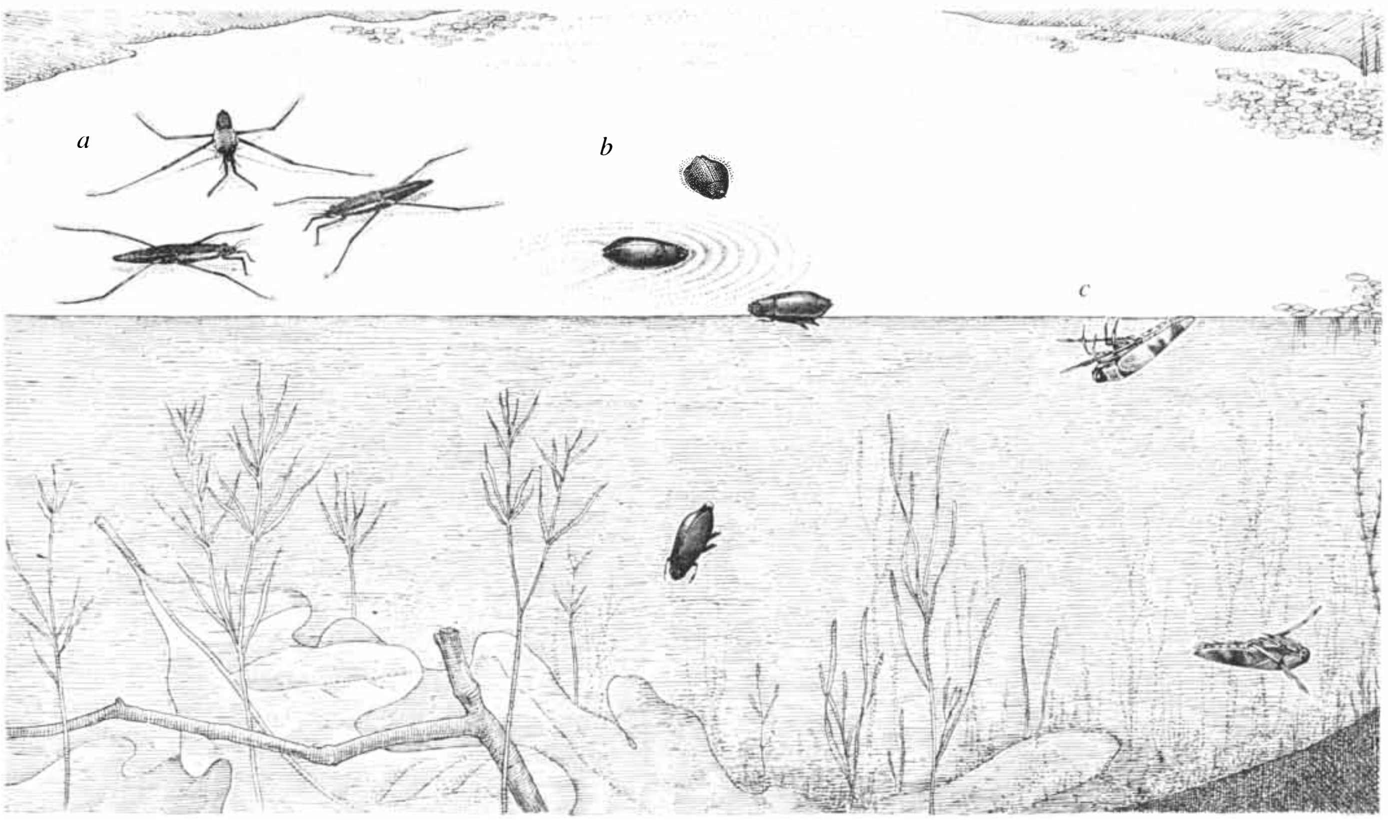
A female within the range of the waves may or may not respond to them. When she does respond, she rows toward the male, often signaling as she approaches him. The female's signals are low-amplitude waves, created by vertical movements of the forelegs; the frequency of the signals is usually between 22 and 25 waves per second. As soon as the male detects the approaching female he sends out similar vibrations. The female indicates her willingness to mate by either crossing one of her longer legs over one of the male's or by grasping one of his hind or middle legs with her forefeet and pulling for a moment.

The male backs away from the object he has been holding, and the female moves forward to grasp it for support with her forefeet. The male mounts her and they copulate for about a minute. The male then backs away and, while he is still facing the female, generates post-mating signals that are indistinguishable from the courtship signals. At this point the female usually releases the supporting object and then grasps it again with her hind legs as she excavates a hole in the object for her eggs. It appears that female water striders will only leave their eggs on objects their mate has been holding. The only objects to which seagoing female striders are known to affix their eggs are feathers: either floating ones dropped by passing birds or feathers on live sea birds asleep on the water at night. A female may deposit dozens of eggs on the underside of a bird's tail before the bird wakes up and flies off.

The male stands close by while the female lays her eggs. He protects her by shaking the water film if another male strider approaches. The frequency of this signal is either steadily higher (23 to 30 waves per second) or steadily lower (nine to 13 waves per second) than the frequencies of the courtship signals. The male employs the same hostile signal to defend a good mating site against other males. He holds his territory by shaking the water film and rushing out to fight any male that comes within eight to 10 centimeters. A male strider without territory to protect may also generate this pattern of vibration to threaten other males on the water surface.

In certain circumstances a strider may employ a kind of locomotion quite different from its usual rowing. For example, a fish might observe a water strider feeding on some wettable prey in the water. If the fish tries to steal the food, the strider will probably jump away. In that case the strider's middle and hind legs work simultaneously to provide the propulsive power for a series of hops that push the insect as much as four inches into the air, usually in a forward direction.

Many species of freshwater striders include both winged and wingless adult



SOME INSECTS GATHER IN GROUPS on the water surface. Water striders (*a*), whirligig beetles (*b*) and springtails (*d*) often congregate in large numbers, but the fierce cannibalistic backswimmers (*c*) are usually found alone in the water. Ripples in the surface film alert all of these insects to prey or predators in their vicinity. Konrad Wiese of the Technische Hochschule in Darmstadt has determined that a single wave with an amplitude of only a thousandth of a millimeter traveling across a water surface can alert half of the insects on top of

individuals, but the winged striders usually fly only in the fall or when the water below them threatens to dry up or become choked with vegetation. Male winged striders can fly for miles, and curiously they travel mostly at night. After surviving the winter in some new location, they absorb their own wing muscles to obtain energy to hunt for food and a mate.

The jet black whirligig beetles *Gyrinus* and *Dineutes* live on a plane lower than that of the water strider. The whirligig's body has a hard, evenly curved exterior that is as slippery as a watermelon seed and is usually dry on top and wet on the bottom because the insect lives half in the water and half out. Although the whirligig can dive and fly and also crawl on a solid surface, it spends most of its time plowing a sinuous path through the water-surface film.

The largest of the whirligig beetles are slightly more than 27 millimeters long and live in pools in torrential mountain streams of Southeast Asia. The largest found in North America and Eurasia are never more than 16 millimeters long. These small insects all seem to be capable of traveling at prodigious speeds through the surface film. Rather than cleaving the water cleanly the whirligig

the surface. The whirligig and water strider generate their own ripples: whirligig by pushing through film and strider by shaking film with its middle legs to signal other striders.

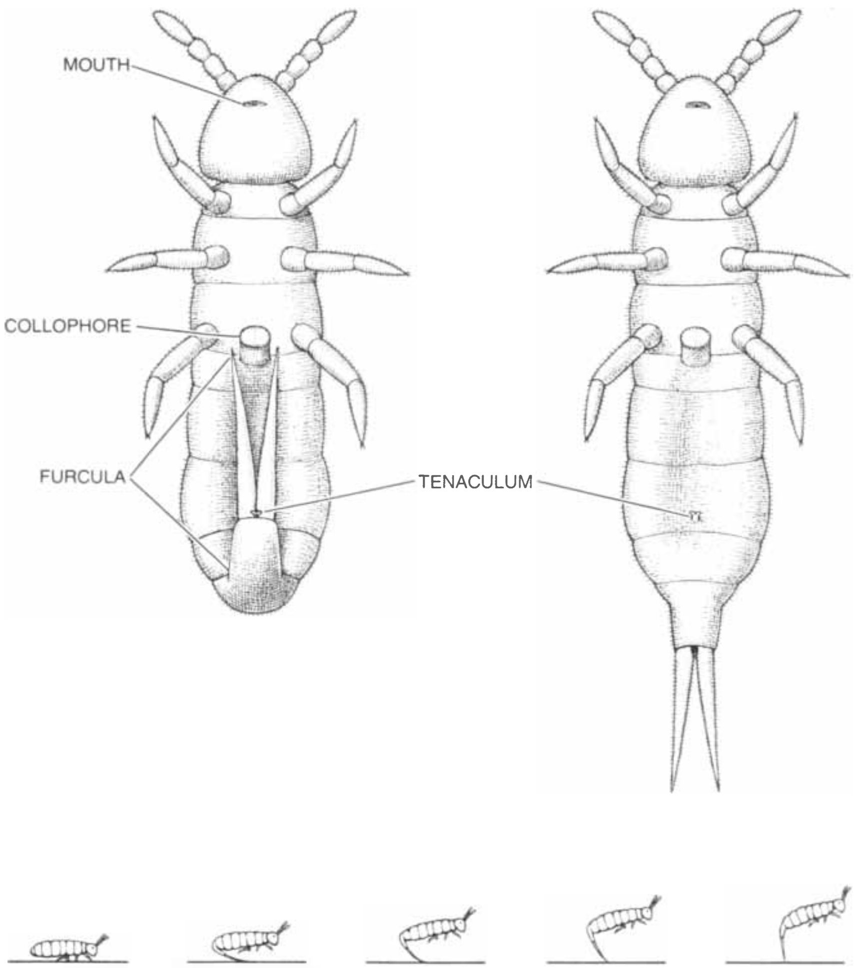
pushes up a series of ripples, proportionately far more impressive than the bow wave of any racing speedboat. Propulsion is provided by the insect's middle and hind pairs of legs, each leg shaped something like a table-tennis paddle. These legs can beat slowly or rapidly to drive the whirligig forward. (The insect's two forelegs, longer and slenderer than the other four, are kept extended in front of the insect, ready to grasp prey.) To stop, the insect lets its middle and hind legs hang down or wave gently.

The whirligig curves to one side or the other when it moves forward, like an inexperienced rower rowing a boat. Sometimes the insect ends up circling in one place, but more often it corrects or overcorrects its course; as a result it zigzags through the water. Thoreau wrote of seeing whirligigs on Walden Pond "ceaselessly progressing over the smooth surface, a quarter of a mile off; for they furrow the water slightly, making a conspicuous ripple bounded by two diverging lines."

When a whirligig paddles at high speed, it can wrinkle the water into as many as 14 consecutive ripples. The ripples form a pattern that remains stationary with respect to the insect, extending up to about six and a half body lengths in front of it. Vance A. Tucker of Duke University has found these patterns in photographs of whirligigs moving at any speed from 23 millimeters per second (.05 mile per hour) to 400 millimeters per second (.9 mile per hour), the maximum whirligig speed he has recorded. These speeds are quite fast for such a small swimmer. The best speed a human swimmer can sustain over a 1,500 meter course is not much more than four times a whirligig's top speed.

A whirligig observes where it is going with compound eyes that are divided into two parts. The lower part of each eye detects events in the water and the upper part detects events on or above the water surface. Glands on the whirligig's body keep its exposed body surface and the upper part of its eyes greased so that they repel water. The underside of its body and the lower part of its eyes are wettable.

The whirligig gathers most of its information about its environment not with its eyes but with the two short, three-part antennae it holds horizontal ahead of its body at the top of the water film. The antennae detect any of the whirligig bow ripples that have been reflected back by an obstacle ahead, for example a smaller insect trapped in the surface film. A larger object in the water causes the film to curve slightly upward, lifting the antennae; the edge of a stream or pond creates an irregular contour in the curved film. The whirligig's responses are so swift that it can correct its course to meet or bypass an obstacle when it is only half an inch away from it. Moreover, the wave-reflection system is so effective that the whirligigs can detect boundaries and obstacles in complete darkness. The system also prevents collisions when dozens or even hundreds of whirligigs whiz about in no discernible formation in a small area of quiet water.

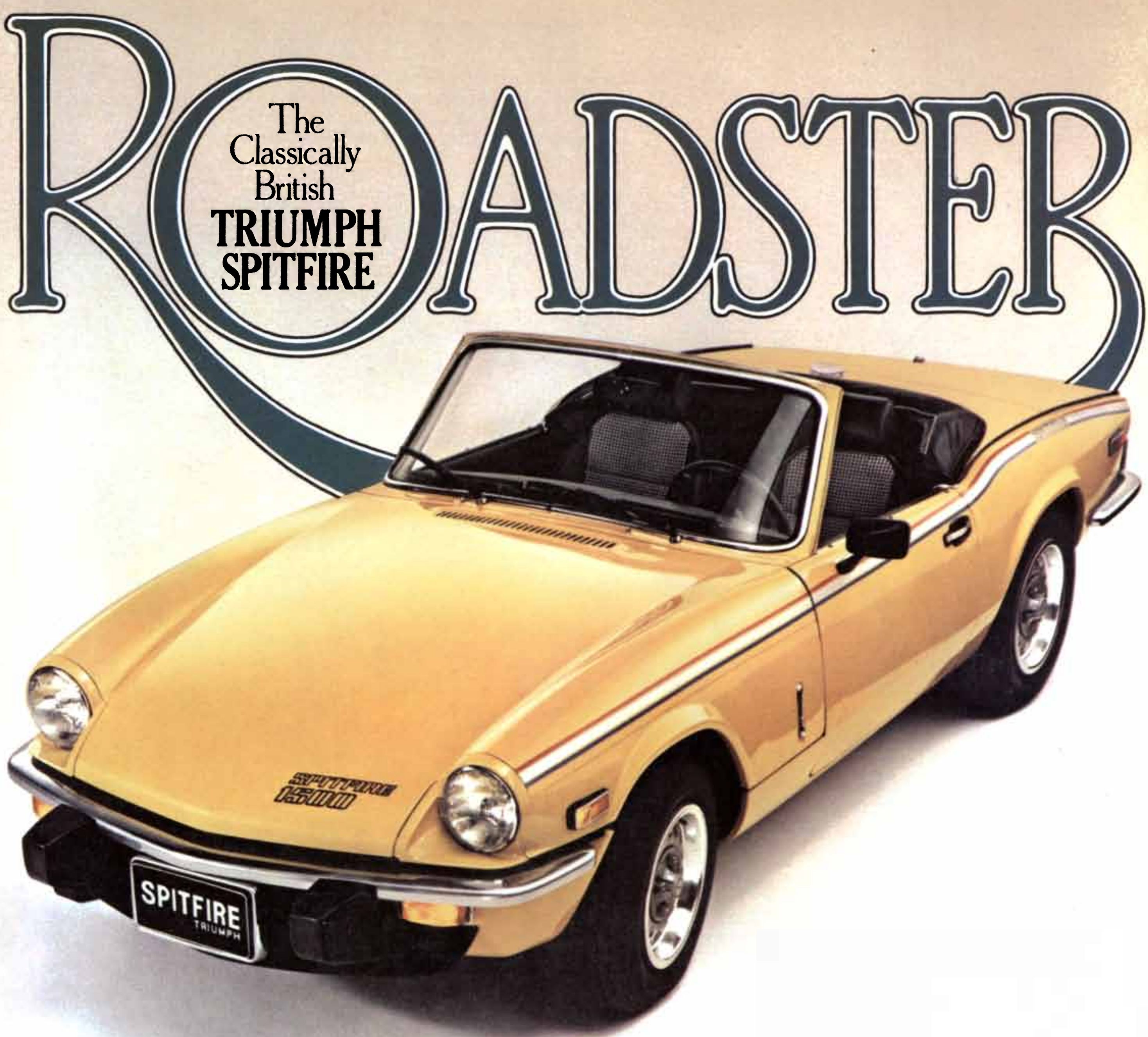


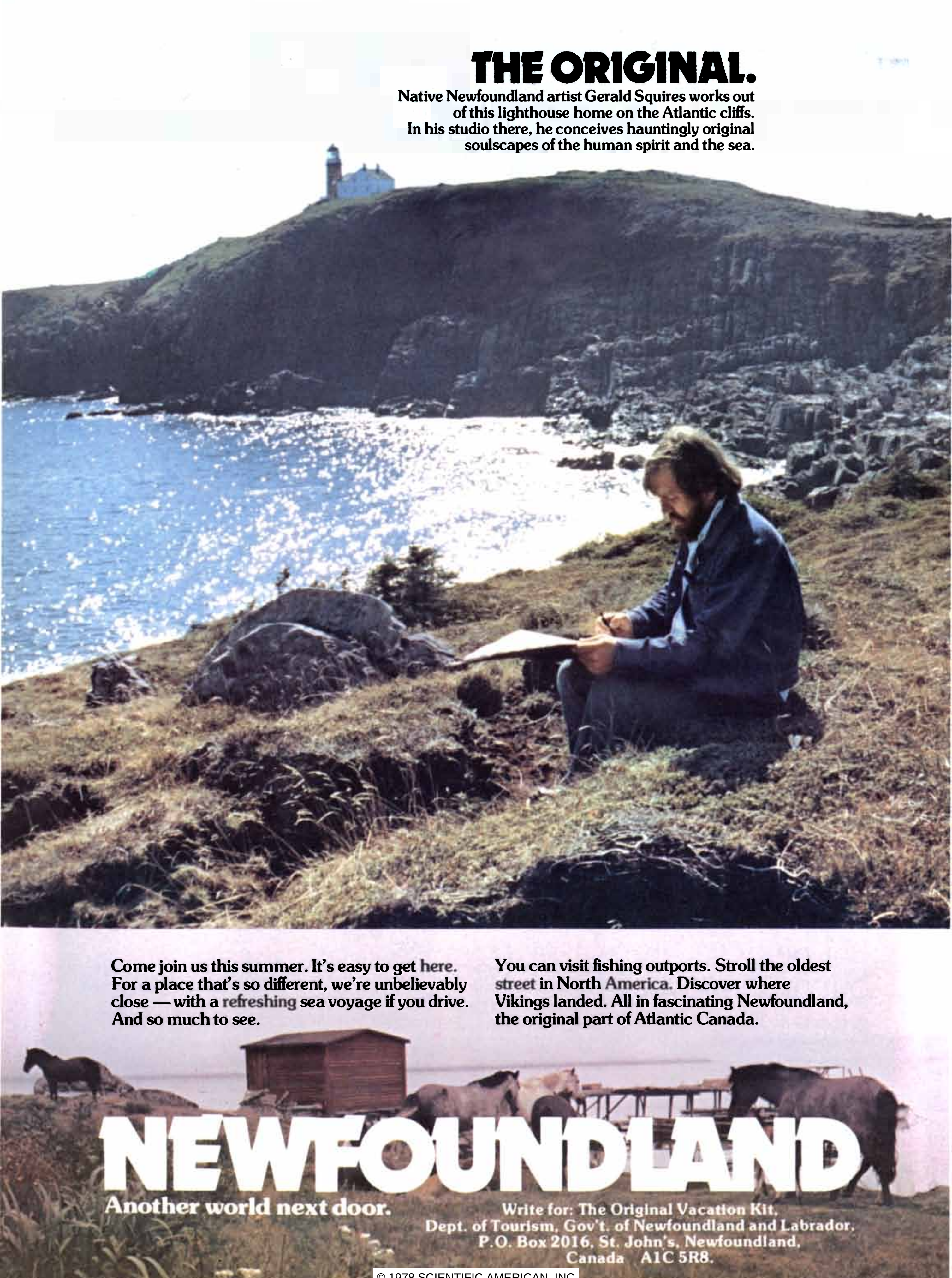
SPRINGTAIL has a tiny body, no more than nine millimeters long, that is water-repellent. The insect crawls on top of the surface film on fresh or salt water and feeds on microscopic plant spores that are blown onto the film. The freshwater springtail (*shown here*) is equipped with a special springing apparatus. The abdominal tip of its body widens into an appendage, usually forked, that is called the furcula. The furcula is curled under the body, held in place by a small appendage called the tenaculum. When the tenaculum is released, the tensed furcula slaps against the surface film, tossing the insect forward into the air as far as 15 body lengths. The only wettable part of the springtail's body is on the stumpy organ called the collophore on its ventral surface. The blunt tip of the collophore is wettable and is kept pressed against the surface film to anchor the springtail, preventing lightweight insect from being blown across water. It appears that the springtail also discharges wastes and takes in water through collophore.

Below the world of the whirligig beetle, in freshwater ponds and in quiet pools that alternate with rapids in streams, insects with torpedolike bodies lie belly up under the surface of the water. These insects are the fierce backswimmer bugs of the genus *Notonecta*. The backswimmers are true bugs, most of them less than 15 millimeters long. They swim belly up because their main air supply is held in two long grooves on their ventral surface. The backswimmer can lie motionless at a shallow depth or move forward to a new position by swinging its outstretched hind legs.

These oarlike hind legs can propel the backswimmer down through the water, away from the surface film. The insect explores a shallow pond from top to bottom and often dives for its prey. Adult backswimmers have wings and are able to fly (and to crawl right side up on land), and yet when there is a disturbance in the water, the backswimmer swims away and often hides by clinging to a plant or stone on the bottom. It is curious that the adult backswimmer should have the habit of hiding, because it is quite capable of taking aggressive action to defend itself. For example, if a human swimmer brushes against a backswimmer in the water, the insect can drive its conical beak through the skin and deliver a painful charge of saliva that can cause infection and swelling.

S. J. R. Birket-Smith of the University


ROADSTER
The Classically British
TRIUMPH SPITFIRE
SPITFIRE 1500
SPITFIRE
TRIUMPH


THE ORIGINAL.
Native Newfoundland artist Gerald Squires works out of this lighthouse home on the Atlantic cliffs. In his studio there, he conceives hauntingly original soulscapes of the human spirit and the sea.
Come join us this summer. It's easy to get here. For a place that's so different, we're unbelievably close — with a refreshing sea voyage if you drive. And so much to see.
You can visit fishing outports. Stroll the oldest street in North America. Discover where Vikings landed. All in fascinating Newfoundland, the original part of Atlantic Canada.
NEWFOUNDLAND
Another world next door.
Write for: The Original Vacation Kit, Dept. of Tourism, Gov't. of Newfoundland and Labrador, P.O. Box 2016, St. John's, Newfoundland, Canada A1C 5R8.

of Copenhagen has observed that juvenile backswimmers, which lack wings, react to a disturbance in the water by rushing up to the surface for air. Some juveniles swim to the surface so quickly that they break through the surface film and then tumble back into the water. Young backswimmers with better control stop at the film, where they spread the long protective bristles that cover their abdominal air-storage channels. When these glistening white chambers are refilled with air, the insect dives again.

Like the whirligig and the water strider, the backswimmer gets information from ripples in the surface film. The insect rests the sensitive tips of its middle legs against the underside of the film. Although it appears to be suspended from the surface, it is actually pushed up against the film by its own buoyancy. Indeed, there are tiny bulges in the film where the feet and abdominal tip of the insect are in contact with it.

The backswimmer shows the greatest interest in surface vibrations that are too fast to interest a water strider. The source of these high-frequency vibrations might be a midge or a mosquito on the surface, beating its wings 100 to 150 times a second before taking off into the air. The backswimmer attacks its prey from below, stabbing it with a sudden thrust of its powerful beak. After the backswimmer sucks out its meal, the empty shell of the midge or mosquito drifts away. (Immature mosquitoes are the preferred prey of the voracious backswimmers; the backswimmers are rarely given the credit they deserve for preventing so many of these potential transmitters of human disease from reaching maturity.)

Occasionally a backswimmer and a water strider arrive almost simultaneously at some insect that has fallen into the water and is caught in the surface film. In that case the backswimmer gets the prize, sometimes by pulling the insect down into the water beyond the reach of the strider. Although the backswimmer and the water strider, each with its ventral surface toward the water film, look almost like mirror images of each other, the backswimmer is the stronger and sturdier of the two.

The fierce backswimmers are rarely sociable. Each of these insects hunts alone and usually accepts only one other individual of its kind, chiefly at mating time. The backswimmer's choice of prey contributes to the insect's isolation. Toward the end of spring, when there are few mosquitoes and other small prey in the water, the adult backswimmers begin to consume younger backswimmers, often including their own offspring. By early summer cannibalism is a way of life for these insects, limiting the size of the backswimmer population. A backswimmer seldom survives long enough to be threatened by starvation. Each backswimmer either succumbs as prey or remains alone, the sole survivor in its vicinity.

The water-surface film is an environment of great antiquity, dating from when the first rains fell on the land and formed puddles and pools of water. From about 375 million years ago to about 250 million years ago only the minute, wingless springtail insects inhabited the film. The springtails, never more than about nine millimeters long, are wonderfully fitted to this environment. Their small, water-repellent bodies are easily supported by the surface film. They feed on the microscopic plant spores that blow onto the film and float there for a time.

Both freshwater springtails of the genera *Podura* and *Sminthurus* and saltwater springtails of the genus *Anurida* can crawl on top of the surface film. Only the freshwater species, however, employ the distinctive mode of sudden locomotion for which the springtails are named. The abdominal tip of the freshwater springtail's body is broadened, almost to the proportions of a beaver's tail; the insect can slap this "springtail" against the surface film to flip itself away from danger.

The springing mechanism works as follows. The abdominal tip is curled forward and held in place under the springtail's body by a special appendage rather like the catch on a mousetrap. The tip is held tensed against the catch so that when the catch is released, the abdomen whacks against the water, tossing the lightweight insect forward into the air. One such spring can throw the springtail as far as 15 body lengths. (Springtails with slightly different anatomies live on land and leap in much the same way.)

Although the springtail can stand and move on the surface film because its exterior is water-repellent, there is one area of the insect's body that serves an important purpose because it is wettable. On the underside of the springtail's abdomen near its hind legs there is a stumpy organ called the collophore. The springtail presses the blunt tip of the collophore against the surface film. It is the tip of the collophore that is wettable, and it anchors the springtail, preventing the lightweight insect from being blown across the water by the wind. It appears that the springtail also takes in water and discharges wastes through the collophore.

The springtail's relation to the water below the surface film is not well understood. A century ago a Swedish naturalist, Charles de Geer, confined several springtails of the species *Podura aquatica* in a deep dish of water. He reported that several times a day the tiny creatures crawled down to the bottom of the aquarium and seemed to suck juices from the submerged plants there. To this day no one knows how springtails penetrate the water film or whether, as is suspected, they do so each fall to hibernate in the mud at the bottom of their pool or pond.

The blue-black springtail *Podura aquatica* is found on freshwater surfaces across Temperate Zone Eurasia and North America and far into the arctic tundra. In fact, it is the most abundant species of springtail found on fresh water. In New England during October and November we see multitudes of these springtails along the edge of almost every pond and slow stream. Until we come too near they mill about in close proximity, forming a seething strip of dark color three to five inches wide. A wave of a hand suffices to alarm a large group of them. Snapping their tensed abdominal tips against the water the springtails take off in different directions and fall back so far apart that their tiny bodies vanish completely from view. If we stand motionless, the gregarious springtails slowly reassemble.

Marine springtails are equally sociable. For example, the blue-gray springtail *Anurida maritima*, found in marshes and along rocky shores, is only about three millimeters long, and yet the insects gather in sufficient numbers to coat the entire surface of a tide pool or marsh puddle. Ralph Dexter of Kent State University reports that in certain seasons it is usual to find in these environments 100 members of the species per square inch.

A century ago an English naturalist, Thomas Belt, commented on the remarkable similarity between the animals and plants of the freshwater surface in Nicaragua and those familiar to him in Europe. The conditions of life are indeed much more uniform on freshwater surfaces around the world than they are on the land areas surrounding them. When someone who has a keen interest in the plants and animals around him is visiting a part of the world far removed from his own, he will feel very much at home if he rows a boat out on a pond or stream. The springtails, water striders, whirligigs and backswimmers will all be there.

The insects of the water surface frequent the transient pools of the polar tundra and are found above the trout in freshwater pools farther south. We have seen them from a dugout canoe in Panama, flourishing in spite of the ant armies, birds and other rain-forest predators on the nearest shore. We have seen these insects or others like them from a narrow dike between two rice paddies in the Philippines and from the edge of a shallow pool in sandstone in an arid area of Australia. They are no less evident in their gossamer habitat in Hawaii, New Zealand, Iceland, Africa and in the Western Hemisphere from Alaska to the Falkland Islands off the southern tip of South America.

Atmospheric Halos

Rings around the sun and moon and related apparitions in the sky are caused by myriad crystals of ice. Precisely how they are formed is still a challenge to modern physics

by David K. Lynch

Anyone who spends a fair amount of time outdoors and keeps an eye on the sky is likely to see occasionally a misty ring or halo around the sun or the moon. The phenomenon is well established in folklore as a sign that a storm is coming. Actually the halo is only one of a number of optical effects that arise from the same cause, which is the reflection and refraction of light by crystals of ice in the air. Whenever cirrus clouds or ice fogs form, arcs of light appear overhead, woven into the veil of cirrus in a splendid variety of circles, arcs and dots.

The effects are best seen when the clouds are thick enough to fill the air with ice crystals but not so thick as to hide the sun. The commonest effect is the 22-degree halo, so named because its radius subtends an angle of 22 degrees from the eye of the observer. The halo appears as a thin ring of light (about 1.5 degrees wide) centered on the sun; sometimes it is pale white and sometimes it is brightly colored, with red on the inside and blue on the outside. The colors are clearest when the clouds form a uniform, featureless haze. If the clouds are patchy, the halo may be incomplete.

Frequently the 22-degree halo appears in company with two "sun dogs," which are bright and sometimes colored patches of light on each side of the halo, either on it or just outside of it. The formal name for them is parhelia, from the Greek for "with the sun." Occasionally visible is a larger, fainter halo with a radius of about 46 degrees.

If the cloud cover is uniform, one can sometimes see a ring of light encircling the sky parallel to the horizon. It is the parhelic circle. It passes through the sun and the parhelia and, if they are visible, through the anthelion (a whitish patch opposite the sun) and the paranthelia (which are like the anthelion but are located at azimuths of plus and minus 120 degrees from the sun).

Other phenomena commonly observed are the circumscribed halo and the circumzenith arc. The circumscribed halo surrounds the 22-degree halo and is bilaterally symmetrical with it. At the top and bottom the two are tangent. The circumzenith arc appears as an inverted rainbow centered on the zenith, facing the sun.

Many other phenomena of this kind have been identified, and I shall describe a number of them. As the optical effects of atmospheric ice crystals are enumerated, however, a point is reached where their existence and properties become questionable. Rare, one-of-a-kind observations haunt the published material, and quantitative measurements are almost unknown. Was the halo real? Could it have been a known halo mistaken for a new one? Was it described accurately? Theoretical work by Robert G. Greenler of the University of Wisconsin at Milwaukee and his colleagues predicts certain arcs that have not been seen; have they been overlooked or is the theory incomplete?

Even though the theory of halos is primarily encompassed by classical optics, the principles of which have been known for centuries, the present understanding of these wonderful arcs is imperfect. Hence the study of halos is as fascinating now as it was 100 years ago.

The belief that a halo signifies the onset of bad weather has a basis in fact. A falling barometer is usually caused by an advancing low-pressure system. Violent convection carries moist surface air to altitudes of from 9,000 to 15,000 meters (30,000 to 50,000 feet), where the temperature is well below freezing. The air becomes supersaturated with water vapor, which condenses out and forms cirrus clouds. High-velocity winds above the system carry the wispy cirrus ahead of it, which is why halos can be seen in these clouds as the lovely first harbingers of foul weather.

Since the ice in the clouds is the source of the optical effects, one is led to consider its structure. The delicate snowflakes of winter show that ice is a hexagonal crystal. Such a crystal has four axes of symmetry: three *a* axes, which are of equal length and intersect at an angle of 120 degrees, and a *c* axis, which is of a different length and is perpendicular to the plane of the *a* axes.

Although many forms of ice can occur, only about four are important in meteorological optics. The others are either too rare or do not have smooth, regular optical faces. The important forms are the plate, which resembles a hexagonal bathroom tile, the column, the capped column and the bullet (a column with one pyramidal end).

In each of these forms, except for the single pyramid, the two end faces are parallel and lie perpendicular to the *c* axis. They are termed the basal faces. The angles between the crystal faces are always the same: 120 degrees for adjacent prism (side) faces, 60 degrees for alternate prism faces and 90 degrees for the junctions between ends and sides. The 60-degree and 90-degree combinations are responsible for nearly all of the halo phenomena.

Atmospheric ice crystals form by direct sublimation from air that is supersaturated with water. The type of crystal depends primarily on the air temperature, although the degree of saturation relative to ice can be a factor when the saturation is less than 108 percent. Then only plates and columns can form. Most of the optically interesting crystals form when the saturation is between 100 and 140 percent and the temperature is between minus 5 and minus 25 degrees Celsius. When the saturation is higher than 140 percent, the growth of crystals is so rapid that rime (an amorphous deposit of frozen droplets) grows on the crystals and destroys their optical faces.

The relations of temperature and saturation can be represented in a diagram in which different regions are designated by Roman numerals [*see illustration on page 151*]. Composite crystals are formed when the growing crystal moves from one region to another. For example, a capped column begins as a simple column. During its period of formation it passes from Region VII to Region II, perhaps because it is descending through a stratified cloud. Once it is in the plate-forming region the columnar growth stops but the basal faces continue to develop outward from the *c* axis. If the column initially has two flat ends, it grows a cap at each end and comes to resemble a spool. A plate will not form on a column that terminates in a pyramid. It is important to realize that a capped column and all other mixed forms originate as single crystals and go through two separate periods of growth; they are not separate crystals that came together after they were formed.

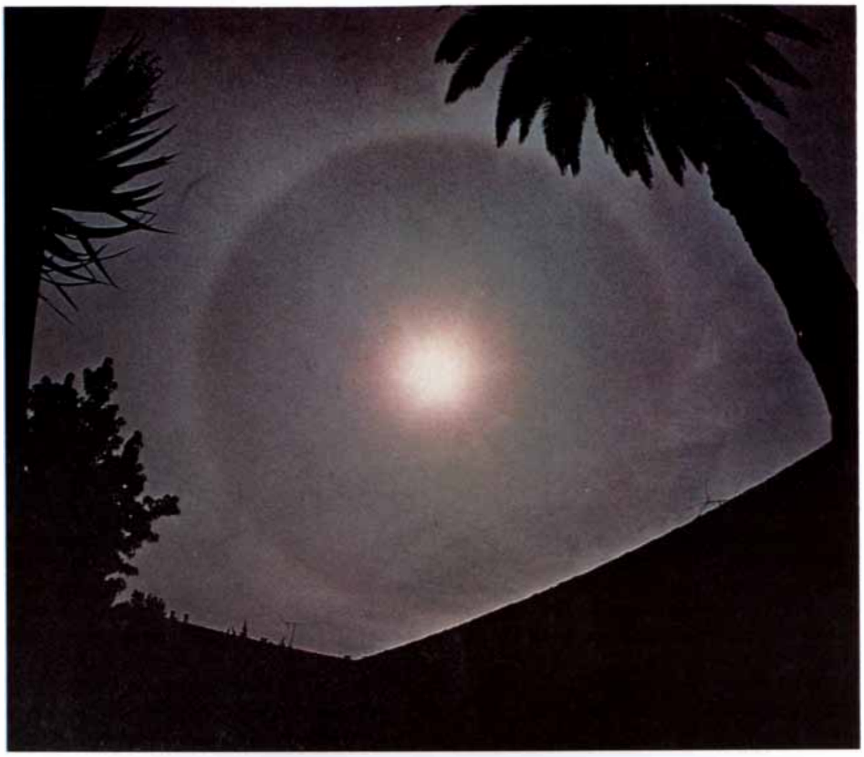
The orientations of the ice crystals as they fall through the air are responsible for the wide variety of halos. Exceedingly small crystals (less than about 20 micrometers in diameter) are subject to Brownian motion, the random movement resulting from the impact on the crystals of the molecules of the air. The random collisions with air molecules cause the crystals to tumble constantly, so that all orientations are present.

When the crystals reach a size of from 50 to 500 micrometers, aerodynamic lift dominates the Brownian motion and forces the crystals into certain positions relative to the direction of their fall. If all of the ice particles are of one kind, they become aligned with one another. (Imagine what such a cloud looks like on a microscopic scale: billions of sparkling prisms lined up uniformly, glinting in the sunlight, each one producing its own family of tiny halos. Most halos are formed in this way.) When the ice crystals reach a size of from .5 millimeter to three millimeters, they tend to spin as they drift downward. These whirling crystals produce yet another class of halos, to which I shall return.

The 22-degree halo is formed by sunlight that passes through alternate side faces of randomly oriented crystals. All the crystals are less than 20 micrometers in diameter and all have 60-degree faces. Since the sunlight strikes the ice crystals at every possible angle, it may seem strange that a cloud composed of countless independent crystals should direct light chiefly at an angle of 22 degrees.

The principle underlying this effect is called the principle of minimum deviation. It is a cornerstone of classical optics and finds application in many areas of meteorological phenomena, including rainbows. Since the crystal faces are inclined to one another by 60 degrees, the problem of the 22-degree halo becomes one of understanding the passage of light through an ordinary spectroscope prism in a plane perpendicular to the *c* axis.

The angle between the incident ray and the emergent ray is termed the deviation; it is the angle by which the light changes direction in the crystal. As the angle of incidence increases from zero degrees (perpendicular to the face) the deviation decreases steadily, reaches a broad minimum and then increases again. In the vicinity of the minimum a



COMMON HALO, known as the 22-degree halo because its radius subtends an angle of 22 degrees from the eye of the observer, surrounds the sun. The photograph was made in Pasadena, which is why one sees a border of palm fronds. The halo is the result of randomly oriented ice crystals in cirrus clouds. This photograph and the one below were made with a wide-angle lens.

HALO COMPLEX was photographed near the South Pole. Visible are the 22-degree and 46-degree halos, the parhelic circle and its parhelia, the upper Parry arc and the circumzenith arc. The variety shows that the cirrus clouds had various shapes and orientations of ice crystals.

SUN AND PARHELION were photographed from White Mountain in California. The parhelion, or "sun dog," is the bright spot at the right. The formal name is derived from the Greek for "with the sun." Frequently two parhelia are visible on opposite sides of the sun.

THREE CROSSED ARCS meet at the antisolar point, which is always below the horizon when the sun is up. The arcs result from multiple reflections off the side faces and one end face of column-shaped crystals. The optics are much the same as in an ordinary kaleidoscope.

change in the angle of incidence produces no change in the deviation. Light therefore accumulates at the angle of minimum deviation. The 22-degree halo is this concentration of light, and it is circular because all orientations are present. Since the deviation of light can be more than 22 degrees but not less, the halo is actually doughnut-shaped, with a bright and sharp inner edge due to minimum deviation and a diffuse outer region resulting from the rays that traverse the crystal at other angles.

Two quantities determine the angle of minimum deviation: (1) the angle between the faces and (2) the index of refraction. The mean index of refraction for ice is about 1.31; as in all solids, however, it varies slightly with color, that is, with wavelength. This property is termed dispersion. It causes white light to be split up so that each component color travels in a slightly different direction. Hence the angle of minimum deviation is a bit different for each color, being smallest for red light. Thus the halo is in fact composed of a continuum of superposed halos, each of a slightly different color and size.

Without dispersion the overlapping halos would combine and would appear pure white, like the light from which they originated. Since the reddish halos are smaller than the others, however, they are seen at the inner edge of the composite halo. Being slightly separated from the rest, their colors are washed out the least. Other colors are considerably smeared because red rays near the minimum deviation can fall on them at the minimum deviation, whereas the opposite effect is impossible.

The 46-degree halo is formed in exactly the same way except that the crystal faces that refract the light are a basal face and a side face that share an edge. Such faces always intersect at 90 degrees rather than 60 degrees. Other halos formed by minimum deviation in randomly oriented crystals are rarely observed, but they can be explained in terms of the prisms with pyramidal terminations. They include six halos ranging in size from eight to 32 degrees.

The commonest optical effects caused by oriented crystals are the parhelia. They are at least as common as the 22-degree halo and are much easier to see because they are brighter. The crystals responsible for these "mock suns" are capped columns, bullets of moderate size and plates, all with vertical *c* axes. Aerodynamic lift forces the crystals to descend in this position.

As before, the light passes through alternate side faces. Since the faces are vertical and the sun is above the horizon, however, sunlight enters the crystal obliquely, and the plane on which the light travels is not perpendicular to the *c* axis. Thus a true minimum deviation does not occur. A "quasi-minimum" deviation does take place, and it concentrates light; the angle is always more than 22 degrees, however, so that the parhelia are formed outside the 22-degree halo. Only when the sun is on the horizon are the conditions for true minimum deviation fulfilled, and then the parhelia do lie on the 22-degree halo.

The colors of the parhelia, which are often dazzling, result from refraction, as within the 22-degree halo. Like most oriented crystals, the plates and capped columns tend to wobble around their mean orientation; the movement smears out the optical effect. The brilliance of a parhelion can be affected by the degree of alignment of the available crystals. Parhelia often show a bluish-white tail that extends horizontally away from the sun. It is caused by the rays that traverse the crystal near but not at the quasi-minimum deviation. The tail is most evident when the sun is low.

Analogous to the parhelia of the 22-degree halo are the parhelia of the 46-degree halo. They are rarely reported and indeed may not exist at all. If the phenomena reported as parhelia of the 46-degree halo are really that, they are formed by a quasi-minimum deviation in a 90-degree prism of a crystal that

VARIETY OF OPTICAL EFFECTS appear in another photograph made near the South Pole. The sun, which is obscured by the flag, is surrounded by parts of the 22- and 46-degree halos and the colorful circumzenith arc, which is the concave arc near the top of the photograph.

TWO EFFECTS, a column and an upper tangent arc, are visible in this photograph. The column rises upward from the sun, and the arc is the bright spot above it in the sky. Such an arc is sometimes seen tangent to a 22-degree halo; from the air one sometimes sees the lower arc.

has its *c* axis horizontal and its refracting edge vertical.

One of the loveliest members of the halo family is the elusive circumzenith arc. Although it is formed in the same crystals (capped columns and bullets) as the parhelia are, it is observed far less often because it can occur only when the sun is below 32.2 degrees of elevation. (Moreover, people seldom look straight up, which is the direction of the arc.)

In forming such an arc light enters the upper horizontal face of a crystal and emerges through a vertical side face. At elevations larger than 32.2 degrees the light is totally reflected internally. At 32.2 degrees the emerging light travels straight down. Hence the circumzenith arc appears as a bright spot at the zenith. As the sun drops below this elevation the spot opens up into a splendid arc of color centered on the zenith and facing the sun. Although the circumzenith arc is not a mimimum-deviation phenomenon, it does achieve its maximum brightness when the sunlight passes through the crystal at the minimum deviation. That happens when the solar elevation is 22.1 degrees, at which point the 46-degree halo and the circumzenith arc are tangent.

Complementing the circumzenith arc is the colorful circumhorizon arc. It is formed in the same crystal but by light that enters through a vertical side face and leaves through the bottom horizontal face. On the basis of symmetry one can readily infer that the circumhorizon arc cannot appear when the sun is below an elevation of 57.8 degrees (90 degrees minus 32.2 degrees). The arc starts out as a ring of color on the horizon. As the sun rises so does the encircling circumhorizon arc. The maximum brightness again occurs at the minimum deviation, when the elevation of the sun is 67.9 degrees (90 degrees minus 22.1 degrees). Because this arc is a high-sun phenomenon it is one of the few halos that cannot be seen from any place on the earth. It is visible only in the Temperate Zone latitudes from 55.7 degrees north to 55.7 degrees south, where the sun can get high enough. Since the sun's elevation is highest in the summer, when cirrus clouds are less likely to form, the circumhorizon arc will probably always remain a rare sight.

Analogous to the circumzenith and circumhorizon arcs for the 60-degree faces are the upper and lower Parry arcs, named for an arc phenomenon described by the British explorer Sir William Parry in 1821. The arcs are formed just above and below the 22-degree halo. They change shape dramatically with changes in the elevation of the sun. This relation has undoubtedly caused confusion in identification, since the arcs can masquerade in many forms. The associated crystals are columns that are oriented with their *c* axes horizontal and with two side faces also horizontal, one on the top and one on the bottom.

Two halos, the parhelic circle and the solar pillar, deserve special attention because they are formed primarily by external reflection from oriented crystals. They are therefore colorless. The parhelic circle is formed by reflection from the vertical side faces of capped columns and plates (*c* axes vertical) and from the end faces of horizontal columns. Since there is no preferred azimuthal orientation, the side faces scatter light in all horizontal directions while preserving the vertical component. Thus light appears to come to the observer from every point of the compass but from a single altitude. The parhelic circle is seen as a horizontal ring of light running through the sun and encircling the sky parallel to the horizon. It is seldom seen in its entirety because the clouds usually do not cover the sky uniformly.

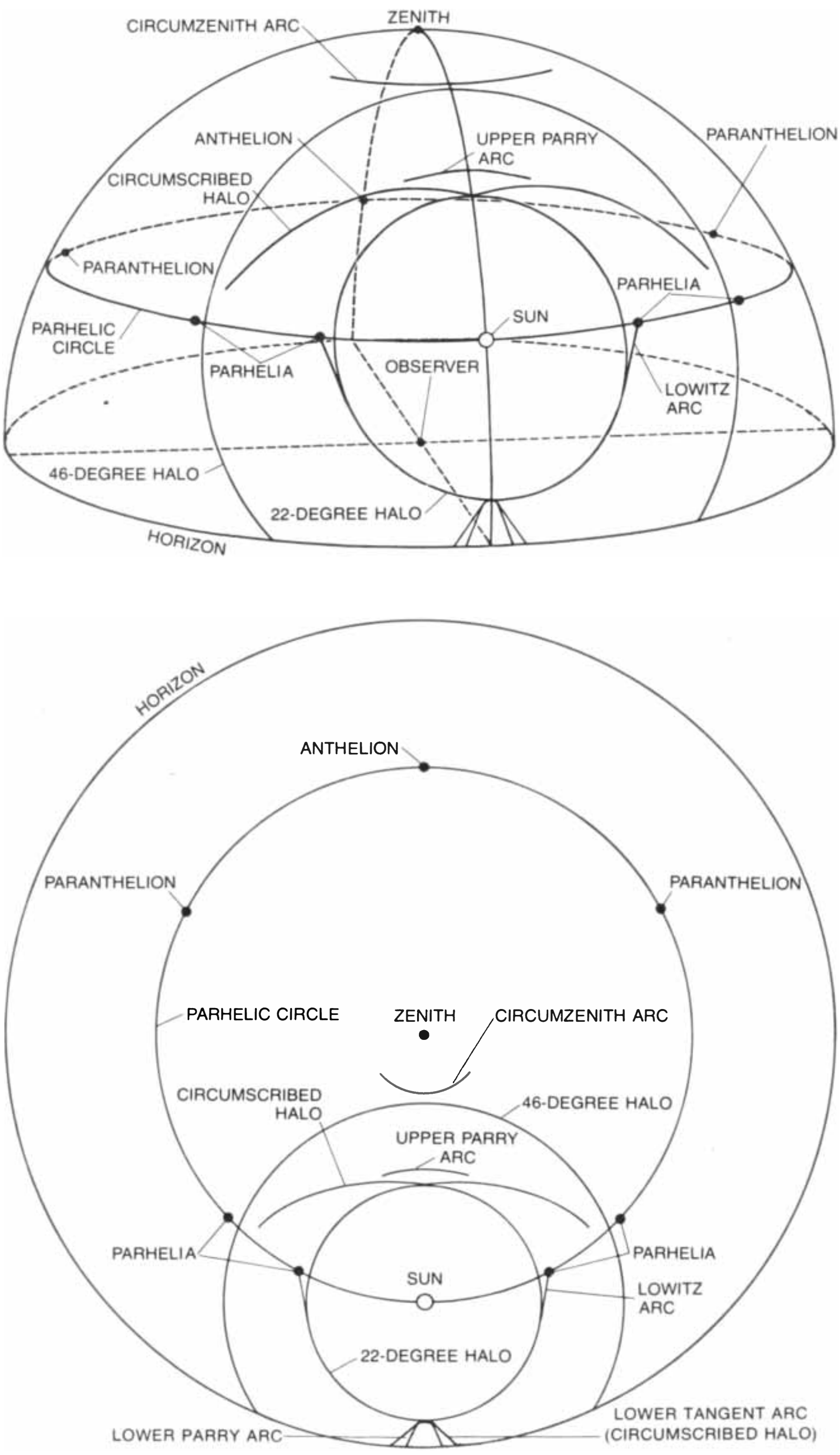
The solar pillar, a commoner phenomenon, is a vertical shaft of light extending upward from the sun. It is most often observed above the rising or setting sun. Occasionally it is tilted or seen below the sun. The pillar is caused by reflection from the basal faces of plates and capped columns. As the crystals descend (with their *a* axes horizontal, like a leaf) they wobble around the mean orientation and smear the reflected solar image out vertically. Pillars therefore provide strong evidence of oriented crystals and also show that the crystals oscillate. Although pillars produce no color of their own, they take on the color of the sun and so often appear to be orange or red.

When the pillar, the parhelic circle and the 22-degree halo appear together, they often form crosses in the sky. This effect has undoubtedly led some people to interpret halos as signs from heaven. The most famous account of a well-timed cross resulted from an incident in the Swiss Alps in the summer of 1865. Edward Whymper and his companions were returning from the first ascent of the Matterhorn when four of them fell and were killed. Some hours later Whymper saw a circle with three crosses in the clouds, "a strange and awesome sight, unique to me and indescribably imposing at such a moment."

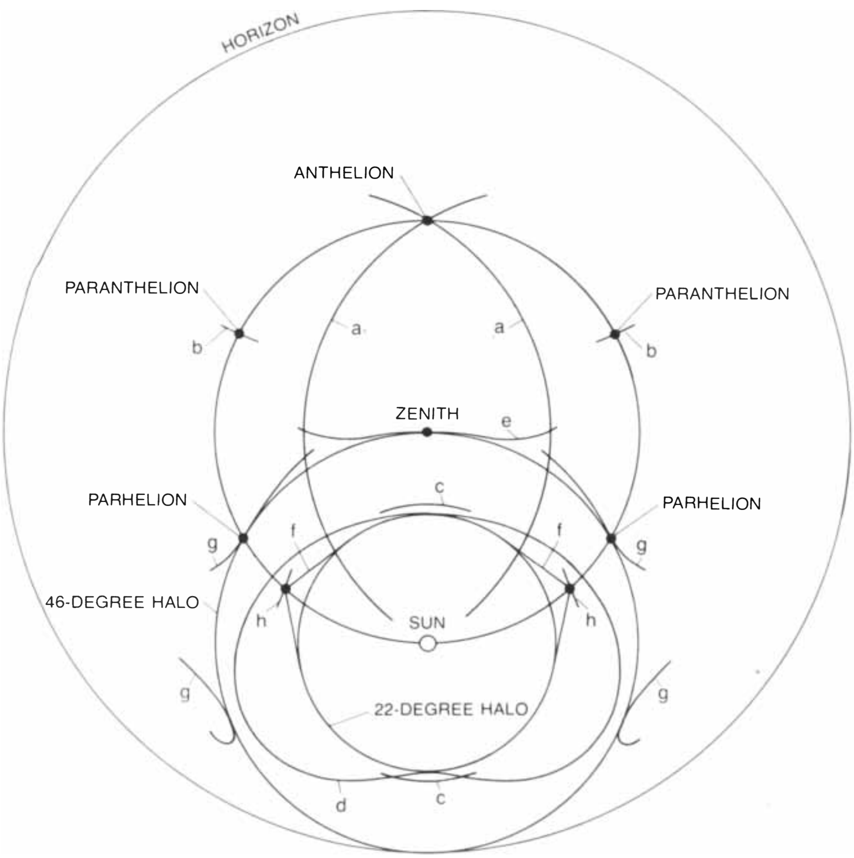
Most of the halos I have described are visible in the general direction of the sun. Looking the other way one finds several interesting phenomena. The colorless anthelion (counter-sun) and the two paranthelia (with the counter-sun) are often visible, looking like beads strung out along the parhelic circle. They also can appear when the parhelic circle is absent. Both are at the elevation of the sun. The anthelion is at an azimuth of 180 degrees relative to the sun, and the parhelia are at plus and minus 120 degrees.

Another class of halos arises from spinning crystals. Nine different arcs, attendant on either the 22-degree or the 46-degree halo, can be identified in this group. Because they are refractive phenomena they can be brightly colored.

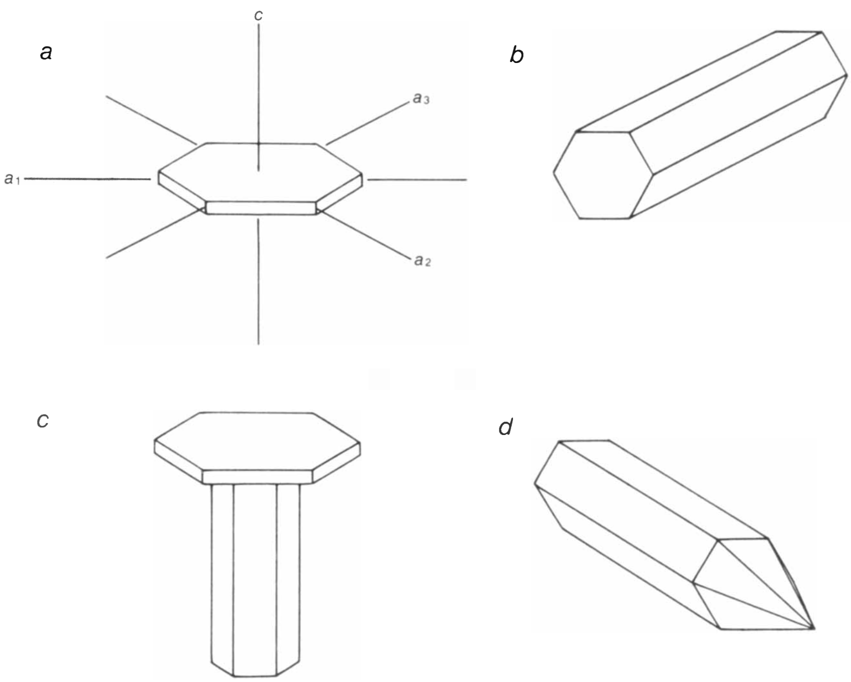
The mechanism begins with crystals of moderate size that are drifting down through the air. They quickly become oriented and reach a terminal velocity of about 20 centimeters per second. At that point the force of gravity is balanced by lift and viscous drag. Air flows smoothly around the crystal and remains relatively undisturbed after its passage. As the crystal grows, the grace-



COMMONEST HALOS occur in the general direction of the sun and are portrayed here in two ways: in a perspective (*top*) from outside the hemisphere and in a view (*bottom*) straight upward to the observer's zenith. The same atmospheric optical effects appear in both views.



OTHER EFFECTS can be seen in every part of the sky. As identified here by letters they are the anthelic arcs (*a*), the paranthelic arcs (*b*), the upper and lower Parry arcs (*c*), the circumscribed halo of the 22-degree halo (*d*), the upper tangent arc of the 46-degree halo (*e*), the supralateral arc of the 22-degree halo (*f*) and the supralateral arc of the 46-degree halo (*g*). An extra arc, the mesolateral arc (*h*), runs through the parhelion of the 22-degree halo. For reference some optical effects shown in the illustration on the preceding page are repeated.



ICE CRYSTALS usually responsible for atmospheric optical effects have these four forms. They are the plate (*a*), with its four axes indicated, the column (*b*), the capped column (*c*) and the bullet (*d*). Although the crystals are drawn to the same scale, they occur in a variety of sizes.

ful streaming becomes increasingly unstable. In time the flow acquires an entirely different character: it becomes turbulent. The crystal leaves a wake of vortexes and eddies, which cause it to spin as it falls.

The halos that result from spinning crystals are lateral arcs and tangent arcs. They can be divided readily into two categories: the arcs attending the 22-degree halo (the infralateral or Lowitz arc, the supralateral arc, the mesolateral arc and the circumscribed halo) and the ones that accompany the 46-degree halo (the infralateral arc, the supralateral arc and the upper and lower tangent arcs). There are actually nine arcs because at low solar elevations the circumscribed halo looks like an upper and lower two-tangent arc to the 22-degree halo. All of the arcs can be brightly colored because they are refractive phenomena.

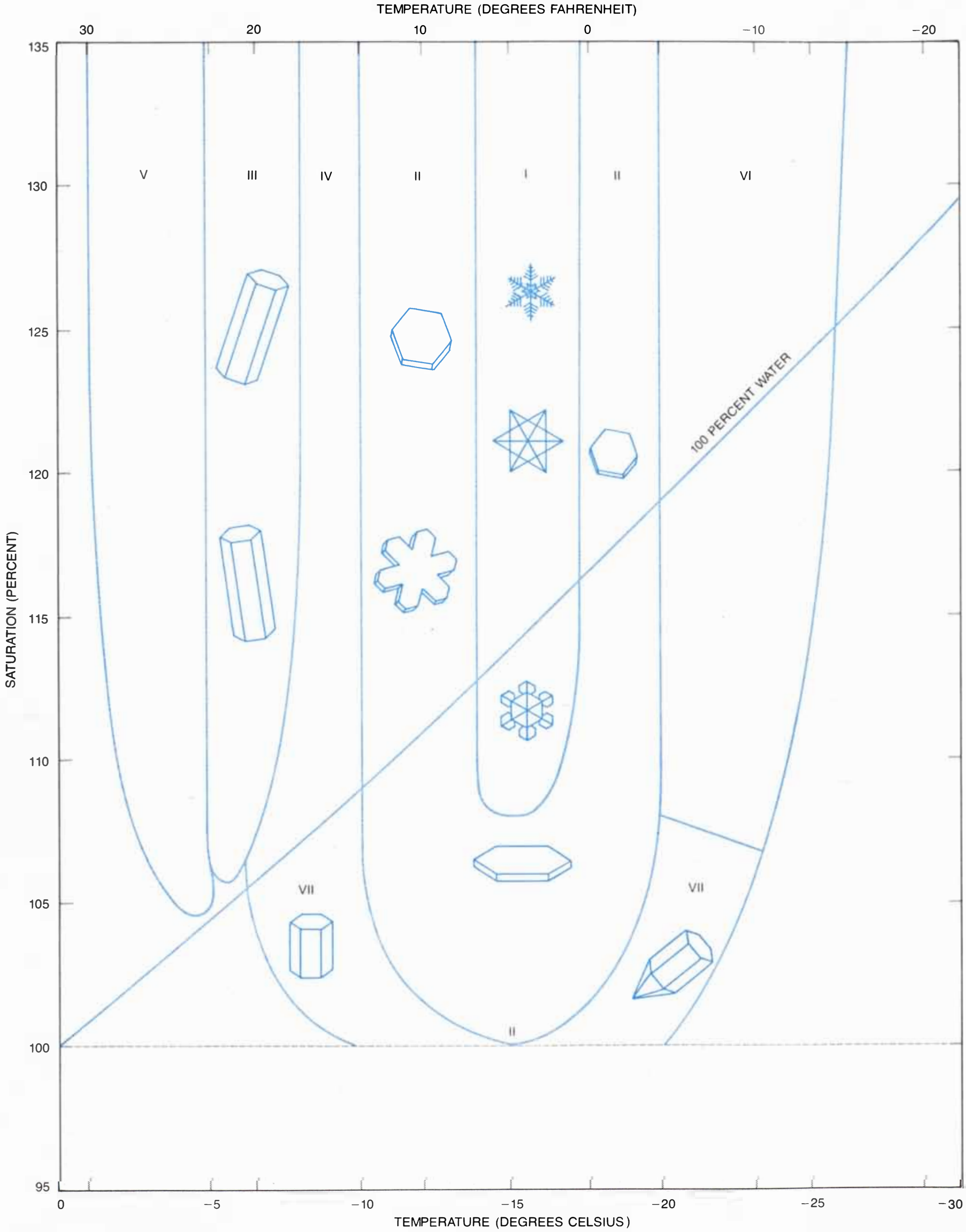
Some halos are formed below the horizon. In order to see them one must look down into the ice crystals. Until airplane travel became common such halos could be seen only from high mountaintops and cliffs. Bright halos can sometimes be seen below the horizon because of reflection from horizontally oriented ice faces. A sub-sun is frequently observed by people in airplanes and is a sure sign that sub-halos are about. I once saw (but alas did not photograph) a splendid sub-halo complex over Canada, consisting of a sub-sun flanked by two sub-parhelia.

The sun is not the only source of light for halos. At night the moon often has a halo. In northern regions where ice fogs occur pillars can be seen standing over street lamps and runway lights. When snowflakes lie horizontally, pillars and sub-lights can appear below the headlights of an automobile.

Many halos, some common and some rare, remain a puzzle. Anthelic arcs appear frequently, but attempts to explain them have not fully succeeded. They have been observed in so many forms that more than one crystal may be involved. A number of other halos are likewise not satisfactorily explained. Plainly much remains to be learned.

Research on halos is inching its way into the 20th century. So far, however, no work has been done beyond classical optics, and little progress has been made in the areas of polarization and diffraction. What is needed is a large number of observations, on which new theoretical work can be based.

This situation offers an opportunity to the interested skywatcher, who can do valuable research with modest equipment: a camera, a notebook and a sharp eye. Calibrated photographic observations are badly needed. The observer should record carefully the radius of any halo, the angular shapes and extensions of the arcs, the color characteris-

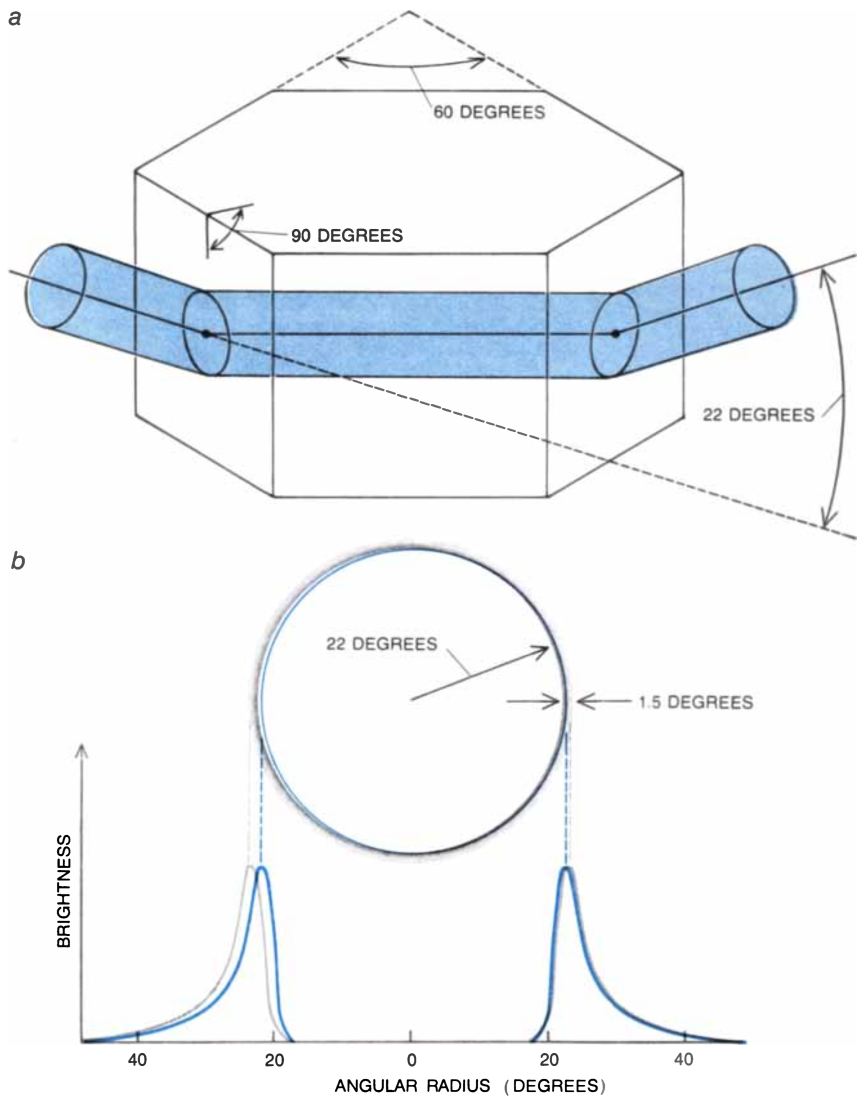


FORM OF CRYSTAL depends on the air temperature and the degree of water saturation of the ice. When the saturation is more than 108 percent, the temperature alone determines the type of crystal. The Roman numerals delineate regions with different forms, which can best be followed by reading the Celsius temperature scale from left to right. Irregular forms are not shown. The forms are irregular needles at minus 3 degrees (*V*), regular needles at minus 7 degrees (*III*), cups or scrolls at minus 9 degrees (*IV*), plates at minus 12 degrees (*II*), snowflakes at minus 15 degrees (*I*), plates at minus 17 degrees (*II*) and irregular plates at minus 23 degrees (*VI*). When the saturation is below 108 percent, only plates and columns grow. At saturations above 140 percent crystals grow so fast that they accumulate rime.

tics and the period of observation. Photographs made through a Polaroid filter at several orientations to the vertical are most important. The filter's orientation for each photograph should be noted. The meteorological conditions at the time of the observations should be recorded, along with the date, the time, the altitude of the sun and the geographic location. Neither a large monetary grant nor a laboratory full of advanced equipment can compete with hundreds of energetic observers who are at the right place at the right time to see and report on a halo complex.

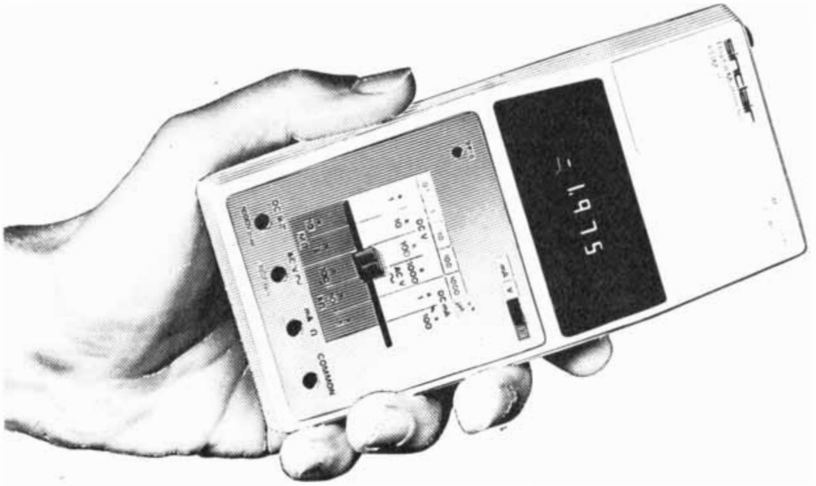
Only in a few fields of physics will a casual glance at the laboratory reveal anything about the experiments in progress. When a halo is sighted in the icy laboratory overhead, however, one immediately knows the temperature of the cloud; the state of the water; the size, shape and orientation of the ice crystals; the conditions of temperature and humidity in which the crystals are formed, and the subtle optics that are producing the halo. If several arcs are observed, even more is known.

Halos stir one's mind and soul, since they probe both the physical environment of the cloud and one's awareness and appreciation of the natural world. From the chaos of billions of pale, microscopic, angular crystals of ice nature spins a colorful fabric of expansive, graceful curves. All of them are accessible to the observer who takes the time to look up.



MECHANISM OF 22-DEGREE HALO is depicted. The halo is formed by the refraction of sunlight passing through the 60-degree faces of ice crystals. The effect is shown here (*a*) for a single crystal. The average deviation (the angle between the incident ray and the emergent ray) is 22 degrees, which is the mean radius of the halo. The halo is circular because the light is passing through billions of randomly oriented crystals. Because of dispersion (*b*) the 22-degree halo is actually a continuum of overlapping halos, the smallest one red and the largest violet. The inner edge is reddish and sharply defined; the outer edge is bluish-violet and much fuzzier.

The Sinclair PDM35.
A personal digital multimeter for only $49.95



Technical specification

DC Volts (4 ranges)
Range: 1 mV to 1000 V.
Accuracy of reading 1.0% ± 1 count.
Note: 10 MΩ input impedance.

AC Volts (40 Hz-5 kHz)
Range: 1 V to 500 V.
Accuracy of reading: 1.0% ± 2 counts.

DC Current (6 ranges)
Range: 1 nA to 200 mA.
Accuracy of reading: 1.0% ± 1 count.
Note: Max. resolution 0.1 nA.

Resistance (5 ranges)
Range: 1Ω to 20 MΩ.
Accuracy of reading: 1.5% ± 1 count.
Also provides 5 junction-test ranges.

Dimensions: 6 in x 3 in x 1½ in.
Weight: 6½ oz.
Power supply: 9 V battery or Sinclair AC adapter.
Sockets: Standard 4 mm for resilient plugs.
Options: AC adapter for 117 V 60 Hz power. De-luxe padded carrying wallet. 30 kV probe.

Now everyone can afford to own a digital multimeter

A digital multimeter used to mean an expensive, bulky piece of equipment.

The Sinclair PDM35 changes that. It's got all the functions and features you want in a digital multimeter, yet they're neatly packaged in a rugged but light pocket-size case, ready to go anywhere.

The Sinclair PDM35 gives you all the benefits of an ordinary digital multimeter – quick clear readings, high accuracy and resolution, high input impedence. Yet at $49.95 it costs less than you'd expect to pay for an analog meter!

The Sinclair PDM35 is tailor-made for anyone who needs to make rapid measurements. Development engineers, field service engineers, lab technicians, computer specialists, radio and electronic hobbyists will find it ideal.

With its rugged construction and battery operation, the PDM35 is perfectly suited for hand work in the field, while its angled display and optional AC power facility make it just as useful on the bench.

What you get with a PDM35

3½ digit resolution.
Sharp, bright, easily read LED display, reading to ± 1.999.
Automatic polarity selection.
Resolution of 1 mV and 0.1 nA (0.0001 μA).
Direct reading of semiconductor forward voltages at 5 different currents.
Resistance measured up to 20 MΩ.
1% of reading accuracy.
Operation from replaceable battery or AC adapter.
Industry standard 10 MΩ input impedance.

Compare it with an analog meter!

The PDM 35's 1% of reading compares with 3% of full scale for a comparable analog meter. That makes it around 5 times more accurate on average.

The PDM35 will resolve 1 mV against around 10 mV for a comarable analog meter – and resolution on current is over 1000 times greater.

The PDM35's DC input impedance of 10 MΩ is 50 times higher than a 20 kΩ/volt analog meter on the 10 V range.

The PDM35 gives precise digital readings. So there's no need to interpret ambiguous scales, no parallax errors. There's no need to reverse leads for negative readings. There's no delicate meter movement to damage. And you can resolve current as low as 0.1 nA and measure transistor and diode junctions over 5 decades of current.

The Sinclair credentials

Sinclair have pioneered a whole range of electronic world-firsts – from programmable pocket calculators to miniature TVs. The PDM35 embodies six years' experience in digital multimeter design, in which time Sinclair have become one of the world's largest producers.

Tried, tested, ready to go!

The Sinclair PDM35 comes to you fully built, tested, calibrated and guaranteed. It comes complete with leads and test prods, operating instructions and a carrying wallet. And getting one couldn't be easier. Just fill in the coupon, enclose a cheque/PO for the correct amount (usual 10-day money-back undertaking, of course), and send it to us.

We'll mail your PDM35 by return!

For Instant Service, CALL TOLL FREE: 1-800-528-6050, Ext. 1052.
Ariz. Res. Call Collect 602-955-9710 or send coupon:

Starshine Group.
Dept. 293, 924 Anacapa Street, Santa Barbara, Calif. 93101
Please promptly send ______ Sinclair PDM35 Digital Multimeter(s) @ $49.95 (plus $3 shipping and insurance) each. I have the option of returning it within 1 week of receipt for an immediate refund.
☐ Also Send AC Adaptor(s) @ $4.95 ea.
☐ Also Send Deluxe Padded Carrying Case @ $4.95 ea.
☐ Check or Money order Encl. (Calif. Res. add 6% Sales Tax)
☐ Charge to The Credit Card Checked Below:
☐ BankAmericard/Visa ☐ MasterCharge (Interbank # ______)
☐ American Express ☐ Carte Blanche ☐ Diners Club
Credit Card # ______
Exp. Date ______
Name ______
City/State/Zip ______
Signature ______

THE AMATEUR SCIENTIST

The physics and chemistry underlying the infinite charm of a candle flame

by Jearl Walker

I grew up with candles. They provided the illumination that held back the night in the midst of a big Texas storm. I mostly remember lying in bed before going to sleep, hypnotized by the flame and wondering how the candle burned. I had seen other things burn, but they usually did so much faster. The candle burned so slowly that the wick lasted all night or even for two nights.

Other features of the candle intrigued me almost as much. Why was a candle made of wax rather than some other material? Why was most of the flame yellow and a small region at the base of it blue? Why was there a dark inner cone between the wick and the yellow part of the flame? Why did some candles smoke, others burn nicely and others flicker weakly? Finally, why was the soot from a smoky candle black whereas a white vapor came from a candle that had just been blown out?

These are all old questions, and surely some of them have old answers. The fact remains that several of them have yet to be answered with any certainty. In this piece I shall describe how you can make your own candles and how you might explore the physics and chemistry of a burning candle. Answers to some of the basic questions will emerge. In part I shall be retracing the path of investigation followed by Michael Faraday, who was fascinated by both the simplicity and the complexity of candle flames.

In addition to the paraffin from which my grandmother made her candles many other materials have been employed: tallow, stearin, spermaceti (from the sperm whale) and beeswax. Today you can buy paraffin in a grocery store or a hobby shop. Often it is already mixed with stearin (stearic acid) to raise the melting point of the paraffin and therefore provide a stronger and slower-burning candle.

Wicks are also readily available. They are commonly made of plaited cotton yarn that has been mordanted (pickled in a substance that inhibits smoking). The treatment is necessary because normal cotton would burn too fast and smoke. The idea is to make the wick burn in such a way that it tends to lean out of the flame. Then the inorganic salts of the cellulose in the wick combine with the phosphates and borates of the mordanting salts for an ashless decomposition.

Making a candle is not hard. I was shown how by Liz Knepp of Cleveland, but you will find similar procedures and more details in the many books on making candles. Set up a double boiler so that water is boiling in the bottom part and paraffin is heating in the top part. Monitor the temperature of the paraffin with a wax or candy thermometer. If the paraffin has not already been mixed with stearin for strengthening, it will melt in the temperature range from 48 to 74 degrees Celsius (118 to 165 degrees Fahrenheit). If the stearin has not been included, add an amount equal to about 5 percent of the volume of the paraffin. The stearin raises the melting point and will prevent the candles from burning too fast. Church candles are often pure stearin. Do not overheat your paraffin in the boiler and do not expose it directly to an open flame. If the paraffin should catch fire, douse it not with water but with baking soda.

While the wax is melting you should prepare the candle mold. Depending on the shape you want, you can use many types of metal containers, avoiding any from which the removal of the candle would be difficult. Punch a hole in the bottom of the mold. Draw an appropriate length of wick through the hole and tie it off with a knot. Lay a pencil across the open end of the mold and tie the other end of the wick to it so that the wick is taut and straight in the mold. Anoint the inside of the mold with a silicone spray to facilitate the removal of the candle.

The successful burning of your candle will depend in part on the proper selection of the wick. Too wide a wick will result in a smoky flame. Too narrow a wick will yield a weak flame that may die out. A rough rule of thumb is that the wick should be chosen according to the diameter of the candle. If the diameter is from one to three inches, use 15-ply wicks; for four-inch or small tapered candles use 24-ply wicks and for larger candles use 30-ply wicks.

When your wax has melted, carefully pour it into the mold. Let it cool for a while and then push a long, thin screwdriver (or some other tool with a similar shape) down the long axis of the candle near the wick to eliminate air bubbles and promote a uniform contraction of the cooling wax. Fill the hole left by the screwdriver with more melted wax. You might have to repeat this procedure several times to allow sufficient contraction. Wait until the next day and then invert the mold, cut the knots on the wick and tap the candle from the mold. You then have a candle with which you can begin to experiment.

Faraday, whose name is usually associated with his far-reaching work on electricity and magnetism, also did the classical work on the physics and chemistry of candles. In 1860 and 1861, as part of his celebrated Christmas performances for a "juvenile auditory" at the Royal Institution, he gave six lectures designed to show the vast amount of science in the deceivingly simple candle. In trying to understand why a candle burns I have repeated some of Fara-

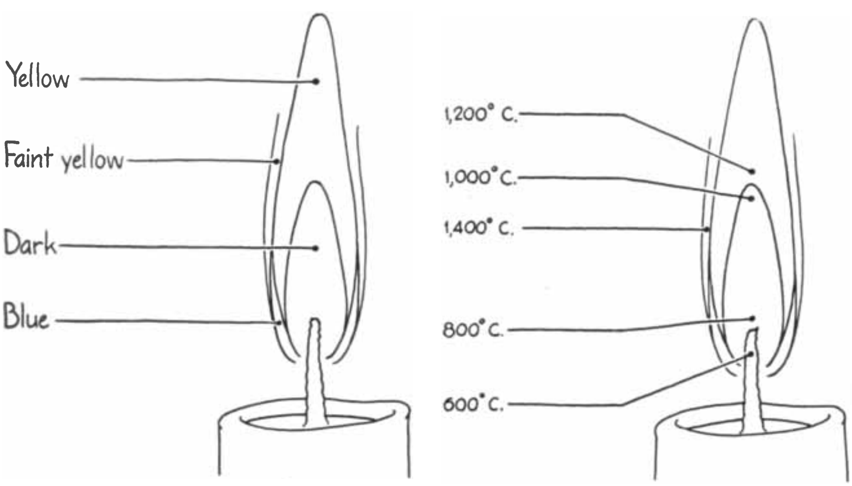


Diagram of colors and typical temperatures of a flame day's experiments, added a few and investigated current theories of flames. Except for modern theories of molecular and atomic emissions the elucidation of the candle flame has progressed surprisingly little since Faraday's time.

Once you light your candle you will quickly notice that the heat radiated from the flame melts the wax near the wick. The liquid wax is pulled up the wick by capillary attraction, which is to say that the attractive molecular forces between the wax molecules and the wick molecules pull the wax upward. Once the wax climbs to near the top of the wick it is vaporized. The released hydrocarbon molecules play a variety of roles in the flame.

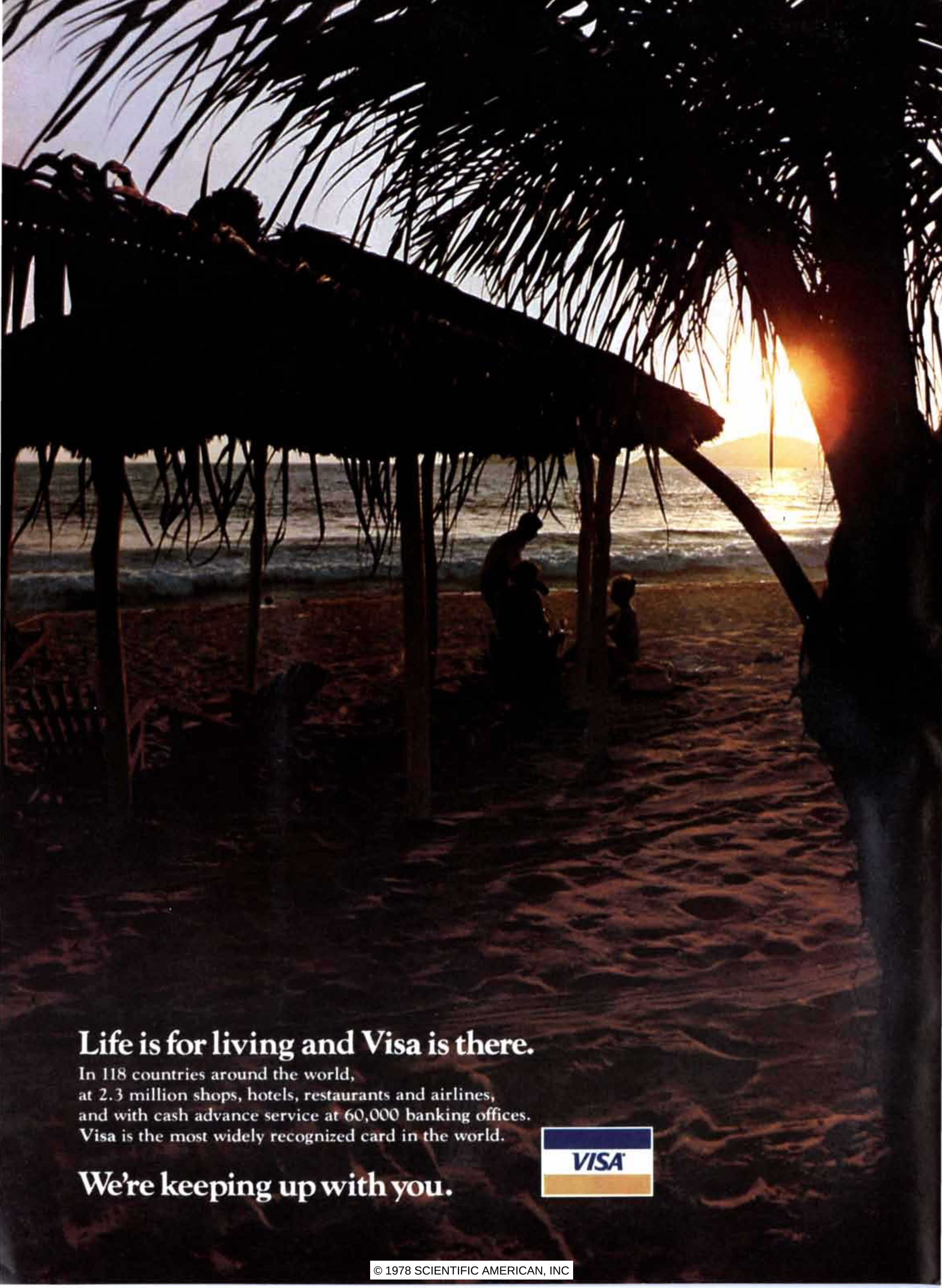
The flame has the interesting features I noticed as a child. Just above the wick you see a dark cone that is topped by the yellow region responsible for most of the light. On the sides of the flame and near the wick are the blue regions. Depending on the diameter of the wick and the height of the flame, the blue regions may be only a fraction of the height of the flame or may reach nearly to the top of the yellow region. It is surprising that the vaporized fuel molecules give rise to these three distinct regions instead of a single region of uniform color.

In the modern classification of flames the type of flame produced by a candle is called a diffusion flame. Another familiar type is called a premixed flame; its commonest examples are seen in the Bunsen burner and the gas stove. In a diffusion flame the rate of combustion is determined by the rate at which the gases diffuse through each other, whereas in a premixed flame the gases are mixed prior to the burning and the rate of combustion depends on the flow rate. For example, stove gas is mixed with air before it reaches the burner outlets. Further classifications of flames include the possibility of turbulence, but here I consider only a nonturbulent diffusion flame.

Perhaps surprisingly the highest temperatures in a flame are not found near the wick. The top of the wick is actually a comparatively cool place, its temperature being only about 600 degrees C. As you consider points progressively higher the temperature rises until it is about 1,200 degrees C. in the central portion of the yellow region. Somewhat higher temperatures, about 1,400 degrees C., lie off center and on the edge of the yellow flame. This nonuniform temperature distribution through the flame is one reason why the details of the light emission from the flame are so difficult to ascertain.

The relatively cool region just above the wick is the dark cone. There the released molecules of fuel are insufficiently heated and have so little oxygen that little or no light is emitted. What form the released molecules take in this region—how they break down into single

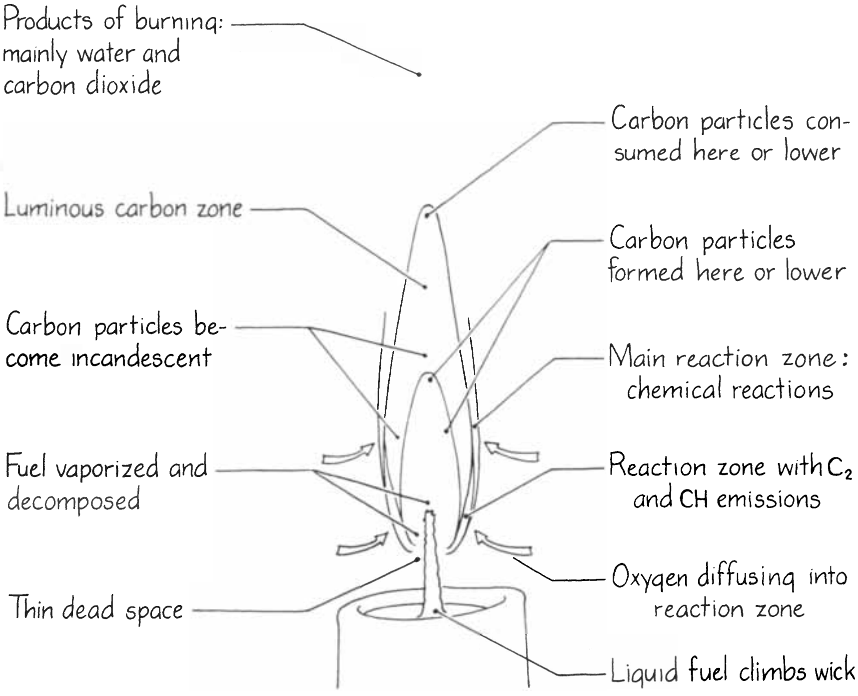
A candle flame


Life is for living and Visa is there.
In 118 countries around the world,
at 2.3 million shops, hotels, restaurants and airlines,
and with cash advance service at 60,000 banking offices.
Visa is the most widely recognized card in the world.
We're keeping up with you.
VISA

atoms or diatomic molecules and where they combine to form solid carbon particles—is not understood. One of the most curious features of the flame is that although vaporized fuel is found just above the wick, at no point in the flame does vaporized fuel ever come in contact with any significant amount of oxygen. Contrary to my own intuitive picture of the flame, neither the heat nor the light of the flame results from a simple oxidation of the fuel vaporized from the wick.

The blue regions of the candle flame are part of what is called the reaction zone. In that zone the large hydrocarbon molecules vaporized from the wick are broken down into smaller molecules, which then react chemically with one another and with the oxygen diffusing in from the air outside the flame. The bluish light is primarily due to the emission of two excited molecules, molecular carbon (C_2) and a hydrocarbon (CH), that are produced in the chemical reactions. These molecules, instead of being excited thermally by the hot environment, may actually be produced in their excited states by the chemical reactions creating them. Thus the blue light may not be an emission resulting directly from the heat.

Whereas single atoms emit light at particular wavelengths, molecules emit light in what are called band emissions, which are rather closely spaced groups of wavelengths. Unless you were to examine the emitted light by greatly spreading the spectrum, a molecular-band emission would appear to extend uniformly over a large range of wavelengths compared with what you would find from single atoms. The strongest emission of the CH molecule lies in a band around a blue wavelength (432 nanometers), with other emissions occurring in bands at shorter wavelengths in the deeper blue. The C_2 molecule strongly emits in what are called the Swan bands, a compact system of bands in the green, with less intense contributions lying toward the blue and deep blue. In a candle flame the net effect of these molecular emissions is a bluish emission from at least the lower portion of the reaction zone.



Processes in the flame

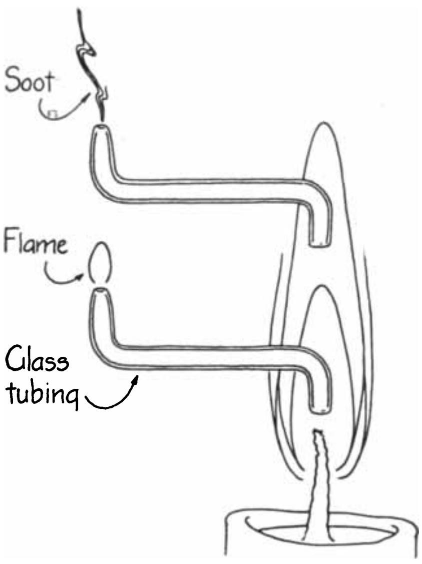
To me the most interesting portion of the candle flame is the yellow region, which is called the carbon zone or the luminous zone. There solid carbon particles are heated to incandescence by the hot gases and by heat radiated from the reaction zone. It is this incandescence that produces the yellowish light. In actuality the full visible spectrum is emitted, but the emission in the yellow is more intense and dominates one's perception of the light. The emission is not like molecular or atomic emissions, because the full spectrum is produced rather than individual colors. You might have seen something similar if you have ever heated a poker white hot. The poker was then so hot that the thermal radiation from it spanned the entire visible spectrum, emitting all the colors and therefore giving you a net perception of white light.

The solid carbon particles in the luminous zone are from 10 to 200 nanometers in size, most of them no larger than 50 nanometers. If your candle smokes, solid carbon particles are being released by the flame; most of them are larger than 50 nanometers. The first person to explain the yellow light in terms of incandescent particles of solid carbon was Humphry Davy, Faraday's mentor. Although Davy presented his explanation over a century ago, the mechanism that gives rise to the particles is still not understood.

One popular hypothesis pictures the fuel of large hydrocarbon molecules as being first broken down into carbon atoms or diatomic carbon molecules and then somehow nucleating on some agent to form the solid particle. Another hypothesis is that the hydrocarbons first aggregate to form the solid particle and then lose their hydrogen. In either case a key requirement for the formation appears to be the deficiency of oxygen in the dark zone above the wick. Whatever the true mechanism, the particles of solid carbon are formed by the time they reach the luminous zone, become incandescent and produce the candle's delicate yellow flame.

As the particles rise through the luminous zone they are consumed by reacting with water and carbon dioxide (the principal products of the flame) to yield carbon monoxide. In a smokeless flame the particles are totally consumed by the time they reach the top of the flame. Otherwise they are released as soot. Sooting usually occurs if the wick is too large in diameter, so that it transports upward too much fuel to be consumed in the flame.

Now for the experiments. One of the simplest of Faraday's demonstrations was to place the candle in strong sunlight and examine the shadow cast by the flame. The darkest portion of the shadow came from the luminous region, which is the brightest area of the flame itself. The darkness of the shadow results from the fact that the collection of solid particles is densest there. To update Faraday somewhat I played a beam from a low-power (20-milliwatt)



Effects with glass tubing

helium-neon laser over the flame, expecting to see laser light scattered by the solid carbon particles. The scattering was not sufficiently intense, however, to be noticeable to the unaided eye.

Over a Bunsen burner or a gas stove bend a short length of narrow glass tubing as is shown in the lower left illustration on page 157. Taper one end and put the other in the lower portion of the dark zone of a candle flame. Vaporized hydrocarbons enter the tube and leave the tapered end as a stream that is transparent or, if the vapor condenses somewhat to form wax droplets, white. The vapor would have produced carbon particles if it had been allowed to rise into the luminous zone, but it can still burn if you light the free end of the tube.

Instead of burning the vapor in this fashion you can collect it in a cool glass or flask. The vapor will condense to form a thin coating of wax on the inner walls. Such a recapture of solid wax implies that the vapor entering the tube had not had a chance to complete the mechanism giving rise to the solid carbon particles in the luminous zone.

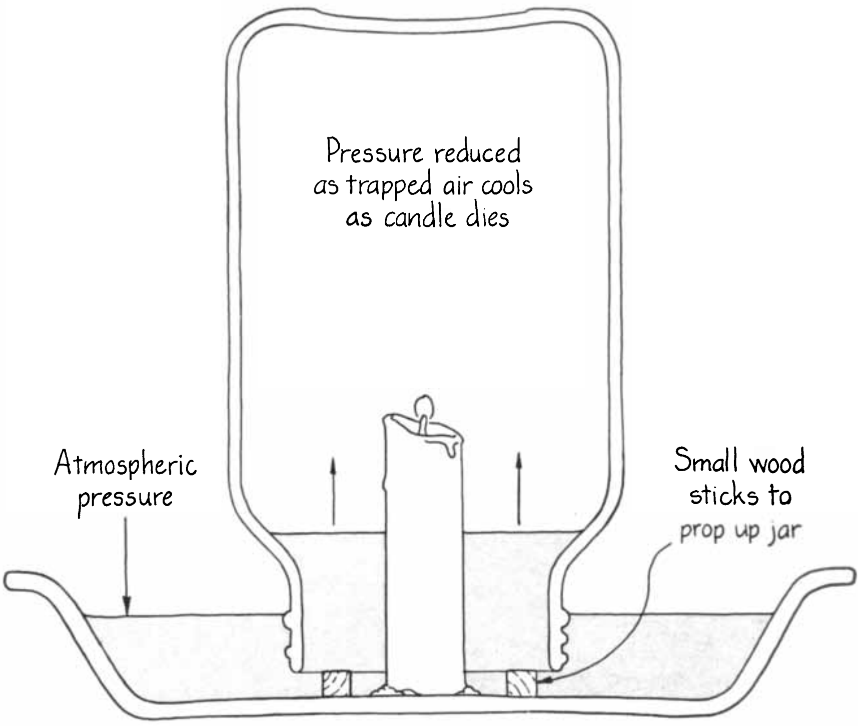
Next lift the flame end of the glass tubing to the luminous zone. Now soot rather than vapor will come from the free end. The glass tubing is no longer collecting vaporized hydrocarbons but is diverting some of the solid carbon particles from the luminous zone. Since the carbon particles have not had a chance to be consumed by reacting with water and carbon dioxide, they leave the tubing as they are: just soot.

Any kind of obstacle, particularly a cool one, placed in the luminous zone can cause sooting because the obstacle interrupts the consumption of the solid carbon particles. Try a straightened paper clip. It is immediately covered with soot. A piece of wire gauze placed across the yellow region can fully interrupt the burning of the carbon particles and eliminate the flame higher than the gauze.

Another demonstration by Faraday showed how combustible the vaporized hydrocarbons are. Carefully blow out the flame of a candle with a quick exhalation, but do not otherwise disturb the rising stream of white vapor that remains. Hold a lighted match above the wick and in the rising stream. If you do it soon enough, the flame will leap downward from the match to the wick and relight the candle. When you blow out the candle, some of the wax continues to be vaporized by the remaining hot gases and the wick. It is this rising stream of vapor that is combustible.

I tried pouring melted wax into a small Pyrex beaker, heating the beaker on an electric stove until the released vapor was noticeable and then lighting the vapor. The flame would catch inside the beaker but would quickly disappear because of the lack of oxygen in the small confines of the vessel. The vapor rising from the wick in the flame of a candle is combustible but does not burn and emit light because of the lack of oxygen there. Hence just above the wick the flame is dark. Only when the solid carbon particles are formed and become incandescent does the flame yield light.

You can demonstrate the rapid consumption of small particles, such as the



The inverted-jar experiment

carbon particles in the luminous zone, by blowing a fine combustible powder into the candle flame. Faraday used lycopodium powder, which is still available from scientific supply houses such as the Fisher Scientific Company (1231 North Monroe Street, Chicago, Ill. 60622) at about $1.90 per ounce. Alternatively you can use sifted flour. When you blow the powder lightly across the flame, the powder burns rapidly and with a noticeable sizzle. If you were to toss the clumped particles of unsifted flour through the flame, they would not burn as readily and perhaps would not burn at all. The smaller grains burn more readily because more surface area is exposed. This effect leads to a serious danger in mines and industrial plants having a large amount of fine airborne dust in them. In such an atmosphere an accidental spark can cause an extremely rapid combustion that sets off a violent explosion.

The temperature distribution of a candle flame can be roughly demonstrated by laying a sheet of paper across the flame. Just before the paper itself begins to flame, notice the blackened regions on it. They lie roughly in a circle around the center of the candle flame, indicating that the center of the flame is cooler than the outside edge.

I tried interfering with the mechanism of my candle by shielding the top of the paraffin with aluminum foil to block the heat radiated from the flame. Predictably the wax flow to and up the wick decreased and the flame waned. I could have accomplished the same thing if I had covered the paraffin with an infrared filter, that is, a filter blocking infrared radiation but passing visible radiation, because nearly all of the candle's radiation lies in the infrared. Less than .4 percent of the heat is radiated as visible or ultraviolet radiation.

The effect of electric fields on candle flames has long been noted. I placed my candle between the oppositely charged poles of a Wimshurst machine, the ancient hand-crank generator of high-voltage static electricity. The flame was attracted toward the negatively charged pole and repelled from the positively charged one.

Apparently when the flame is in the electric field that lies between the poles, the free electrons in the flame are deviated to the positive pole. Since you cannot see such an electron stream, the effect is unnoticed. The flame is then left, however, with an abundance of positive ions, which are relatively heavy and not as mobile as the electrons. The positively charged flame is therefore repelled from the positive pole and attracted toward the negative one.

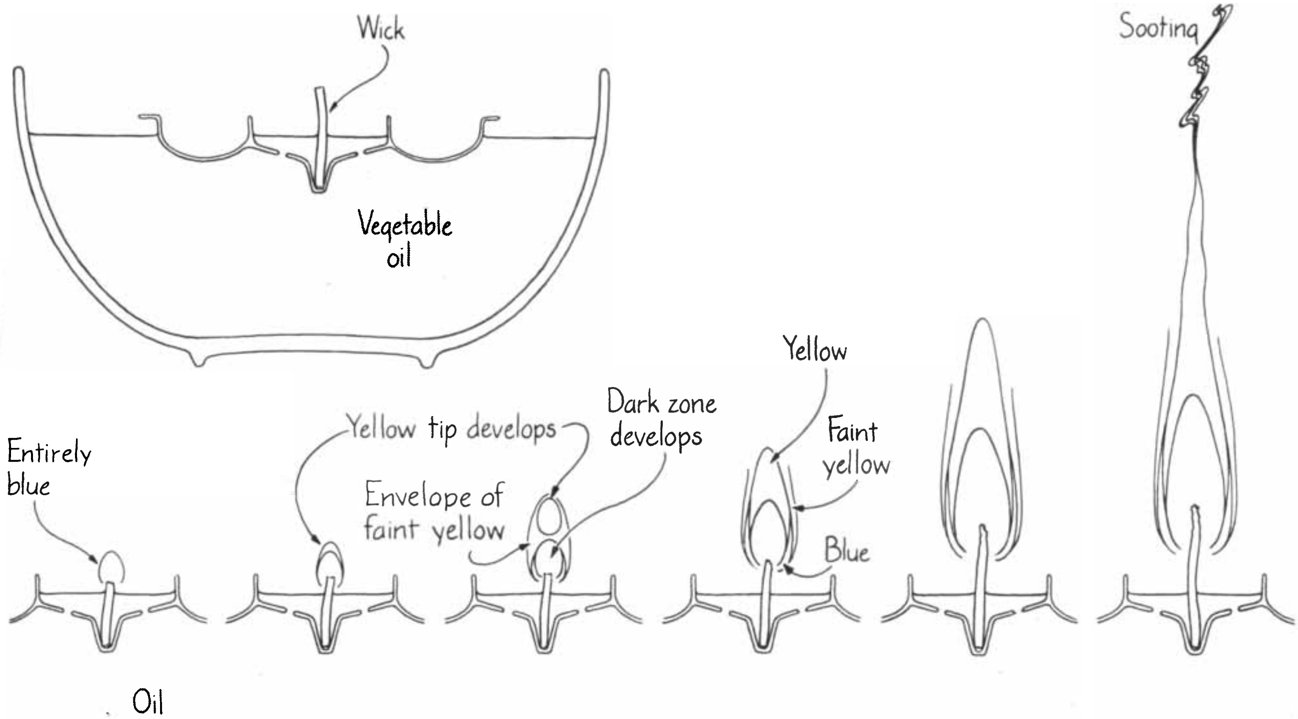
One of the products of the burning of the hydrogen released by the vaporized fuel is water. You can collect it by passing a cold spoon over the top of your candle's flame, taking care to avoid accumulating soot. To measure the amount of water put a funnel over the flame and collect the water in a cool flask by means of a rubber tube. You might want to measure how much water is produced by a given amount of depletion of the candle's wax.

A common demonstration involving the burning of a candle has occasionally been interpreted wrongly. Fix a candle in the bottom of a pan with a bit of hot wax. Pour water into the pan to a depth of about a quarter of the candle's height. Light the candle and put a large clear jar over it upside down. The mouth of the jar should be submerged but held above the bottom of the pan (perhaps by small pieces of wood) to allow water to flow into the jar. As the candle flame consumes the oxygen in the inverted jar and eventually is extinguished, watch the water level inside the jar. It will rise dramatically, perhaps even flooding the candle.

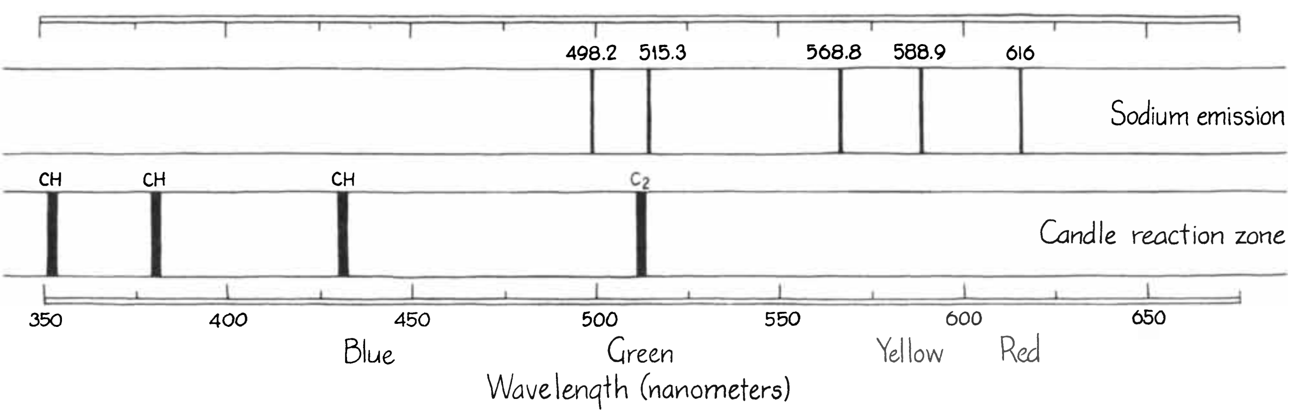
This effect has sometimes been erroneously attributed to the loss of the oxygen from the trapped air as the candle burns. That cannot be the reason, because the burning itself releases gases and vapors, notably carbon dioxide and water. The rise in the water level is actually due to the cooling of the air inside the inverted jar as the candle flame dims and dies out. When you first place the jar over the candle, the trapped air is heated; it expands, and some of it may even bubble out through the submerged mouth of the jar. As the flame dims, the air cools and contracts, and then the atmospheric pressure outside the jar forces the water up into the jar.

By making a series of candles that differ only in the diameter of their wick you can demonstrate the possible types of candle flames, from the kind that barely keeps burning because of poor fuel flow to the kind that smokes abundantly because there is too much fuel flow. To save time I used a commercially available candle called "The Uncandle," which consists of a wick placed in a plastic holder that floats on a pool of fuel such as vegetable oil. The floating holder has two compartments. The center section has holes to allow the oil to flow to the wick. A separate outer section has no holes, so that the platform remains afloat. When the wick is lit, the oil rises through it and is vaporized and burned in the flame.

When I pushed the platform down slightly, the flame dimmed. When I pulled the platform up slightly, the flame lengthened, eventually so much so that soot was produced. You could probably make such a floating platform



Variations among flames



Spectra from a sodium lamp and a candle flame

from an appropriate plastic lid. Whether you make one, buy one or just make a series of candles with different wicks, examine the characteristics of the flames. The dimmest possible flame is entirely a reaction zone and thus has only blue light. In a larger flame the luminous carbon region begins at the top of the reaction zone or just inside it and grows upward to split the reaction zone apart. In a higher flame the blue reaction zone is around the base, and the top of the luminous carbon zone is exposed directly to the outside air. Eventually the flame rises so high that the inner core of the luminous carbon zone cools; the carbon particles are no longer consumed by the time they reach the top of the flame, and so they are released as soot.

I checked the spectrum of my candle's emission with an inexpensive diffraction grating (50 cents) and with a glass prism (a few dollars). Both devices can be obtained from the Edmund Scientific Company (7778 Edscorp Building, Barrington, N.J. 08007). The luminous carbon zone dominated the light emission, and I saw a complete spectrum from red to deep blue, with no hint of any molecular emissions. To see the C_2 and CH emissions I had to use a spectrometer of the type often employed in introductory physics classes. Light from the candle entered a narrow slit and was directed by a lens onto either a diffraction grating or a prism, which deflected and dispersed the light to show individual colors. The spectrometer had a movable telescope with which to examine the dispersed light.

When I allowed light from the luminous carbon zone to enter the slit, I saw a complete visible spectrum again. When I allowed only the light from the blue reaction zone on one side of the flame to enter the slit, however, the full spectrum from the incandescent particles was not as bright, and I could distinguish several molecular emissions. Once I noted the angular positions of these emissions I calibrated the spectrum of colors in terms of wavelengths by removing the candle, replacing it with a sodium lamp and noting the angular positions of the sodium emissions. The wavelengths of sodium emissions are listed in reference books. I was then able to identify the molecular emissions as being from the C_2 and CH molecules by calculating the wavelengths associated with the angular positions I had recorded and then referring to previous work listing the wavelengths of the emissions from those molecules.

The construction of spectrometers has been described previously in "The Amateur Scientist." An excellent diffraction-grating spectrometer was described in September, 1966. With the spectrometer you would be able to photograph the flame's molecular emissions in great detail, including emissions in the near ultraviolet.

Many companies don't call AMP until they need electrical connectors. But why wait till then? If you involve us early enough, we can help you save money. And not just on connectors.

Here's an example:

Printed circuit boards solved a lot of problems for the electronics industry. And created others.

The biggest problem was connecting the thousands of circuits.

Using solder was slow, expensive and unreliable. So we found a better way—a solderless connector with our unique ACTION PIN contact.

Because 10,000 contacts can be inserted in an hour, production rates increased while costs decreased. And it's proved so reliable it was adopted as a standard in the major telephone system.

Can we produce the same kind of savings for you?

If your company manufactures electrical/electronic products, the answer is probably yes. But it's important for you to call us in early.

That's when our experience in developing connectors, switches and application equipment for hundreds of industries worldwide can help you the most.

Early Involvement.

At AMP, we call this approach to solving our customers' problems "Early Involvement." It's our better way. And it's what makes a 3¢ AMP contact worth so much more than 3¢.

Ask for a copy of our brochure "AMP Has A Better Way." Call (717) 564-0100, Ext. 8420.

Or write to AMP Incorporated, Harrisburg, PA 17105.

AMP
INCORPORATED

AMP has a better way.

Barcelona • Brussels • Buenos Aires • Frankfurt • Harrisburg • Helsinki • 's-Hertogenbosch • London • Luzern • Mexico City • Montreal
Paris • Puerto Rico • Sao Paulo • Stockholm • Sydney • Turin • Toronto • Tokyo

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

MATHEMATICAL GAMES

On Significance in Music. Susanne K. Langer in *Philosophy in a New Key.* Harvard University Press, 1957.

Noise: Sources, Characterization, Measurement. A. van der Ziel. Prentice-Hall, 1970.

"1/f Noise" in Music and Speech. Richard F. Voss and John Clark in *Nature,* Vol. 258, No. 5533, pages 317–318; November 27, 1975.

"1/f Noise" in Music: Music from 1/f Noise. Richard F. Voss and John Clark in *The Journal of the Acoustical Society of America,* Vol. 63, No. 1, pages 258–263; January, 1978.

NUCLEAR POWER, NUCLEAR WEAPONS AND INTERNATIONAL STABILITY

Moving Toward Life in a Nuclear Armed Crowd? Albert Wohlstetter, Thomas A. Brown, Gregory Jones, David McGarvey, Henry Rowen, Vincent Taylor and Roberta Wohlstetter. Pan Heuristics, April 22, 1976.

Nuclear Power Issues and Choices. Nuclear Energy Policy Study Group. Ballinger Publishing Company, 1977.

The National Energy Plan. Executive Office of the President: Energy Policy and Planning. Government Printing Office, April 29, 1977.

Nuclear Proliferation and Safeguards. Office of Technology Assessment, Congress of the United States, June, 1977.

MICROVASCULAR SURGERY FOR STROKE

Experimental Augmentation of Regional Cerebral Blood Flow by Microvascular Anastomosis. Jack M. Fein and G. Molinari in *Journal of Neurosurgery,* Vol. 41, pages 421–426, 1974.

Microvascular Anastomosis for Cerebral Ischemia. Edited by Jack M. Fein and O. H. Reichman. Springer-Verlag, in press.

Microvascular Neurosurgery—1977. Edited by Jack M. Fein. The C. V. Mosby Company, in press.

THE TRACKS OF MOVING CELLS

Cold Spring Harbor Conferences on Cell Proliferation, Vol. 3, Books A, B and C: Cell Motility. Edited by R. D. Goldman, T. D. Pollard and J. L. Rosenbaum. Cold Spring Harbor Laboratory, 1976.

Daughter 3T3 Cells: Are They Mirror Images of Each Other? Guenter Albrecht-Buehler in *The Journal of Cell Biology,* Vol. 72, pages 595–603; 1977.

The Phagokinetic Tracks of 3T3 Cells. Guenter Albrecht-Buehler in *Cell,* Vol. 11, No. 2, pages 395–404; June, 1977.

THE FOOD-SHARING BEHAVIOR OF PROTOHUMAN HOMINIDS

Primate Behavior and the Emergence of Human Culture. Jane B. Lancaster. Holt, Rinehart & Winston, 1975.

Earliest Man and Environments in the Lake Rudolf Basin: Stratigraphy, Paleoecology and Evolution. Edited by Yves Coppens, F. C. Howell, Glynn Ll. Isaac and Richard E. F. Leakey. University of Chicago Press, 1976.

Humankind Emerging. Edited by Bernard G. Campbell. Little, Brown & Company, 1976.

Human Origins: Louis Leakey and the East African Evidence. Edited by Glynn Ll. Isaac and Elizabeth R. McCown. W. A. Benjamin, Inc., 1976.

THE BIRTH OF MASSIVE STARS

The O Associations in the Solar Neighborhood. Adriaan Blaauw in *Annual Review of Astronomy and Astrophysics: Vol. 2,* edited by Leo Goldberg. Annual Reviews, Inc., 1964.

Several H II Regions in the Near Infrared. M. Beetz, H. Elässer, C. Poulakos and R. Weinberger in *Astronomy and Astrophysics,* Vol. 50, No. 1, pages 41–46; July 1, 1976.

Shock-Driven Implosion of Interstellar Gas Clouds and Star Formation. Paul R. Woodward in *The Astrophysical Journal,* Vol. 207, No. 2, Part 1, pages 484–501; July 15, 1976.

Sequential Formation of Subgroups in OB Associations. Bruce G. Elmegreen and Charles J. Lada in *The Astrophysical Journal,* Vol. 214, pages 725–741; 1977.

KIMBERLITE PIPES

Lesotho Kimberlites. Edited by P. H. Nixon. Lesotho Development Corporation, Maseru, 1973.

Occurrence, Mining and Recovery of Diamonds. A. A. Linari-Linholm. De Beers Consolidated Ltd., Kimberley, 1973.

Model of a Kimberlite Pipe. J. B. Hawthorne in *Physics and Chemistry of the Earth,* Vol. 9, pages 1–15; 1975.

Sub-cratonic Crust and Upper Mantle Models Based on Xenolith Suites in Kimberlite and Nephelinitic Diatremes. John B. Dawson in *Journal of the Geological Society, London,* Vol. 134, Part 2, pages 173–184; November, 1977.

INSECTS OF THE WATER SURFACE

Wave-making by Whirligig Beetles (Gyrinidae). Vance A. Tucker in *Science,* Vol. 166, No. 3907, pages 897–899; November 14, 1969.

Communication by Surface Waves: Mating Behavior of a Water Strider (Gerridae). R. Stimson Wilcox in *Journal of Comparative Physiology,* Vol. 80, pages 255–266; 1972.

Can *Halobates* Dodge Nets? II: By Moonlight? Lanna Cheng and James T. Enright in *Limnology and Oceanography,* Vol. 18, pages 666–669; 1973.

The Mechanoreceptive System of Prey Localization in *Notonecta,* II: The Principle of Prey Localization. Konrad Wiese in *Journal of Comparative Physiology,* Vol. 92, pages 317–325; 1974.

ATMOSPHERIC HALOS

Light and Colour in the Open Air. M. Minnaert, translated by H. M. Kremer-Priest. G. Bell and Sons, Ltd., 1959.

Field Guide to Snow Crystals. Edward R. LaChapelle. University of Washington Press, 1969.

Introduction to Meteorological Optics. R. A. Tricker. Elsevier North-Holland, 1971.

THE AMATEUR SCIENTIST

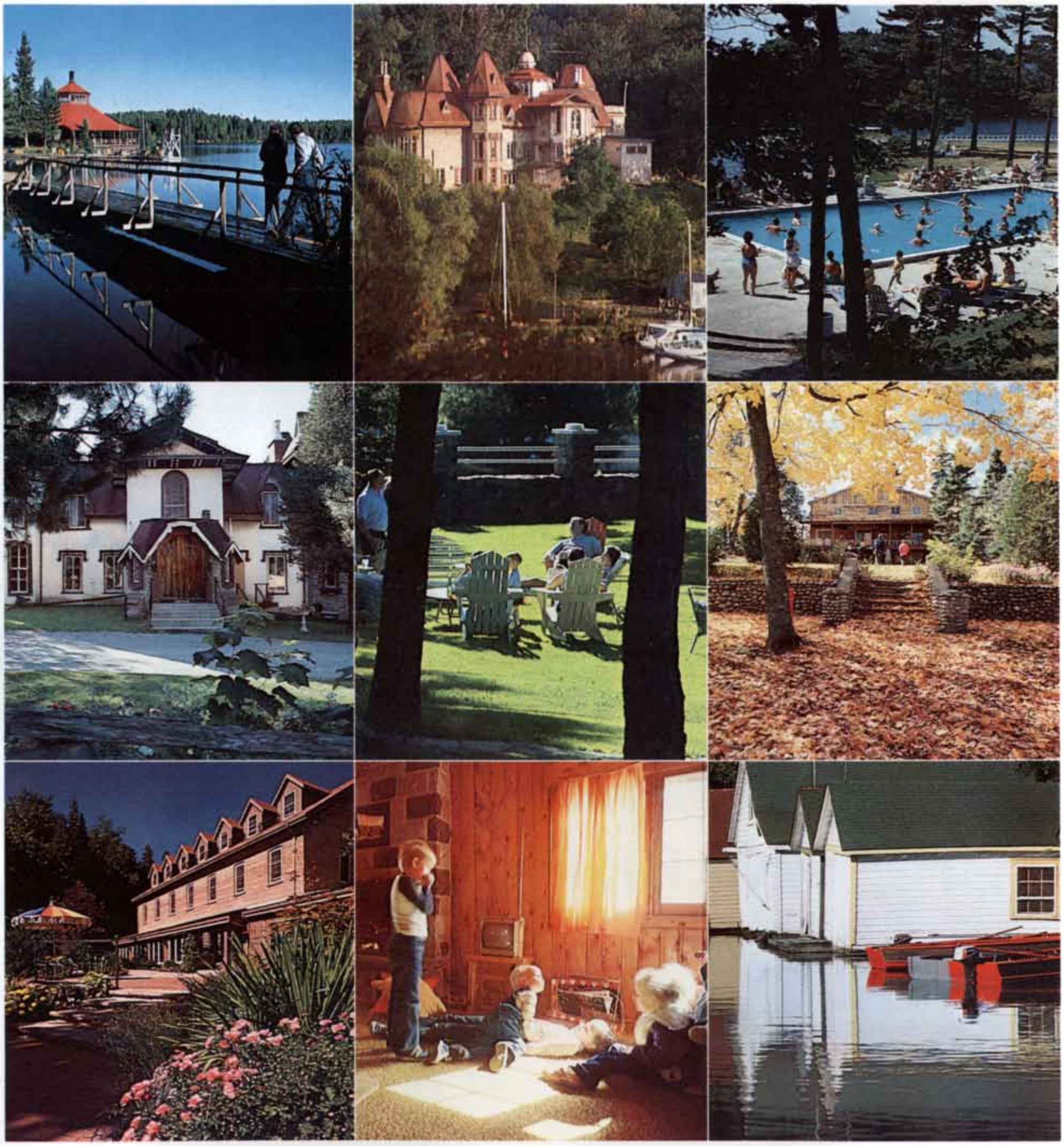
The Chemical History of a Candle. Michael Faraday. Collier Books, 1962.

Combustion and Flame. Robbin C. Anderson in *Journal of Chemical Education,* Vol. 44, No. 5, pages 248–260; May, 1967.

Flames: Their Structure, Radiation and Temperature. A. G. Gaydon and H. G. Wolfhard. Chapman & Hall Ltd., 1970.

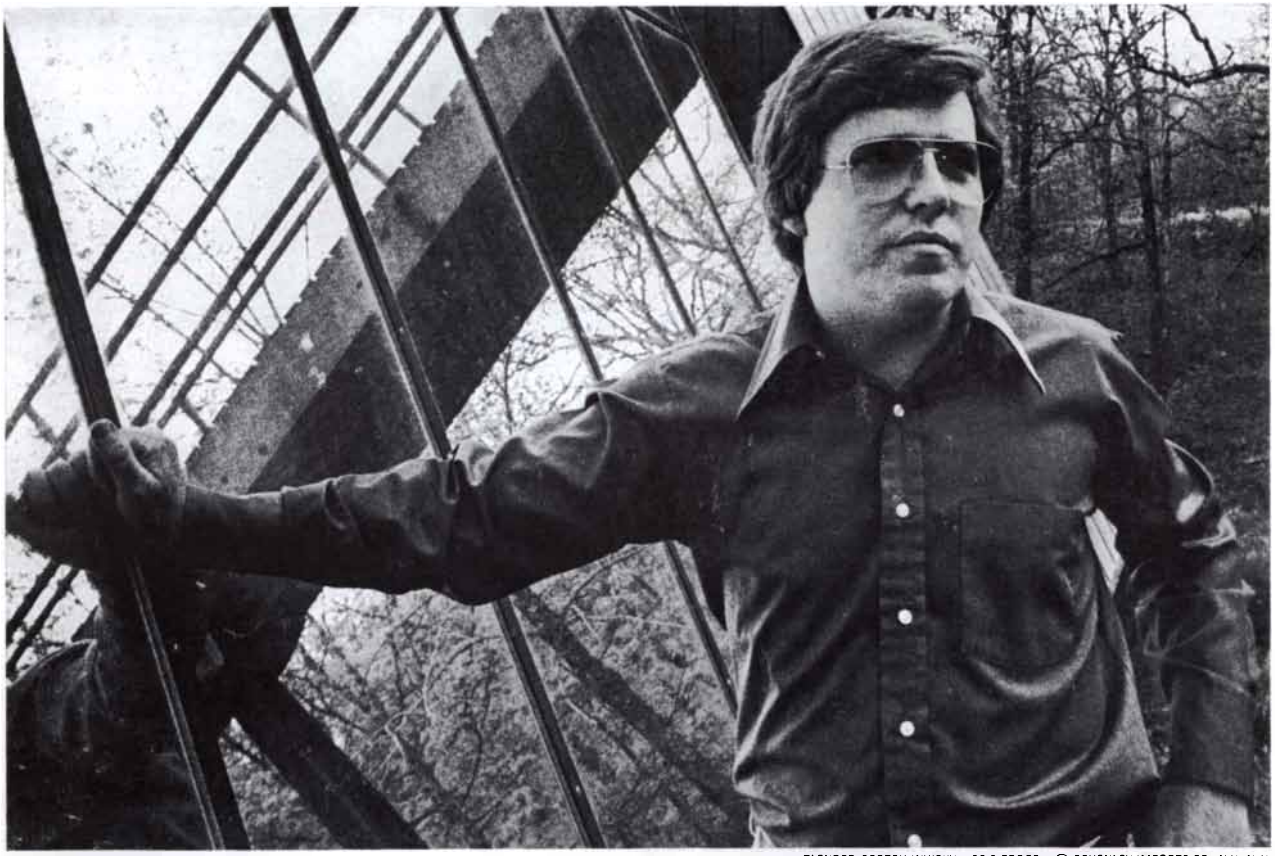
Creative Candlemaking: From the Simple Basic Candle to the Most Intricate Sculptural Form. Thelma R. Newman. Crown Publishers, Inc., 1972.

Flames: A Study in Molecular Spectroscopy. Richard A. Ashby in *Journal of Chemical Education,* Vol. 52, No. 10, pages 632–637; October, 1975.



DEWAR'S® PROFILES

(Pronounced Do-ers "White Label")



BLENDED SCOTCH WHISKY • 86.8 PROOF • © SCHENLEY IMPORTS CO., N.Y., N.Y.

JAMES LAMBETH

HOME: Fayetteville, Arkansas

AGE: 34

PROFESSION: Architect, solar expert

HOBBIES: Photography, writing, scuba diving.

MOST MEMORABLE BOOK: "The Art of Walt Disney" by Christopher Finch

LATEST ACCOMPLISHMENT: Designed one of the most technically and economically successful solar residences in the U.S.A., according to the American Institute of Architects Research Corporation in 1976.

QUOTE: "It is not enough to save energy if it is to be at the expense of beauty. We must maintain a balance between the two and satisfy our need for new energy sources in an aesthetically pleasing manner."

PROFILE: Intensely motivated yet casual and easygoing. A strong innovator who's fast emerging as a major creative force in solar design through his unique feel for combining form with function.

HIS SCOTCH: Dewar's "White Label"®