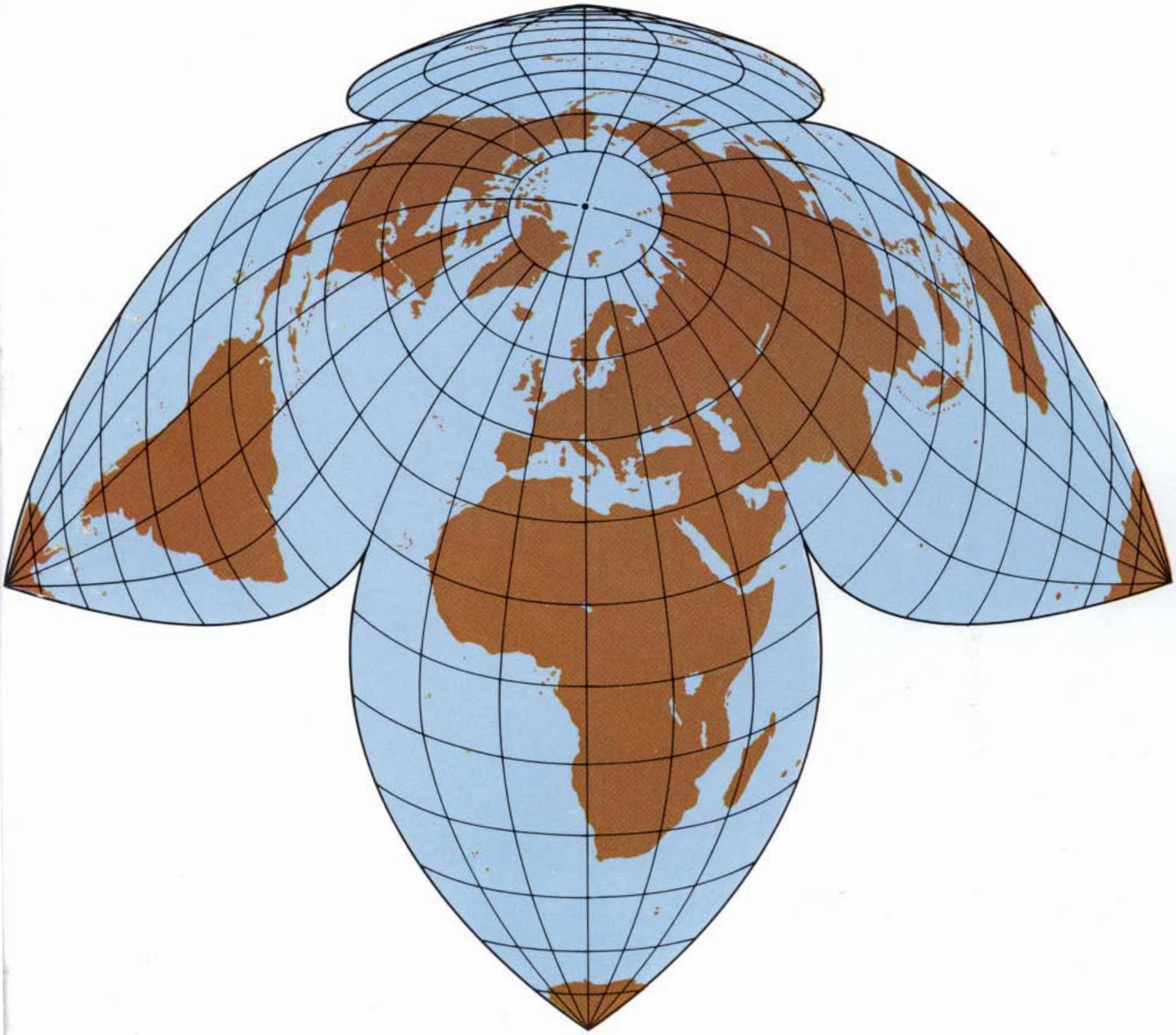
SCIENTIFIC AMERICAN

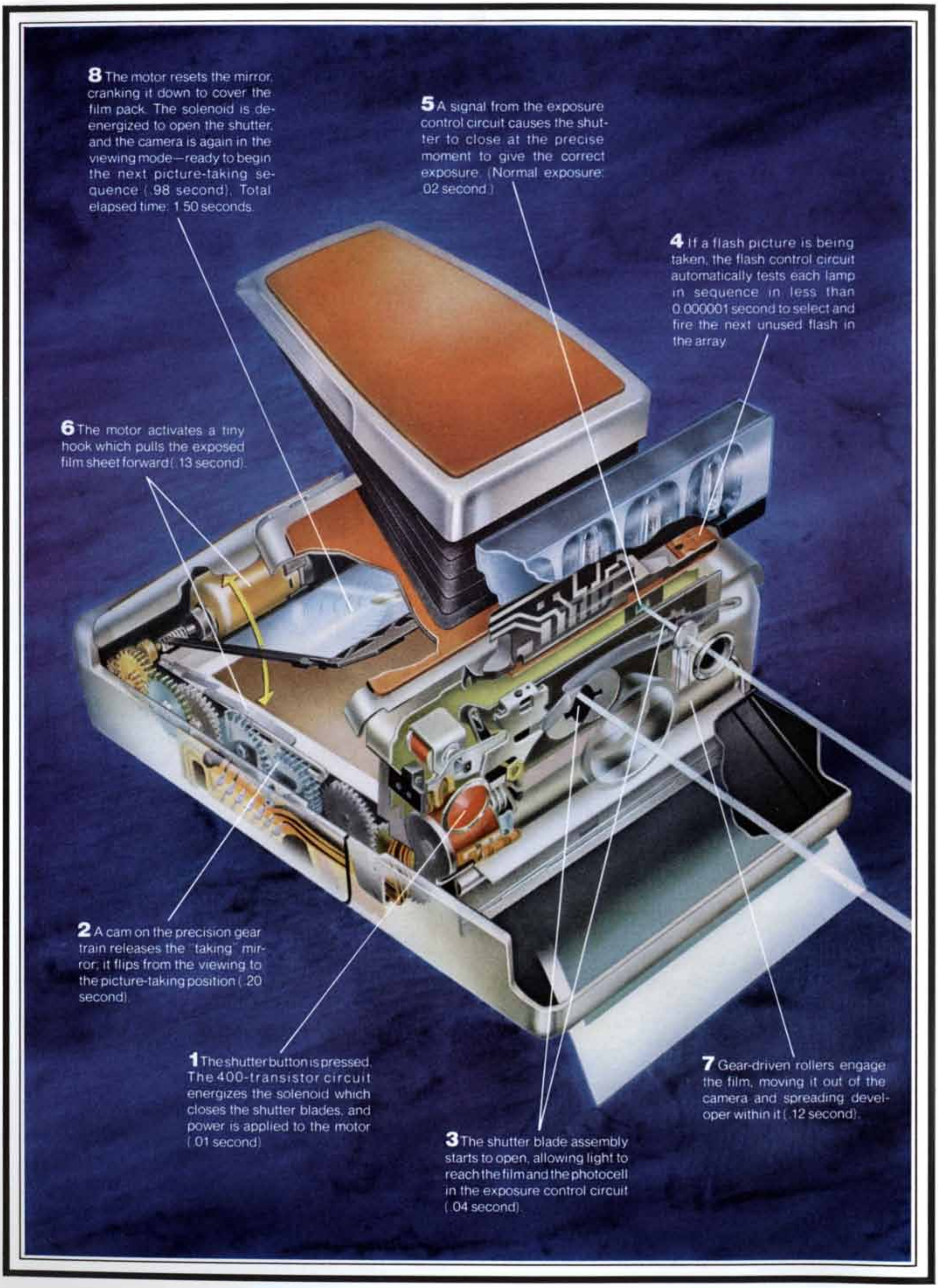


NOVEL MAP PROJECTIONS

$1.25

November 1975

THE 1.5 SECOND OPERATING SEQUENCE OF POLAROID'S SX-70.



We created an unprecedented electronic nerve system to control a complex mechanism...

More inventiveness went into the design of the SX-70 Land camera than any other amateur camera ever developed.

Consider what happens when the shutter button is pressed:

The shutter, which is open to allow viewing through the camera's single-lens reflex system, is closed by a solenoid, darkening the interior of the camera.

A 2-sided, hinged mirror is then released by jogging the camera's electric motor with a surge of current. The mirror swings up, driven by a spring.

The upper side, a Fresnel mirror which is part of the viewing system's light path, is now out of the way. The lower side, a plane mirror which has been sealing off light from the film, becomes part of the light path from lens to film.

The shutter starts to reopen, and the electric-eye circuit—which uses a fast, accurate silicon photocell—measures the ambient light and commands the shutter to start closing when most of the light needed for exposure has passed through.

Now the motor restarts. It moves a hook that pulls the exposed film into the grip of two rollers. The motor rotates one of the rollers via a gear train, driving the exposed film out of the camera and spreading developer uniformly within it.

During this process, the motor and gear train pull down the dual mirror, covering the next film and protecting it from light. The shutter reopens and the viewing path is restored.

All this takes place in 1.5 seconds.

If a flashbar is used, further circuitry comes into play:

Inserting the flashbar automatically puts another, smaller solenoid in control of the shutter blades. This solenoid engages a linkage connected to the focus wheel, thus setting the aperture. Now the distance of the subject from the camera—not ambient light—determines the size of the shutter opening, which is slightly larger than would normally be required for a simple follow-focus flash system. The exposure control circuit then closes the shutter when the proper portion of the flash output has passed through.

The camera has a scanning circuit which selects and fires the first unused flashbulb, so that a partially used flashbar automatically works when put back on the camera.

If the photographer inadvertently tries to take an eleventh picture (there are just 10 in a pack), another circuit prevents the flash from going off.

Still other circuits accomplish the following:

When a new film pack is inserted and the loading door is closed, the camera goes through a cycle that ejects the dark slide which covers the first film. And—if a flashbar is in place, it is prevented from firing.

The power source.

For all of these actions, the camera requires a power source. Designing one for the restricted confines of a pocket camera was a difficult problem, and the solution was ingenious: a compact, flat battery was made part of each film pack.

The battery supplies a nominal 6 volts and is constructed like a printed circuit, with four laminated flat cells. Its outer dimensions conform to that of the film pack,

Continued on next page.

and it is only .12″ thick. Thus, the battery has a high ratio of surface area to volume. This minimizes internal resistance, and has two desirable effects: It's possible to get large currents out of this small battery, since short bursts of high power are needed to accelerate the motor. Yet the voltage does not drop so low as to upset the logic circuits, which are designed to function from 6½ to 3 volts.

Everything possible was done to save energy so that the battery could be made as small as possible—even to details such as completely disconnecting the battery when the camera is closed.

The motor.

We developed a motor that could accelerate to 12,000 rpm in 8 revolutions and is more than 50% efficient. This helps conserve the battery.

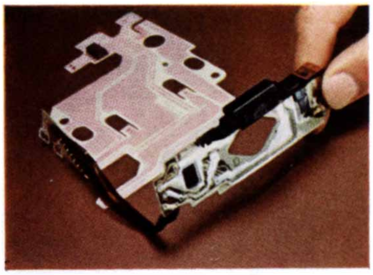
To start the motor, a control circuit turns on a big IC power transistor which offers almost no resistance to the battery's full current. The instant the motor pulls power there is a 2 ampere spike which drops to a steady .8 ampere as the picture is being driven out of the camera. When the picture is all the way out, a discrete transistor "switch" short-circuits the motor, stopping it in a mere 8 revolutions. The energy of dynamic braking is dissipated in this transistor and the motor windings.

The solenoids are also relatively heavy power consumers. The large solenoid draws about one ampere; the small one about .6 ampere. To conserve the battery's energy, we designed the solenoid drivers (transistors that act as heavy-duty switches) for a dual mode of operation: full power while the solenoids are pulling in and reduced power (about 10% of full current) while they are holding.

The circuitry.

Three electronic assemblies—the exposure control module, the motor control and the flash circuit—meet the camera's needs. All are interconnected by flexible, flat cable with printed-circuit conductors.

This miniaturized, solid-state system involving 400 transistors represents an unprecedented association of functions, combining amplifiers that handle picoam-



peres with amplifiers handling 2 amperes.

In developing the system, we mastered the art of integrated circuit design. We combined fundamentally different kinds of circuitry: digital and analog. (Such as the binary signals of the logic circuits and the analog signals of the photocell circuit.) We synchronized the operating sequences with an 8 kilohertz clock and overcame an electrically noisy environment.

For instance, the DC motor generates a lot of electrical noise due to commutation. Solenoids produce noise peaks, too. We timed the more sensitive electronic functions so that they would not coincide with the noise. As an example, the motor comes completely to rest after releasing the mirror, and only then is the signal from the photocell integrated in a capacitor.

We physically separated the IC's handling low-current photo-signals from the high-current solenoid driver circuits. And we put the motor control in a package of its own near the motor.

The IC chip that controls power was the greatest challenge of all. It's the central brain of the system. It receives exposure data, time-delay signals and transducer switch inputs. The chip processes all these in its logic section and commands the solenoids and motor to perform the appropriate operations.

A second IC chip contains the clock—an 8 kHz oscillator with 17 flip-flops forming a counting chain. The circuitry generates several time intervals. They range from a 40-millisecond delay during which the mirror comes to rest, to a 20-second period, the maximum time the shutter is allowed to stay open under low ambient light conditions.

We created a unique flash exposure control system that *first* sets the size of the aperture and *then* controls the time the shutter is open according to the amount of light reflected back from the scene.


THE LUXURY CAR THAT SHOWS YOU HAVE MORE THAN MONEY.
PSR-600
A luxury car has long been the symbol of having arrived.
The Volvo 164 shows you used your brains to get there.
For one thing, it shows you're too intelligent to buy a luxury car, and then pay extra to add the luxuries.
The Volvo 164 comes with air conditioning, automatic transmission, power-assisted steering, 4-wheel power disc brakes, power front windows and leather to sit on… all at no extra cost.
The Volvo 164 also comes with other things an intelligent person can respond to.
Like a responsive three liter engine with computerized fuel injection. Electronic sensors automatically monitor temperature, speed and altitude to determine the proper fuel mixture for varying driving conditions.
Even Volvo's front bucket seats show a deep-seated commitment to intelligent design. They're infinitely adjustable. You can firm or soften them against your back. When it's cold, the driver's seat heats up. Automatically.
Rust and rattles are not luxurious. So Volvo fuses the body into one solid unit with nearly 8,000 spotwelds. Rustproofing isn't just sprayed on. It's drawn into the metal with a powerful magnetic charge. And there are two separate coats of undercoating.
These are just a few of the things that make the Volvo 164 what it is.
A luxury car that shows you're too smart to be impressed by a luxury car price tag. Unless an intelligently thought-out car comes attached.
VOLVO 164
The luxury car for people who think.

Established 1845 **SCIENTIFIC AMERICAN** November 1975 Volume 233 Number 5

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: $15 PER YEAR, U.S., POSSESSIONS AND CANADA; $18 PER YEAR, ALL OTHER COUNTRIES. EUROPEAN SUBSCRIPTION COPIES MAILED C/O PUBLICATIONS DISTRIBUTION SERVICE, ST. KATELIJNEVEST 14, B2000 ANTWERP, BELGIË.

THE COVER

The illustration on the cover exemplifies a fundamental problem of mapmaking: How can the earth's surface (or a portion of it) be drawn on a plane so that all distances are accurately represented (see "Mathematical Games," by Martin Gardner, page 120)? The problem is essentially unsolvable, but over the past four centuries cartographers have devised many useful partial solutions. The most famous is the cylindrical projection used in the 16th century by the Flemish geographer Gerhardus Mercator. In the map on the cover the surface of the earth is represented in an orthographic projection on an interrupted hyperboloid, a projection devised by the late Erwin Raisz. This unusual projection is characterized by the cartographer Richard Edes Harrison, who renders it here, as "the most elegant of all world maps."

THE ILLUSTRATIONS

Cover by Richard Edes Harrison

Page	Source
26–34	Ilil Arbel
36	Carolyn Cohen, Brandeis University
37	B. Bullard, University of Oxford
38	George V. Kelvin (*top and third from top*); Iwao Ohtsuki, University of Tokyo (*second from top and bottom*)
39	George V. Kelvin
40	Carolyn Cohen, Brandeis University (*top and bottom*); George V. Kelvin (*middle*)
41	George V. Kelvin
42	Carolyn Cohen, Brandeis University (*top*); George V. Kelvin (*bottom*)
43	George V. Kelvin
44	Carolyn Cohen, Brandeis University (*top and middle*); Sugie Higashi-Fujime, University of Nagoya (*bottom*)
45	George V. Kelvin
46–47	Ralph Morse
48–53	Dan Todd
54	Ralph Morse

Page	Source
65	Carol Donner
66–67	Total Communications Industries, Ltd.
68–71	Carol Donner
72–78	Total Communications Industries, Ltd.
80–86	Tom Prentiss
88	EROS Data Center, U.S. Geological Survey
90–98	George V. Kelvin
103	B. V. Derjaguin, Institute of Physical Chemistry, U.S.S.R. Academy of Sciences
104	Ilil Arbel
105	Total Communications Industries, Ltd.
106	B. V. Derjaguin, Institute of Physical Chemistry, U.S.S.R. Academy of Sciences
107–108	Ilil Arbel
111	Eckhard H. Hess, University of Chicago
112–118	Total Communications Industries, Ltd.
121–125	Richard Edes Harrison
129–131	Jerome Kuhl

LETTERS

Sirs:

How delightful to learn from Howard C. Berg's article "How Bacteria Swim" [SCIENTIFIC AMERICAN, August] that the flagella of bacteria rotate like propellers! This opens up the possibility that evolution is capable of producing more advanced animals that are propelled through water or air by similar means.

Students of Oz are familiar with just such creatures. On the island of Orkland, in the Nonestic Ocean not far from the eastern coast of Oz, live the orks—enormous birds whose tiny wings would be inadequate for flight if they were not assisted by a propeller at the rear. I enclose a picture of an ork, taken from *The Scarecrow of Oz*, a book by L. Frank Baum that was published in 1915 [*see illustration below*].

MARTIN GARDNER

Hastings-on-Hudson, N.Y.

Sirs:

Stillman Drake has conjectured that Galileo was able to estimate the relative time of events accurately because, as a musician, he could keep a constant tempo ["The Role of Music in Galileo's Experiments," by Stillman Drake; SCIENTIFIC AMERICAN, June]. And Samuel Katzoff tells of Benjamin Franklin's measuring time by counting as fast as possible to 10 and indexing the 10's on his fingers ["Letters," SCIENTIFIC AMERICAN, August].

I suggest that such methods must be used quite naturally by many and need not be thought of as the special solution of geniuses. When I was an undergraduate at Harvard, I took a course in experimental psychology for which we had

An ork of Oz

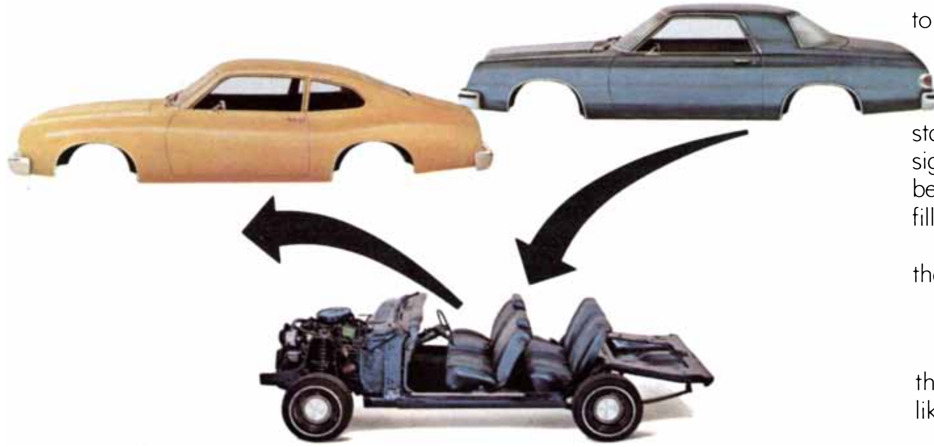
virtually no apparatus beyond string, wire and the contents of our pockets. We got excellent estimates of real-time delays in seconds by counting "one steamboat, two steamboats..." up to 10 and indexing groups of 10 by finger counts (digital storage). In my first year as an instructor I was faced with 70 students in an experimental psychology class and no apparatus. We were able to do an astonishing number of controlled experiments with such devices as playing cards, sandpaper, glass marbles and rulers. As an example we estimated very accurate motor reaction times by means of a ruler and the relation $s = (\frac{1}{2})gt^2$, where s is distance, g is the gravitational constant (9.80665 millisecond^{-2}) and t is time. The experimenter held a ruler against the wall with a vertical edge coincident with the lower edge of a strip of electrical tape stuck to the wall. The subject stopped the ruler in its flight and the distance fallen was used to compute a simple reaction time. I checked this method again just now on a colleague in the hallway. In three trials I measured falls of 18, 16 and 19.6 centimeters, which give times of 191.6, 180.6 and 199.6 milliseconds, from the relation $t = (2s / g)^{1/2}$, the mean value of which, 190.6 milliseconds, is in good agreement with classical results for unpracticed motor reaction times.

An interesting and original use of subjective judgment in fundamental measurement was that of Marin Mersenne (1588–1648), a fellow student of Descartes's and a pupil of Galileo's. Mersenne noted that it took him exactly one second to shout "*Benidicam dominum.*" He then found a distance from a wall such that a shout was tail to head to the echo. Thus he heard "BENIDICAMDOMINUMbenidicamdominum" when he was 158 meters from the wall, which distance doubled gave him $c = 316$ meters per second^{-1} as the velocity of sound. The modern value is 332 meters per second^{-1}.

Much later, with the aid of an electromagnetic fork, this method of coincidences was used to compute the velocity of sound with a timing accuracy said to be about 1/400 second. So for Kahl in 1864 (see A. B. Wood's *A Textbook of Sound,* Third Edition, 1955, Bell, London) an observer with a tapper moved away from a wall noting the points of coincidence of direct sound and echo (perceptual fusion). The velocity is $c = 2 \cdot N \cdot d$, where N is the number of signals per second and d is the mean distance between coincidences. The precision of these results is reasonable in view of the

NE

What does



The Detroit concept: Make it look new.

It's probably the most powerful word in advertising.

And often the emptiest.

Because it's been abused so much.

Make a modest change in a product and right away it's NEW!

Make two little changes and it's NEW NEW!

Some advertisers have even gone so far as to label their products ALL NEW! Which, if you stop for a moment and analyze it, is somewhat redundant.

Detroit vs. Webster

Let's take the American car manufacturers. And mind you, we don't mean to disparage their craftsmanship, because the fact is, they do make some pretty fine automobiles.

But their misuse of the word "new" would make Webster turn over in his grave.

Last year, almost every major American car manufacturer introduced new outsides on old insides. And spent gigantic advertising budgets promoting their NEW cars.

(One manufacturer is actually about to launch a major NEW car introduction for an automobile that's been around in Europe for some time now, that they're merely making a few changes and slapping a sleek American body on.)

Are these cars really new? Hardly.

Sure, legally they can get away with calling them NEW. However, if the lawyers really wanted to be accurate about it, their advertising should carry a sentencethat reads "OUTSIDE NEW ONLY."

The Secret Everybody Knows

Now, we all know what "new" really is. It's no great, dark secret that you have to go to the top of the mountain to find out.

It's simply that which hasn't existed before.

In the case of an automobile, it's starting from scratch and totally redesigning just about every single part to best fill your needs. Or rather, to best fill the needs of the driver.

Which is exactly what we did with the Volkswagen Rabbit.

Five Long, Hard Years

Five years ago, we set out to design the car of the future. Which may sound like a cliché, but it happens to be true.

We wanted to build the perfect car not only for today, but for the next twenty (maybe more) years.

To do that properly, we had to start from ground zero, taking everything into consideration—primarily economy, handling, safety and comfort.

Let's take economy.

With the price of gas skyrocketing—and no relief in sight—we felt we had to build a car that didn't get good, but *great* gas mileage.

And so we did. The standard transmission Rabbit gets an EPA-estimated 39 miles per gallon on the highway—and 25 in the city. (Actual mileage may vary depending on type of driving, driving habits, car's condition and optional equipment.)

Big Mileage: No Big Deal

Now there's nothing that extraordinary about getting high gas mileage—if you want to sacrifice performance (which is exactly what most cars do). But we didn't want to. We felt we couldn't. More and more superhigh-

© Volkswagen of America.

W.

it mean?

The VW concept: Make it new.

ways are being built every day and our car had to be zippy enough to negotiate them

Well, our engineers figured out a way, despite the 39 miles per gallon, to get the Rabbit from 0 to 50 in 8.2 seconds.

To our knowledge, there is no other car in the world—none—that can give you this much gas mileage and this much acceleration together. And there may never be another one.

A Good Handling Car Is a Safe Handling Car

As far as handling goes, we didn't just stop at things like front-wheel drive for better tracking and rack-and-pinion steering (though they make the car handle so well we probably could have). We designed, for example, a totally unique "independent stabilizer rear axle." Rather than bore you with the details right now, we think it will suffice to say that this axle significantly increases the stability of the car on rough roads. And therefore the safety.

And speaking of safety, we gave the Rabbit features that you'll find on few other cars in the world. Like something called "negative steering roll radius," which helps bring the car to a straight stop in the event of a front-wheel blowout. "Dual diagonal brakes," which means that if either brake circuit fails, directional stability is maintained. And a uniquely designed double-jointed steering column that breaks aside in the event of impact.

How We Did the Impossible

Our engine, by the way, is what's called a "transverse engine." Which means it's mounted sideways. That's how we were able to keep the Rabbit so compact on the outside, yet so big and comfortable on the inside (it actually has the same amount of head and leg room as some mid-sized American cars!).

Curl Up With a Good Ad

Most of the incredible features that we've incorporated into this revolutionary automobile we really don't have the space to go into right now. However, you'll get a chance to read about them in detail in future ads we're planning to run. We're certain you'll be quite impressed.

But what will impress you even more is stopping in at a VW dealer and actually seeing the Rabbit in the flesh. And, of course, driving it.

You see, if you're in the market for a new car, we think your hard-earned money deserves more than just the word NEW with an exclamation point after it.

It deserves new, period.

The Amazing Rabbit

We're drilling for plastic.

When somebody mentions crude oil and natural gas, most people think of gasoline, motor oil, and fuel for home and industry.

Diamond Shamrock sees more than these products in a barrel of crude oil or a thousand cubic feet of natural gas. Because crude oil and natural gas are also raw materials in the production of chemical feedstocks. At our petroleum manufacturing complex in the Texas Panhandle, facilities are now under construction to furnish these vital feedstocks for our future growth in plastics, and many other products.

Plastics used to make building and construction materials, home furnishings and household goods, consumer goods such as rainwear, sporting goods and toys, automobile and transportation equipment,

medical goods, packaging for foods, hard and soft goods, and a multitude of other things.

Long before most people were aware of the impending energy shortage, Diamond Shamrock significantly increased its budget for domestic oil and gas exploration and production. Today, we have expanded our net leasehold acreage by 52%, and brought important new sources of crude oil and natural gas into production.

Offshore in the Gulf, Diamond Shamrock owns interests in 54 tracts, covering 244,300 acres of Federal leases offshore Texas and Louisiana. Natural gas is flowing from 34 wells drilled from two platforms and drilling operations are underway from two others. Six other platforms are under construction. In shallower state waters offshore Texas, Diamond Shamrock owns major interests in two production platforms already in operation.

Onshore Diamond Shamrock has assumed a leading position in the Rocky Mountain area, principally in Wyoming.

We've stepped up development of gas fields in the Anadarko Basin of the Texas Panhandle and we are actively evaluating promising areas in Alabama, Arkansas, Florida, Kansas, Louisiana, Montana and Texas. Long-range, we are preparing for participation in such frontiers as offshore Atlantic, the Gulf of Alaska, Bristol Bay, and other domestic areas.

In addition to gasoline sold through branded service stations in 10 Southwestern and Rocky Mountain States, other Diamond Shamrock petroleum products include: diesel fuel, lubricants, jet-turbine fuel, liquefied petroleum gases, fertilizers, asphalt and sulphur.

And when our expanded facilities in the Texas Panhandle are completed in 1976, feedstock from oil and gas operations will be going to our chemical plants in the Houston area.

For more information about Diamond Shamrock, the international corporation that's turning industries around with chemistry and energy, write: Diamond Shamrock, 1100 Superior Avenue, Cleveland, Ohio 44114.

Because when we drill for oil and gas, we're drilling for more than just gasoline, motor oil, or heating fuel.

Diamond Shamrock

ability of the human ear to resolve successive clicks as little as one millisecond apart.

EDWARD CARTERETTE

University of California
Los Angeles

Sirs:

Arthur A. Few's article "Thunder" in your July issue, and particularly his discussion of acoustic techniques for determining the geometry and geography of lightning channels, recalls a study by the late M. G. H. Ligda, about two decades ago, of the use of radar for similar objectives ("The Radar Observation of Lightning," *Journal of Atmospheric and Terrestrial Physics,* Vol. 9, pages 329–346).

In that paper Ligda reports a lightning stroke over 100 statute miles (161 kilometers) long, as far as I am aware a length never yet exceeded and nearly an order of magnitude greater than the longest cited by Few. It would be interesting to know whether any joint acoustic and radar observations exist, and how the two techniques compare.

WILLIAM K. WIDGER, JR.

Laconia, N.H.

Scientific American, November, 1975; Vol. 233, No. 5. 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 30 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 subscription is shown at upper right-hand corner of label.)

NAME

NEW ADDRESS

OLD ADDRESS

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

NOVEMBER, 1925: "Typical of the advances made in biological chemistry is the gradual growth of our knowledge of diabetes, a disease that is manifested by a derangement of the ability of the body to take care of carbohydrates or sugar. Some 50 years ago German investigators found that when a patient dies in the coma that is characteristic of death in this disease, there is overproduction in his system of acetone, or diacetic acid; the body condition is that of acidosis. Later it became apparent that diabetes could be produced by the removal of an organ lying close to the stomach and liver and known as the pancreas. Another investigator showed by studies with the microscope that the pancreas contained a certain form of differentiated tissue in the shape of little islands, which have since been called the islands of Langerhans. Pathologists observed that in cases of diabetes the islands of Langerhans were degenerated. With these studies before them, Doctors Banting and Best conceived the idea of preparing an extract of the islands of Langerhans with which they could treat diabetes. Working with the advice of Doctor J. J. R. Macleod of Toronto and aided in preparing the extract by Doctor J. B. Collip, they at last produced insulin, the culmination of the researches of a half century."

"The latest temporary star, Nova Pictoris, is one of the most conspicuous stars in the southern sky; it is unusual among such objects on account of the exceptional slowness of its changes. A typical 'new star' rises to full brightness in a very few days and fades away rapidly. In the present case the corresponding changes run into months. What little we know of Nova Pictoris fits in well with the belief that a mass of matter is flying violently off from the star's surface in all directions. This extraordinary expansion is evidently caused by some huge explosive liberation of energy at the star's surface, or a little inside it, which drives the outer portions outward at such a rate that after a day or two they get practically beyond the influence of the star's attraction and fly off into space. Since the star remains after the ejected matter has spread out so far that we lose sight of it, and remains with almost the same brightness as before, it seems clear that the outburst, vast as it is, must be only superficial, leaving the main core of the star untouched. What causes the tremendous explosion is still very much a mystery. The amount of energy liberated must be so great that it is tempting to suppose that in some way the energy locked up in the core of the atoms—which is probably slowly liberated in the deep interior of an ordinary star—is here rapidly liberated."

"The dirigible *Shenandoah* was wrecked in a furious thunderstorm that flung her 4,000 feet upward into the heavens as though she were a shaving of wood caught in a spouting geyser of water. When she drove her nose into the uprushing column of air, the sudden upward thrust set up a violent vertical bending stress. She was swept rapidly upward, and Commander Lansdowne valved her freely, pointing her nose down with engines running. She came down with such rapidity that he had to discharge water ballast in large quantities and order the dropping of gas tanks. She was brought to a level keel at about 3,000 feet, but here she suddenly broke in two. An airship is by far the frailest of all mechanical constructions of great size that men's hands have fashioned. The integrity of this delicate framework depends on its being everywhere supported by the upward pressure of the bags of gas within it. Deflate two or three bags and the ship will sag and break her back. Unlike a bridge, the airship cannot run a web of trussing from top to bottom, thus utilizing her depth to gain girder strength. The gas bags are in the way."

"Radio is said to be revolutionizing the methods of Scotland Yard, the famous headquarters of London's police. Experiments there have revealed radio so widely useful in police work that the entire system of crime detection throughout the British Isles is likely to be shifted to a radio basis. The Yard has seven radio-equipped motor cars attached to the criminal investigation flying squad. These radio cars not only aid in detecting crime but also perform a helpful service in regulating heavy traffic along the highways. On Derby Day they were stationed at strategic points on the roads leading to Epsom and by signaling to one another the moment traffic was to be diverted they prevented the arteries of transportation from being choked with automobiles. It is believed that before long detectives engaged in work that requires quick and secret communication with headquarters will carry small portable sets."

SCIENTIFIC AMERICAN

NOVEMBER, 1875: "Dr. Du Chaumont, in a recent lecture, said that up to a quite late date there was an absence of any satisfactory theory as to the relation of food to work. It was supposed that bodily force was due to a chemical change in the muscles themselves, and that the nitrogenous matter in food repaired the waste. The researches of Joule, Playfair, Frankland and others on the conservation of energy have led to the conclusion that active force is produced chiefly by the potential energy stored in the carboniferous food and set free by oxidation."

"It may be safely predicted that the time is not very far distant when vessels carrying perishable cargoes of fruit, meat and other articles of food will make constant and regular voyages between the Tropics and the colder temperate regions. The use of refrigerator cars in transporting the fruit and vegetable productions of California to the Atlantic seaboard, and more recently the export of a quantity of American peaches to England by steamer, may be considered successful experiments leading to a steady commerce. In Texas, on the pampas of South America and in Australia thousands of cattle are slaughtered simply for their hides, the bodies being left totally unutilized. If success in the transportation of this beef can be obtained, an immense trade is at once possible."

"Sir Charles Wheatstone, the distinguished inventor, died in Paris the 19th of October last. He was born at Gloucester, England, in 1802, and in youth was engaged in the manufacture of musical instruments. With the object of improving upon these he was led to study the laws of sound and, imbibing a strong taste for physical science, he proceeded to the investigation of the phenomena of optics and subsequently of electricity. With William Fothergill Cooke of Heidelberg, Germany, Wheatstone perfected a new mechanical and electromagnetic telegraph."

THE AUTHORS

WILLIAM EPSTEIN ("Nuclear-free Zones") has been involved in the work of arms control and disarmament for many years, most prominently in his capacity as director of the Disarmament Division of the United Nations Secretariat, a post he held until his retirement last year. During this period he represented the UN Secretary-General at numerous international conferences, among them the meetings of the Preparatory Commission for the Denuclearization of Latin America, which led to the establishment of the world's first nuclear-free zone by the Treaty of Tlatelolco in 1967. A lawyer by training, he was educated at the University of Alberta and the London School of Economics. His association with the UN goes back to 1945, when he joined the staff of the UN Preparatory Commission in London. He continues to serve as a consultant to the Secretary-General and as a Special Fellow of the United Nations Institute for Training and Research (UNITAR), commuting to New York once a month from British Columbia, where he is currently visiting professor at both the University of Victoria and the University of British Columbia at Vancouver.

CAROLYN COHEN ("The Protein Switch of Muscle Contraction") is professor of biology and chairman of the graduate biophysics program at Brandeis University. A graduate of Bryn Mawr College, she obtained her Ph.D. in biophysics from the Massachusetts Institute of Technology in 1954. She then spent a year as a Fulbright Scholar at King's College of the University of London, returning to M.I.T. in 1955. Three years later she set up the Structural Biology Laboratory at the Children's Cancer Research Foundation in Boston (now the Sidney Farber Cancer Center). In 1972 she and her associates moved their laboratory to the new Rosenstiel Basic Medical Research Center at Brandeis.

HENRY KOLM, JOHN OBERTEUFFER and DAVID KELLAND ("High-Gradient Magnetic Separation") developed the new technique described in their article as an outgrowth of basic research at the Massachusetts Institute of Technology's Francis Bitter National Magnet Laboratory. Kolm has been connected with M.I.T. since 1946, when he entered as a freshman following military service. He obtained his B.S. and Ph.D. in physics in 1950 and 1954 respectively and went on to do solid-state research for a number of years at M.I.T.'s Lincoln Laboratory before becoming one of the founders of the National Magnet Laboratory in 1963. His research since then has focused on the generation of very strong magnetic fields and on their practical applications. Oberteuffer earned his bachelor's and master's degrees at Williams College and his doctorate at Northwestern University, all in physics. He came to M.I.T. originally as a postdoctoral research fellow and switched to the National Magnet Laboratory in 1971 to work on the magnetic-separation project. In 1974 he joined Sala Magnetics, Inc., in order to pursue his work to the point of industrial practice. Kelland is a graduate of Montclair State College, acquiring a B.A. there in 1957 and an M.A. in 1960. He taught physics at Simmons College and Emmanuel College before joining the National Magnet Laboratory in 1967.

JOHN CAIRNS ("The Cancer Problem") works at the Mill Hill Laboratories of the Imperial Cancer Research Fund in London. After obtaining his medical degree from the University of Oxford, Cairns did research in Australia for several years on the multiplication of viruses and on the visualization of DNA molecules by autoradiography. He later came to the U.S. and served as director of the Cold Spring Harbor Laboratory of Quantitative Biology until 1968, when he was succeeded by James D. Watson. For a while, he reports, he continued his work on DNA synthesis, but in 1973 he returned to England, deciding "that I should myself stop doing research and instead try to produce a scientists' guide to cancer research; there is no such book at the moment and it is my hope that, having had a career in medicine followed by virology followed by molecular biology, I at least have the background for producing such a book."

TORKEL WEIS-FOGH ("Unusual Mechanisms for the Generation of Lift in Flying Animals") is professor of zoology at the University of Cambridge, where he is a fellow of Christ's College. A native of Denmark, he studied at the University of Copenhagen, acquiring his doctorate there in 1952. Before being appointed to his present post in 1966, he was professor of zoophysiology and head of Zoophysiological Laboratory B at Copenhagen. Although now an "Anglophile resident" of England, he writes, "I retain a strong connection with my old university in Copenhagen and with the Danish

Pioneer's new High Polymer Molecular transducer technology will alter the course of high fidelity.

There's a significant new development in high fidelity that is destined to play a vital role in sound reproduction. It is intimately tied in with the piezoelectric principle.

The piezoelectric effect deals with certain crystal devices that flex when voltages are applied to them. Now, Pioneer has discovered a totally new application of the piezoelectric effect by applying the principle to ultra-thin aluminum coated high polymer film.

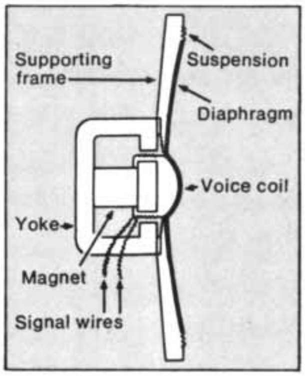


Fig. 1 Principle of conventional dynamic headphones

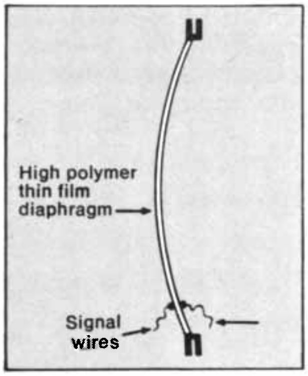


Fig. 2 Principle of new Pioneer High Polymer Molecular headphones

By employing this film as a low-mass diaphragm and applying audio signal voltages, the material expands and contracts uniformly generating acoustic energy. For the first time it becomes possible to transform electrical energy to an accurate acoustical equivalent. Such thin-film diaphragms properly mounted are capable of reproducing all music frequencies by means of an incredible "breathing" effect. The ramifications of this unique refinement of the piezoelectric principle are far reaching. Consider such immediate applications as microphones, cartridges, speaker systems and headphones — in fact, any type of electromechanical transducer requiring resonance-free performance.

There have been many attempts to create sound using diaphragm motion. For example, electrostatic speakers and headphones. But in contrast to the electrostatic principle, the new application of the High Polymer Molecular principle as discovered and perfected by Pioneer, requires no dangerous, high polarizing voltages.

The first totally new concept in headphones in over a decade.

Pioneer has successfully incorporated the High Polymer Molecular transducer principle in two new headphones that are unlike any others. Conventionally designed headphones use moving coils, miniature loudspeaker elements and other mechanical parts — as shown in Figure 1 — all of which come between you and your music. Pioneer's new SE-700 and SE-500 headphones don't. They employ a single thin-film high polymer piezoelectric diaphragm that reproduces sound directly, as shown in Figure 2. Only the diaphragm moves air — and moves it accurately, in exact conformance with the electrical signal applied directly to it. The accurate, low-distortion signals available from any standard headphone jack on your receiver or amplifier are directly translated to equally precise, low-distortion sound by the action of the high polymer film diaphragm. Nothing, absolutely nothing comes between you and the original sound.

Even though you may now own a pair of headphones, you owe it to yourself to hear these new piezoelectric high polymer transducer headphones. In fact, compare them with other types. You'll find a lower level of distortion-free sound than has ever been achieved — even at unprecedented volume levels. The experience of listening with these new Pioneer headphones is a revelation. In addition, the open-back design, light weight and soft, snug fitting earpieces permit hours of comfortable, private listening. You'll come away from your Pioneer dealer thoroughly convinced that Pioneer has altered the course of high fidelity.

SE-700, under $80*; SE-500, under $50*. Both come with a 9¾ foot connecting cable, standard phone plug and storage case.

U.S. Pioneer Electronics Corp., 75 Oxford Drive, Moonachie, New Jersey 07074. West: 13300 S. Estrella, Los Angeles 90248 / Midwest: 1500 Greenleaf, Elk Grove Village, Ill. 60007 / Canada: S.H. Parker Co.

*The value shown is for informational purposes only; the actual resale price will be set by the individual Pioneer dealer at his option.

PIONEER
when you want something better

community, and I do most of my writing, creative thinking and some experimental work in my cottage north of Copenhagen, placed ideally in the midst of fields and forests with an abundance of hover flies, dragonflies and birds on site."

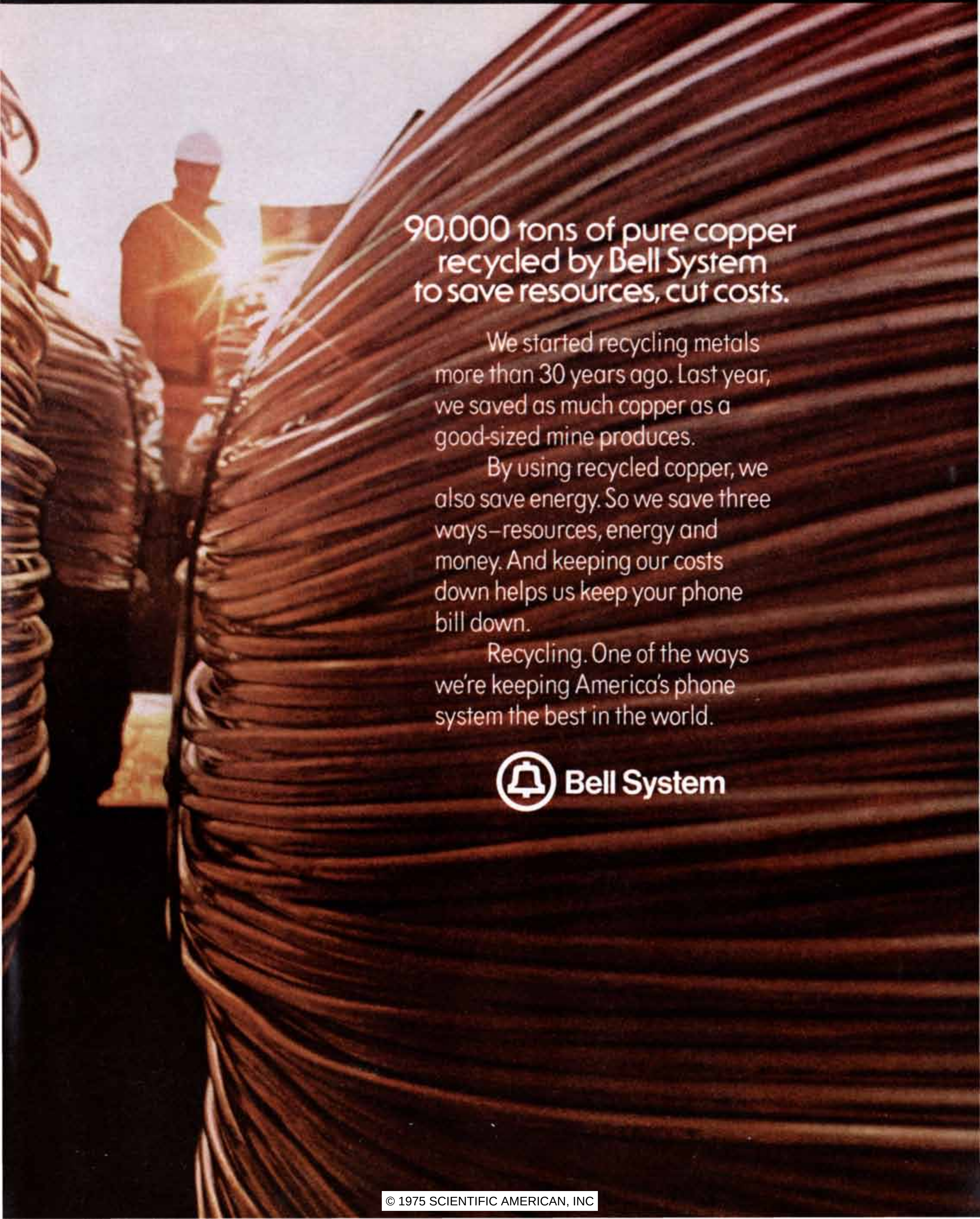
M. NAFI TOKSÖZ ("The Subduction of the Lithosphere") is professor of geophysics at the Massachusetts Institute of Technology, where he heads the George R. Wallace, Jr., Geophysical Observatory. Born in Turkey, Toksöz completed his undergraduate studies at the Colorado School of Mines and went on to obtain a Ph.D. in geophysics and electrical engineering from the California Institute of Technology in 1963. He did research at Cal Tech for several years before moving to M.I.T. in 1965. A specialist in seismology, plate tectonics and the structure and evolution of planetary interiors, he has devoted his recent efforts to "extensive theoretical calculations" in an effort to show that "processes involved in the earth's lithosphere and asthenosphere can indeed produce geological and geophysical impressions at the surface."

B. V. DERJAGUIN and D. B. FEDOSEEV ("The Synthesis of Diamond at Low Pressure") are on the research staff of the Institute of Physical Chemistry of the Academy of Sciences of the U.S.S.R. Derjaguin has served as director of the Institute's Laboratory of Surface Phenomena since its founding in 1935; he has been a corresponding member of the Academy since 1946. Fedoseev has collaborated with him on the experiments described in their article.

ECKHARD H. HESS ("The Role of Pupil Size in Communication") is professor of psychology at the University of Chicago. He writes: "I am primarily an ethologist, having studied such diverse animals as the common fruit fly, fishes, chicks, ducks, Canada geese, turkeys, Japanese quail, rats, cats, guinea pigs, whales and man. I spend several months each spring at a research station in the saltwater marshes of the eastern shore of Maryland, where I study imprinting." Hess has written five previous articles for SCIENTIFIC AMERICAN, the most recent of which was " 'Imprinting' in a Natural Laboratory," which appeared in the August 1972 issue. His interest in human perception from an ethological viewpoint has resulted in a book, *The Tell-Tale Eye*, published by Van Nostrand Reinhold Company; the work described in the present article is an extension of the studies described in that book.


"My favorite airplane!"
N54641
DC-9
MCDONNELL DOUGLAS
"My favorite airplane!"
DC-10
MCDONNELL DOUGLAS DC-10
Many airlines fly the DC-9 and their crew members say there's no other jetliner like it. Many other airlines fly the DC-10 and those flight crews know it's the greatest. Maybe you have your favorite, too: The DC-9 because it holds the highest on-time departure rate (over 99%) of any twinjet in the world—which means you can depend on it to leave on schedule. Or the DC-10 wide-cabin because of its quiet comfort, roomy elegance, and similarly remarkable on-time departure rate (over 97%).
At any rate, passengers like to fly the DC jetliners that so many airlines fly. The quality and performance we've built into more than 3,000 commercial airliners have made them favorites for more than 40 years. With airline presidents or pilots. With experienced air travelers or first-time fliers. So, ask your airline or travel agent to put you aboard a DC-9 or DC-10 the next time you fly.
MCDONNELL DOUGLAS
DC jetliners:
Choice of 64 world airlines.

PHONE SYSTEM THE BEST IN THE WORLD.



Established 1845 SCIENTIFIC AMERICAN November 1975 Volume 233 Number 5

Nuclear-free Zones

The failure of recent attempts to revitalize the Nonproliferation Treaty has led to renewed interest in regional multinational pacts as an alternative way of preventing the spread of nuclear weapons

by William Epstein

The effort to prevent the proliferation of nuclear weapons continues to face numerous obstacles. "Vertical" proliferation—the ongoing development and deployment of more advanced weapons systems by the existing nuclear powers—has not been halted or even slowed by the series of strategic-arms-limitation agreements between the U.S. and the U.S.S.R. Meanwhile "horizontal" proliferation—the spread of nuclear-weapons capability to additional countries—has proceeded in spite of such supposed barriers as the Partial Test-Ban Treaty and the Treaty on the Nonproliferation of Nuclear Weapons. Indeed, the very survival of the Nonproliferation Treaty was jeopardized last year by India's test explosion of a "peaceful" nuclear device, which must be viewed in the context of an energy-hungry world where nuclear power reactors are being installed at an accelerating rate and where access to sophisticated nuclear technology and weapons-grade fissionable material is being acquired by dozens of countries and tens of thousands of people currently outside the "nuclear club" [see "The Proliferation of Nuclear Weapons," by William Epstein; SCIENTIFIC AMERICAN, April].

The review conference of the parties to the Nonproliferation Treaty, held in Geneva in May, turned out to be a particular disappointment. Anticipated by many participants and observers as a rare opportunity to strengthen and extend the basic provisions of the treaty, the conference quickly degenerated into an unresolved confrontation between the nuclear parties and the non-nuclear parties. The net result of the conference appears to have been a widening rather than a narrowing of the gap between the two groups.

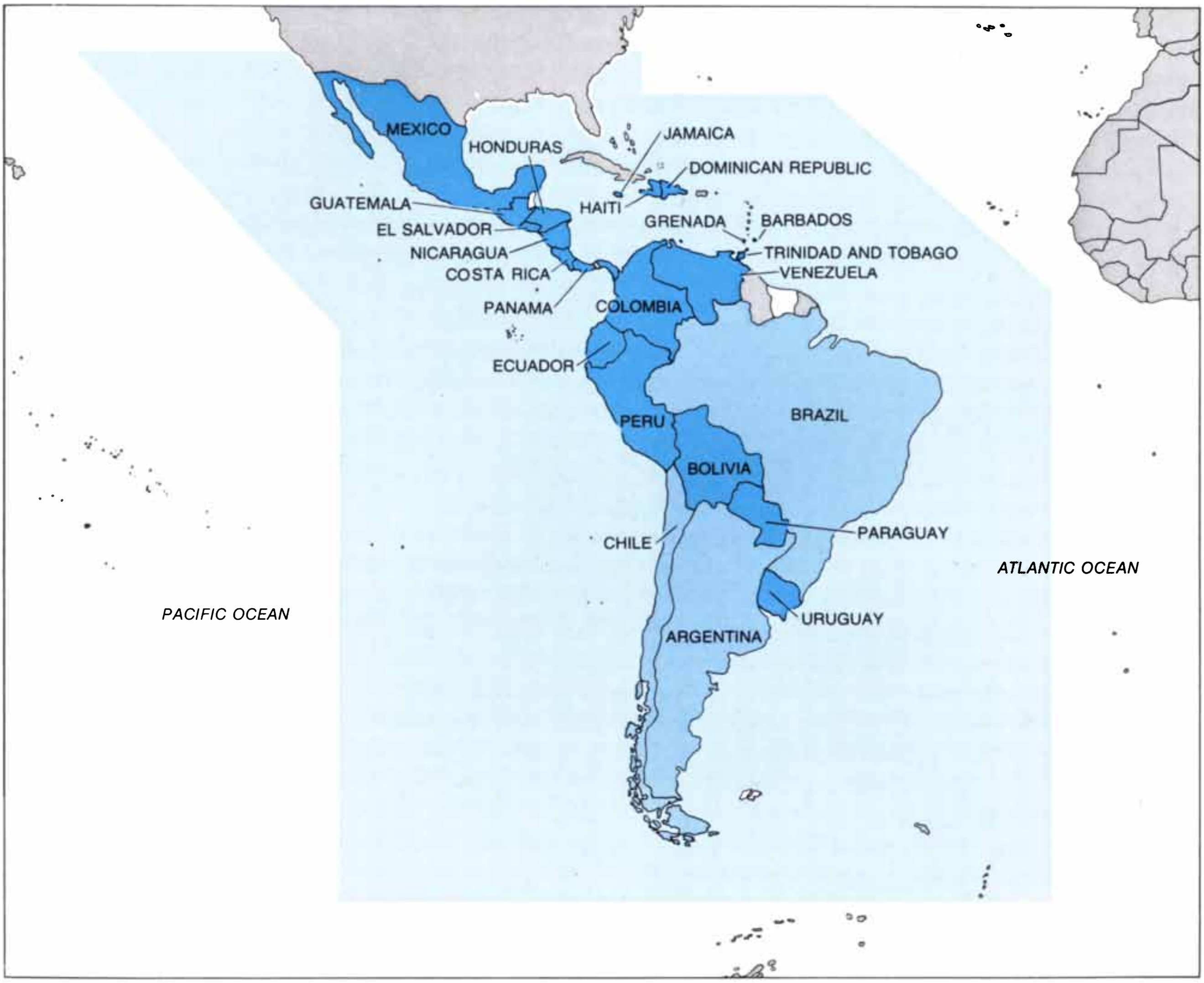
Is the effort to prevent the further proliferation of nuclear weapons therefore hopeless, and must we all learn to live in a world of many nuclear powers? I believe the answer is no, or at least "Not yet." There is still a chance to slow down and limit nuclear proliferation. Even if the likelihood is that the Nonproliferation Treaty will slowly erode and that several new nuclear powers will emerge in the next decade or so, the treaty can still remain an important restraining influence on the spread of nuclear weapons. In addition to the Nonproliferation Treaty, however, there are several other ways to restrict the number of new nuclear powers. Among them is the creation of nuclear-free zones. Provided that such zones are correctly conceived, adequately prepared and properly implemented, they could be even more effective than the Nonproliferation Treaty in curbing the spread of nuclear weapons.

Every international conference on nuclear energy since the end of World War II has been concerned with two basic goals: first, controlling and eliminating "atoms for war," and second, promoting and exploiting "atoms for peace." The basic dilemma has been that the development of nuclear energy for either purpose also helped to enhance its potential for the other. From the beginning all efforts and negotiations toward the first goal, nuclear-arms control and disarmament, had as their aim not only the control of nuclear weapons but also the prevention of their spread to other countries. The immediate objective was to prevent the further proliferation of nuclear weapons both vertically and horizontally, and the ultimate objective was to eliminate them entirely.

With the failure of the Baruch Plan in the late 1940's, little hope remained for achieving the elimination of nuclear weapons, but efforts continued to limit their development, to reduce their numbers and above all to prevent their horizontal proliferation. Subsequently two different approaches were developed to the problem of preventing the spread of nuclear weapons: first, the creation of nuclear-free zones in which all nuclear weapons would be prohibited, and second, the enactment of a treaty that would specifically ban the dissemination of nuclear weapons by the nuclear powers and the acquisition of such weapons by states not possessing them. These two approaches were supplementary to the ban on nuclear-weapons tests, which was considered as being an important aim in itself but which it was recognized would also help to prevent the proliferation of nuclear weapons.

Since 1958, when Poland first proposed the denuclearization of Central Europe (Poland, Czechoslovakia, East Germany and West Germany), various proposals have been put forward for the denuclearization of other geographic areas, including the Balkans (by Romania), the Mediterranean (by the U.S.S.R.), the Middle East (by Iran), the Nordic countries (by Finland) and Asia and the Pacific region (by China). These proposals consisted mainly of general concepts rather than concrete steps. Formal and specific plans were put forward only for Central Europe, Africa and Latin America. All these proposals, except the one concerning Latin America, failed to make significant progress because of the complex political and strategic questions involved. More recently, after the Indian nuclear test, the proposal for a nuclear-free zone in the Middle East was revived by Iran and Egypt, and a new proposal for a nuclear-free zone in South Asia was put forward by Pakistan.

All proposals for nuclear-free zones were supported in principle by the U.S.S.R. and its allies, but they placed particular emphasis on Central Europe, the Mediterranean and Asia, where the two great-power blocs confronted each other and where the danger of nuclear conflict seemed greatest. Proposals for nuclear-free zones in those three regions were aimed mainly at reducing the American nuclear presence there. The U.S. and its allies considered that such reduction of American military power would give some military or political advantage to the U.S.S.R. They conceived of such zones chiefly in the context of preventing the spread of nuclear weapons, and they laid down certain principles regarding their creation: (1) that they should not upset the existing military balance, (2) that they should be initiated by the states in the region, (3) that they should include all the countries of



LATIN-AMERICAN NUCLEAR-FREE ZONE, created in 1967 by the Treaty of Tlatelolco, is at present the only such zone established in a populated region of the world. The 20 countries for which the treaty is in full force are shown in dark color. Two countries, Brazil and Chile, have signed and ratified the treaty but have not waived certain requirements that must be met before a party is bound by the provisions of the treaty. The treaty is also not yet in force for Argentina, which has signed but not ratified it. Territories of foreign countries that have been included in the zone under Protocol I of the treaty are white. The total extent of the nuclear-free zone takes in a large part of the Atlantic and Pacific oceans (*light-color area*). Area is substantially the same as that of "security zone" declared by Latin-American countries on the initiative of the U.S. and accepted by Allied Powers in World War II.

the area if possible, or at least those with significant military power, and (4) that they should be subject to verification to ensure that the zone would remain nuclear-free.

So far the only nuclear-free zone created specifically for the purpose of preventing the spread of nuclear weapons is the one established by the Treaty for the Prohibition of Nuclear Weapons in Latin America (commonly referred to as the Treaty of Tlatelolco, after the borough of Mexico City where it was signed). Other nuclear-free zones—those established by the Antarctic Treaty of 1959, the Outer Space Treaty of 1967 and the Seabed Treaty of 1971—were created in part for the purpose of arms limitation, but they were mainly concerned with the efforts of the world community to regulate the use of these still unexplored, unexploited and uninhabited environments. Unlike the Treaty of Tlatelolco, their arms-limitation aspects were secondary and therefore comparatively easy to achieve.

The Cuban missile crisis of 1962 had suddenly and dramatically confronted the states of Latin America with the fact that their area had become directly involved in the strategic plans and rivalries of the nuclear powers. There were immediate consultations aimed at preventing the recurrence of a similar crisis in some other country of Latin America and also precluding the possibility that a nuclear-arms race might develop among the countries of the area. On April 29, 1963, the heads of state of five Latin-American nations (Bolivia, Brazil, Chile, Ecuador and Mexico) issued a joint declaration favoring a multilateral agreement whereby the countries of the region would undertake not "to manufacture, store or test nuclear weapons or devices for launching nuclear weapons." Later that year the United Nations General Assembly adopted a resolution giving its support and encouragement to the idea. On the initiative of Mexico a Preparatory Commission for the Denuclearization of Latin America was established in November, 1964, in Mexico City.

There ensued intensive negotiations over the next two years on the elaboration of a treaty to create a nuclear-free zone in Latin America. Although the four nuclear powers that were then members of the UN gave their blessing to these efforts, their representatives expressed much skepticism in private. I had been appointed technical adviser to the Preparatory Commission by the UN Secretary-General and took part in the work of preparing the treaty. On several occasions various officials of the four nuclear powers indicated to me that the concept of a Latin-American nuclear-free zone was premature and unlikely to succeed; that it was too difficult to get agreement among all the Latin-American states, some of which were not really serious about the project; that the states of the area had too many problems and differences of their own, and that it was unlikely that all the nuclear powers would give their support to it.

There were indeed a number of hard problems to be solved. Would it be necessary for all the states of Latin America to become parties to the treaty before its entry into force or could a treaty limited to only part of the zone be effective, even if all the areas were not contiguous? Would it be necessary for all foreign countries having responsibility for territories in the zone (France, the Netherlands, Great Britain and the U.S.) to agree that their territories should be included? Would it be necessary for all five nuclear powers to undertake to respect the zone?

These problems were solved in the following way. It was first decided that the treaty would enter into force when it was ratified by all the Latin-American countries; when all the countries mentioned above having responsibility for territories in the zone ratified a protocol to the treaty undertaking to apply the treaty in those territories (Protocol I); when all five nuclear powers ratified a protocol undertaking to respect the nuclear-free status of the zone and not to use or threaten to use nuclear weapons against the parties to the treaty (Protocol II), and when each of the parties to the treaty had concluded agreements with the International Atomic Energy Agency for the application of the IAEA's safeguards system to its nuclear activities. These requirements were far-reaching and it was obvious that a long time would elapse before all of them were met. Accordingly Alfonso García Robles, who was then Under-Secretary for Foreign Affairs of Mexico and chairman of the Preparatory Commission, came up with the ingenious idea that signatory states could waive those requirements in whole or in part so that the treaty would enter into force for such states immediately on the deposit of their instrument of ratification with their declaration of waiver. This proposal was subsequently adopted by the commission.

Concerning the control provisions of the treaty, a comprehensive system of verification was provided. It included the application of the IAEA safeguards system, periodic reports of the parties to the agency established to implement the treaty (OPANAL), special reports when requested by the General Secretary of OPANAL and special inspections in addition to the IAEA's safeguards system in the case of suspicion of a violation.

It was also agreed that nuclear materials and facilities would be used exclusively for peaceful purposes and that nuclear explosions for peaceful purposes could be carried out, including explosions involving devices similar to those used in nuclear weapons, as long as such devices did not constitute nuclear weapons. The Treaty of Tlatelolco, unlike the Partial Test-Ban Treaty and the Nonproliferation Treaty, provided a definition of nuclear weapons. Article 5 of the Treaty of Tlatelolco stated that "a nuclear weapon is any device which is capable of releasing nuclear energy in an uncontrolled manner and which has a group of characteristics that are appropriate for use for warlike purposes." The main obligations of the parties are defined in Article 1 of the treaty. Article 18 sets out the provision regarding explosions for peaceful purposes [*see illustration on page* 29].

Finally, on February 14, 1967, the treaty was signed at Tlatelolco. Twenty-three states of Latin America and the Caribbean have now signed the treaty and 22 of them have ratified it. Of these 22 states the treaty is in force for 20, each of which has deposited a declaration of waiver of the requirements for entry into force. Although Brazil and Chile have ratified the treaty, they have not deposited the declaration of waiver. The treaty is therefore not in force for them. It is also not in force for Argentina, which has signed the treaty but has not ratified it. The Netherlands and Great Britain have signed and ratified Protocol I to the treaty, so that it applies to their territories in the zone. Neither France nor the U.S. has yet signed or ratified Protocol I, even though the General Assembly specifically called on them in 1974 to do so. Protocol II has been signed and ratified by China, France, Great Britain and the U.S., so that they are bound to respect the zone and not to use or threaten to use nuclear weapons against it.

The U.S.S.R. has not yet signed Protocol II, in spite of a number of resolutions by the General Assembly calling on it to do so. Among the reasons it gives is that the treaty allows peaceful nuclear explosions, that it does not apply to the

Panama Canal Zone or to the transit of nuclear weapons and that the area of the zone takes in a large part of the Atlantic and the Pacific [*see illustration on opposite page*]. The U.S.S.R. has said, however, that it will respect the nuclear-free status of each state in the area that remains nuclear-free. Many observers feel that the U.S.S.R. may adhere to Protocol II when the boycott against Cuba is lifted by the Organization of American States and Cuba itself becomes a party to the treaty.

Concerning peaceful nuclear explosions, Ambassador García Robles, who was mainly responsible for the final drafting of the texts of Article 5 and Article 18, has explained that by virtue of the wording of Article 1 and the definition of a nuclear weapon in Article 5 no peaceful nuclear explosion may be carried out if the device has "a group of characteristics that are appropriate for use for warlike purposes," since that would make it a nuclear weapon. This interpretation was expressly supported by the U.S. when it signed Protocol II. On the other hand, when Brazil ratified the treaty, it stated its understanding that the treaty allowed the explosion of nuclear devices for peaceful purposes. A similar declaration was later made by Argentina when it signed the treaty.

After the signing of the treaty in 1967 it was welcomed by the UN General Assembly "with special satisfaction" as "an event of historic significance in the efforts to prevent the proliferation of nuclear weapons." The treaty was also hailed at the Conference of Non-Nuclear-Weapons States in September, 1968, which recommended that other non-nuclear-weapons states study the possibility and desirability of establishing such zones in their areas. The treaty has been frequently described as a model for proposals for the establishment of nuclear-free zones in other areas of the world.

The fact that Argentina, Brazil and Chile, none of which is a party to the Nonproliferation Treaty and all of which are near-nuclear or potential nuclear powers, are not bound by the provisions of the Treaty of Tlatelolco certainly reduces the effectiveness of the regional treaty from the point of view of nonproliferation. It is, however, a matter of importance that not only Mexico, which is a potential nuclear power, but also such countries as Colombia and Venezuela, which have not ratified the Nonproliferation Treaty but which will soon become potential nuclear powers, should be parties to the Treaty of Tlatelolco and be bound by its provisions. Moreover, Chile is a special case, and there are possibilities that it might become a full party by depositing its declaration of waiver without waiting for Argentina and Brazil to do so.

The Treaty of Tlatelolco establishes a nuclear-free zone that encompasses more than 7.5 million square miles inhabited by some 200 million people. What is more important, the number of its parties and supporters, including the nuclear signatories of the protocols, keeps growing year by year. Although there seems little hope for the foreseeable future that either Argentina or Brazil will become full parties, there would appear to be greater pressures and better prospects for their joining such a regional pact, which ensures equal treatment for all parties on a nondiscriminatory basis, than there are for their becoming parties to the Nonproliferation Treaty. They might find some incentive in the negative security assurance that the nuclear powers would be bound not to use or threaten to use nuclear weapons against them, which assurance is not provided to parties to the Nonproliferation Treaty. Moreover, the possibility is not excluded that some compromise formula might be found, or that the advance of technology might facilitate an acceptable solution to the problem of peaceful nuclear explosions. If the nuclear powers would live up to their obligation under the Nonproliferation Treaty to establish a special international regime for the conduct of peaceful nuclear explosions at minimum cost for non-nuclear powers, with special concessionary rates for developing countries, there would be less incentive and less need for either Argentina or Brazil to develop their own capability in this respect.

Since Argentina and Brazil must each be wary of the other's "going nuclear," it is conceivable that they might both be willing to remain non-nuclear or to establish a balance of mutual rivalry and deterrence on the basis of their being peaceful-nuclear-explosive powers with a ready potential of going nuclear rather than nuclear-weapons powers engaged in a mutually ruinous nuclear-arms race.

If there is any chance at all of Argentina and Brazil refraining from going nuclear, it is more likely to be found within the regional context of the Latin-American nuclear-free zone, where they would have a higher consciousness of and receptivity to the feelings, desires and influence of their neighbors, all of whom are developing countries with similar problems, than to the pressures of the rich industrial powers. At any rate the Treaty of Tlatelolco holds out more hope in this regard than the Nonproliferation Treaty, which both countries regard as a discriminatory treaty that the nuclear powers are trying to impose on them.

In spite of the growing interest in nuclear-free zones for several other areas of the world, none of the various proposals has yet got very far. No preparatory committees have been established, no draft treaties are being negotiated and no detailed or formal plans are being discussed by any regional groups. Nevertheless, there are fashions in political affairs, and political, military and strategic developments are moving at such a pace that opportunities may exist tomorrow that appear to be impossible today. A review of the current status of the various proposals for nuclear-free zones that are still on the international agenda may be useful.

The idea of a nuclear-free zone in Central Europe was first suggested in 1957 by Adam Rapacki, who was then foreign minister of Poland. His proposal, which later came to be known as the Rapacki Plan, provided for prohibiting the production and stockpiling of nuclear weapons on the territory of Poland, Czechoslovakia, East Germany and West Germany. Czechoslovakia and East Germany endorsed the plan, and it was formally put forward by Poland in 1958. According to the plan, no nuclear weapons would be manufactured, stockpiled or installed in the territory of the four countries; the use of nuclear weapons against the countries in the zone would be banned; the U.S.S.R., the U.S., Great Britain and France would undertake to respect the nuclear-free status of the zone; a system of ground and air control with inspection bases would be set up to verify the observance of the commitments, and the zone could be established by binding unilateral declarations in order to avoid the political complications of negotiating a formal treaty.

The Rapacki Plan was unacceptable to the Western powers, which regarded it as an attempt to weaken their military position in Central Europe, since it contained no limitations on conventional forces. In addition it would have called for some form of recognition of East Germany, which the Western powers were not willing to accept at that time because they were pressing for a reunified Germany.

In order to meet some of the objections, Rapacki submitted a revised plan that would be implemented in two stages: first, a freeze of nuclear armaments in the proposed zone, and second, a reduction of conventional forces to be carried out together with the complete denuclearization of the zone. A third version of the plan was submitted by Poland in 1962; it envisioned that the proposed zone would be open to any other European state wishing to accede to it. This version was also to be implemented in two stages: the first stage would freeze nuclear weapons and missiles and prohibit the establishment of new military bases in the area of the zone; the second would provide for the elimination of these nuclear arms and for the reduction of armed forces and conventional armaments.

The Western powers were still opposed to any of these versions of the Rapacki Plan on the ground that they were intended to reduce Western nuclear strength in Europe and would thus give the U.S.S.R. a military advantage because of its superiority in conventional forces. Then in 1964 Poland put forward another proposal, which came to be known as the Gomulka Plan, that called for a freeze at the existing levels in the zone of all nuclear weapons "irrespective of the means of their employment and delivery" under a system of control posts at nuclear plants and at rail, road, sea and air points of access. Although the Gomulka Plan did not provide for the reduction or elimination of nuclear weapons already in the area and merely banned any increase in their number, the Western powers found it unacceptable for the same reasons they had rejected the Rapacki plans.

None of these plans for freezing or reducing nuclear and conventional arms and forces in Central Europe came to anything. They were eventually succeeded by the Vienna Conference on mutual force reductions and associated measures in Central Europe. At present it is not clear whether these talks will be broadened to include tactical nuclear weapons. In any case it seems hardly likely that they would go beyond negotiations for the reduction of such weap-

Article 1

1. The Contracting Parties hereby undertake to use exclusively for peaceful purposes the nuclear material and facilities which are under their jurisdiction, and to prohibit and prevent in their respective territories:

(a) The testing, use, manufacture, production or acquisition by any means whatsoever of any nuclear weapons, by the Parties themselves, directly or indirectly, on behalf of anyone else or in any other way, and

(b) The receipt, storage, installation, deployment and any form of possession of any nuclear weapons, directly or indirectly, by the Parties themselves, by anyone on their behalf or in any other way.

2. The Contracting Parties also undertake to refrain from engaging in, encouraging or authorizing, directly or indirectly, or in any way participating in the testing, use, manufacture, production, possession or control of any nuclear weapon.

Article 5

For the purposes of this Treaty, a nuclear weapon is any device which is capable of releasing nuclear energy in an uncontrolled manner and which has a group of characteristics that are appropriate for use for warlike purposes. An instrument that may be used for the transport or propulsion of the device is not included in this definition if it is separable from the device and not an indivisible part thereof.

Article 18

1. The Contracting Parties may carry out explosions of nuclear devices for peaceful purposes—including explosions which involve devices similar to those used in nuclear weapons—or collaborate with third parties for the same purpose, provided that they do so in accordance with the provisions of this article and the other articles of the Treaty, particularly articles 1 and 5.

2. Contracting Parties intending to carry out, or to cooperate in carrying out, such an explosion shall notify the Agency and the International Atomic Energy Agency, as far in advance as the circumstances require, of the date of the explosion and shall at the same time provide the following information:

(a) The nature of the nuclear device and the source from which it was obtained,

(b) The place and purpose of the planned explosion,

(c) The procedures which will be followed in order to comply with paragraph 3 of this article,

(d) The expected force of the device, and

(e) The fullest possible information on any possible radioactive fall-out that may result from the explosion or explosions, and measures which will be taken to avoid danger to the population, flora, fauna and territories of any other Party or Parties.

3. The General Secretary and the technical personnel designated by the Council and the International Atomic Energy Agency may observe all the preparations, including the explosion of the device, and shall have unrestricted access to any area in the vicinity of the site of the explosion in order to ascertain whether the device and. the procedures followed during the explosion are in conformity with the information supplied under paragraph 2 of this article and the other provisions of this treaty.

4. The Contracting Parties may accept the collaboration of third parties for the purpose set forth in paragraph 1 of the present article, in accordance with paragraphs 2 and 3 thereof.

KEY PROVISIONS of the Treaty of Tlatelolco are contained in the three articles reproduced at right. Article 1 defines the main obligations of the treaty; Article 5 provides a definition of nuclear weapons; Article 18 deals with peaceful nuclear explosions.

ons. Any prospect of the elimination of the weapons is therefore remote.

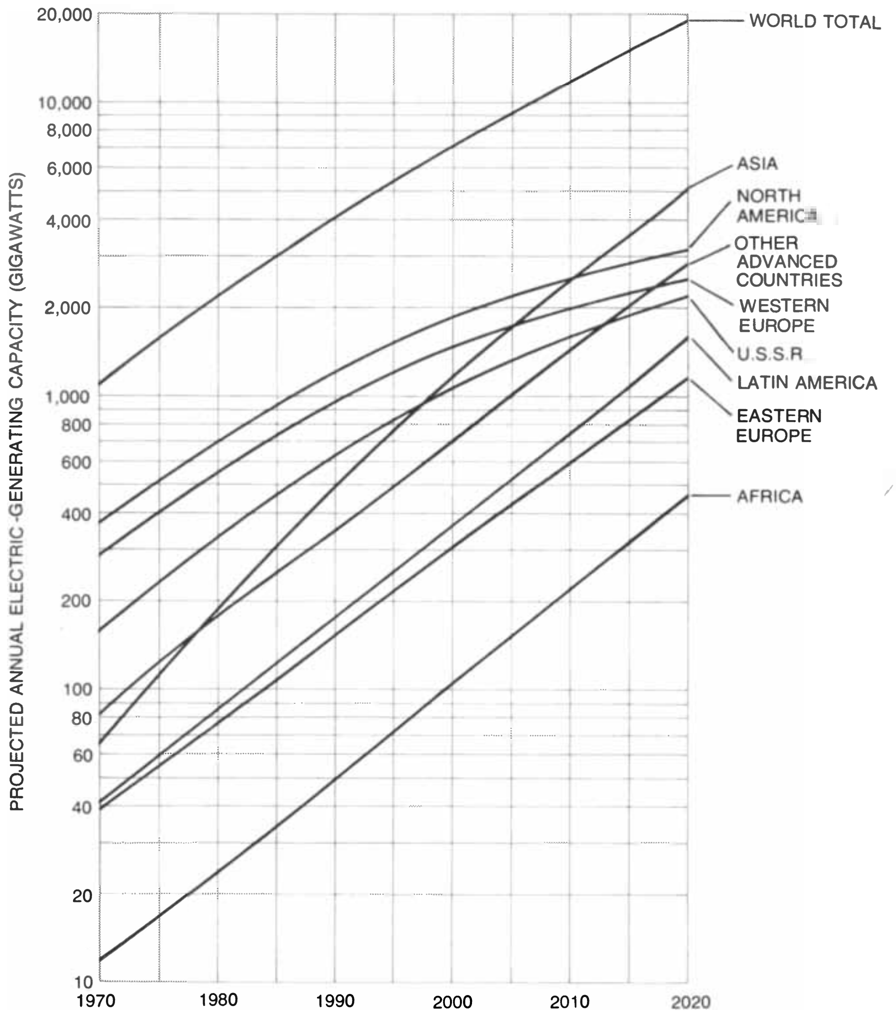
The first proposal for a nuclear-free zone in Africa was made in 1960, after the first nuclear-test explosion in the Sahara by France. At that time eight African countries raised the matter but did not press it. The following year 14 African states formally proposed in the UN General Assembly a resolution for preventing the extension of the nuclear-arms race to Africa and for making Africa a denuclearized zone. The resolution was approved by the General Assembly. It called on all member states to refrain from conducting nuclear tests in Africa and from using the area for testing, storing or transporting nuclear weapons; it also asked them to consider and respect the continent of Africa as a denuclearized zone. The U.S.S.R. supported the proposal, but the U.S. and its allies found it unacceptable on the ground that the prohibition of testing meant an uninspected and uncontrolled moratorium.

In July, 1964, at the first Summit Conference of the Organization of African Unity, the heads of state and of government of the African countries issued a solemn declaration on the denuclearization of Africa and announced their readiness to undertake by treaty not to manufacture or acquire control of nuclear weapons. This declaration was endorsed at a summit conference of nonaligned countries held in October of the same year.

In 1965, 28 African states submitted a proposal in the General Assembly to endorse the declaration on the denuclearization of Africa issued at the summit conference the year before. The resolution was overwhelmingly approved by the General Assembly in an almost unanimous vote, including all the nuclear powers except France. In addition to endorsing the declaration on the denuclearization of Africa, the General Assembly reaffirmed the appeal to all states to respect the continent of Africa as a nuclear-free zone and to abide by the declaration; it called on all states not to use or threaten to use nuclear weapons in Africa, not to test, manufacture, use or deploy nuclear weapons in Africa and not to acquire such weapons or take "any action which would compel African states to take similar action." It urged the nuclear powers not to transfer nuclear weapons, information or technological assistance to the national control of any state in any form that could assist in the manufacture or use of nuclear weapons in Africa. Finally, it expressed the hope that the African states would initiate steps through the Organization of African Unity with a view to implementing the denuclearization of Africa.

Nothing further was done to implement the declaration or the Assembly resolution. Nine years later, however, in December, 1974, 26 African countries again proposed a resolution in the General Assembly that was unanimously adopted by all countries including France, which had some years earlier shifted its nuclear testing from Algeria to the Pacific island of Mururoa. The resolution reaffirmed the previous resolutions and again called on all states to refrain from testing, manufacturing, deploying, transporting, storing, using or threatening to use nuclear weapons on the African continent.

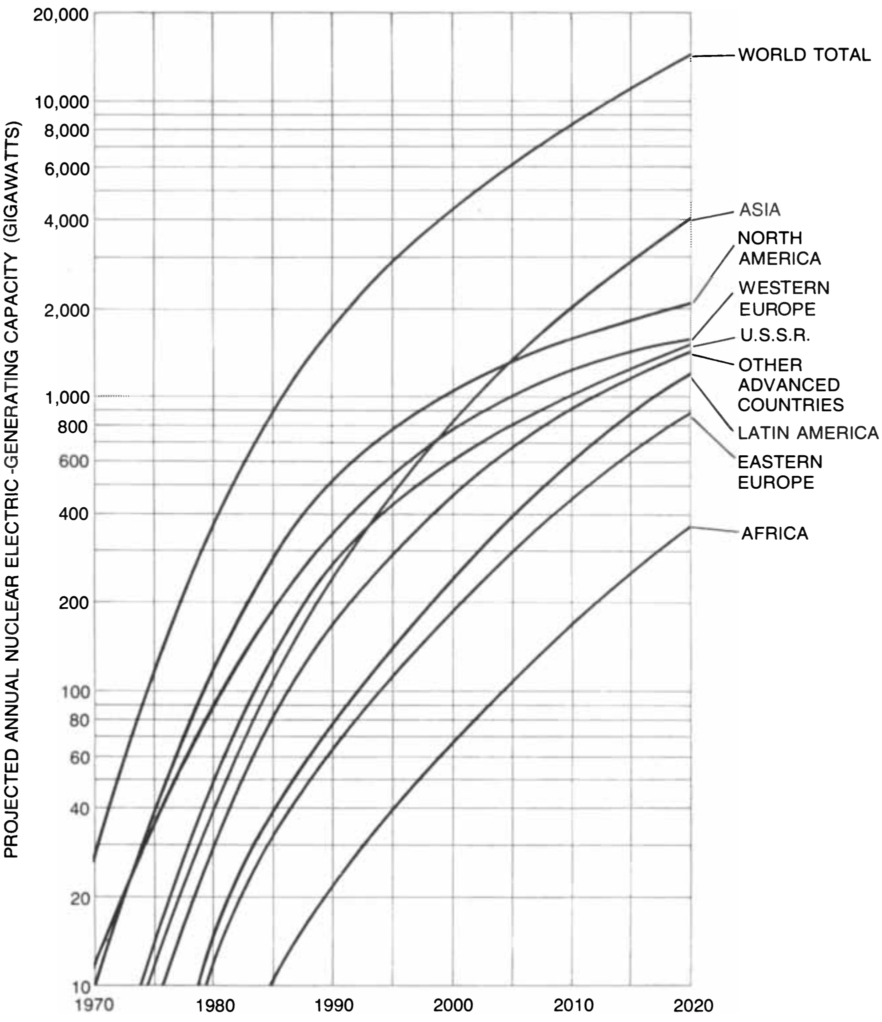
One of the main problems that have hindered the implementation of these declarations and resolutions is that the two African countries most technologically advanced in the field of nuclear energy are in an anomalous position. Although South Africa is a near-nuclear power that can readily go nuclear, most of the other African states have instituted a boycott of South Africa and refuse to have any traffic with it because of its policy of apartheid and its refusal to grant independence to Namibia (South-West Africa). At the other end of the continent Egypt, which is also a potential nuclear power, although it is well behind South Africa, does not wish to tie its hands in nuclear enterprises unless Israel also does so. The other African states are much less advanced in nuclear



PROJECTED GROWTH in the worldwide demand for electrical power is broken down by region in this graph, based on estimates compiled by Bernard Spinrad of Oregon State University. The curve labeled "Other Advanced Countries" includes Japan, Australia, New Zealand and South Africa. The graphs on these two pages appeared originally in a chapter by Spinrad in the book *Nuclear Proliferation Problems*, brought out last year by the Stockholm International Peace Research Institute. Data summarized in graphs represent updating of similar study conducted by Spinrad in 1970 for International Atomic Energy Agency.

technology; none of them can be considered as being a near-nuclear power and very few even as being potential nuclear powers.

Although it would be quite feasible to establish a nuclear-free zone in Africa without the participation of South Africa or Egypt, there seems little urgency and hence not much impetus in that direction. The matter is therefore likely to proceed slowly and perhaps will have to await the achievement of political accommodations that would alter the situation with respect to South Africa and Egypt. Nevertheless, as was demonstrated at the Organization of African Unity conference in 1964, and by the unanimous support of the black African and Arab states of the three General Assembly resolutions, there is a community of interest in Africa for the creation of a nuclear-free zone. It would not require a great deal of effort to achieve agreement even if the zone did not at first include all the countries of the continent. Moreover, it is likely that the 14 African countries that are not parties to the Nonproliferation Treaty would become parties to a treaty creating a nuclear-free zone in Africa. It would seem to be wise to encourage them to pursue their regional efforts. Such efforts would have the maximum chance for success precisely because there is no prospect of any of these countries going nuclear in the immediate future or even acquiring a nuclear option. Once a country acquires that option, it becomes much harder to persuade it to give it up.



ROLE OF NUCLEAR POWER in meeting the growing demand for additional electrical-generating capacity is projected by region in this graph on the assumption that over the next few decades nuclear fission is going to become increasingly the most economical means of generating electrical power. At present nuclear power represents approximately 20 percent of the new electrical-generating capacity installed annually in the world; after 1990 this figure is expected to rise to roughly 90 percent. (The penetration of nuclear power into *total* installed capacity will, according to Spinrad, lag behind by another 10 to 15 years.)

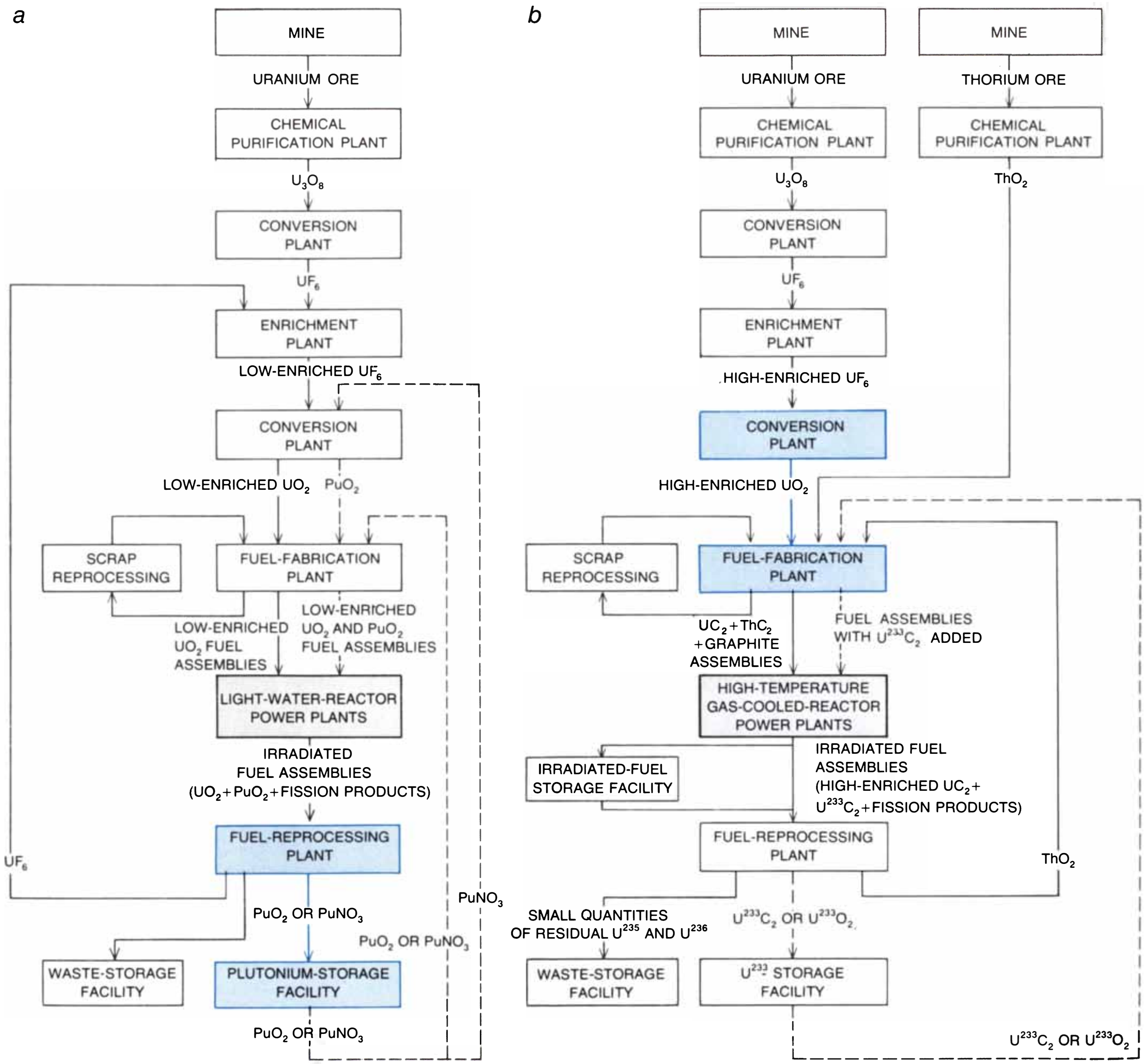
In 1974 Iran, later joined by Egypt, proposed the establishment of a nuclear-free zone in the Middle East. In December of that year the General Assembly commended the idea and called on all parties in the area to proclaim their intention to refrain on a reciprocal basis from producing, testing, acquiring or in any other way possessing nuclear weapons. It also called on the parties in the area to accede to the Nonproliferation Treaty and asked the Secretary-General to ascertain their views and to report to the Security Council and to the General Assembly.

Contrary to what had happened with the Treaty of Tlatelolco and the declaration of the African heads of state and of government, there was no prior consultation with important states in the Middle East and in particular with Israel, the state with the most advanced nuclear technology in the area. Moreover, the General Assembly's approval of the proposal did not call on the parties concerned to enter into consultations with one another but merely asked that the Secretary-General ascertain their views. Because of the state of belligerency existing between Egypt and Israel, the fact that Egypt joined Iran in sponsoring the proposal led to the suspicion that it was maneuvering to gain, if not some military advantage by putting pressure on Israel not to go nuclear, then at least some political or propaganda advantage. Any such suspicion would of course have the effect of aborting the project and practically ensuring that no nuclear-free zone in the area would be accepted by all parties.

As was to be expected, Israel withheld its support from the General Assembly's resolution. It is interesting to note that all the nuclear powers, including India, voted for it. Without Israel's support, however, it is most unlikely that the resolution will lead to a productive result. Since Egypt has announced that it will not ratify the Nonproliferation Treaty unless Israel does so, it is hardly likely that Egypt would become a party to a nuclear-free zone that did not include Israel.

Some political observers have speculated that the fact that Iran accepted Egypt as a cosponsor of its proposal was probably done with the understanding that this would almost certainly ensure the failure of the project, and that this cast some doubts on the sincerity of Iran's motives. In any case, if the idea of a nuclear-free zone in the Middle East is to make progress, it will be necessary either to start anew with prior consultations with Israel or to envision as a beginning a zone of more modest scope that does not include Egypt or Israel. It is not clear whether Iran or other Arab states such as Libya, which has embarked on a program to acquire nuclear reactors, would be attracted by a limited zone. The entire project may have to be postponed until there is some settlement between Israel and the Arab states. On the other hand, Israel's willingness to agree to be a party to such a zone might become a useful bargaining point that could help to facilitate an Arab-Israeli political settlement. The Arab states, and in particular Egypt, would certainly be glad to see Israel give up its nuclear option, and they might conceivably consider paying a price for that.

Soon after India exploded a nuclear device Pakistan proposed the creation of a nuclear-free zone in South Asia. The General Assembly also endorsed this concept but, contrary to what it had done with the resolution for a Middle East zone, invited the states of the region and other neighboring non-nuclear states to initiate necessary consultations with a view to establishing a nuclear-free zone. It urged them to refrain in the



NUCLEAR-FUEL CYCLES required to support three of the main types of nuclear reactor likely to be used to generate electrical power around the world in the next few decades are depicted schematically in the diagrams on these two pages. The three types of reactor system considered are the light-water reactor (*a*), which is expected to continue being used in most nuclear power plants operating in the U.S. at least through 1980; the high-temperature gas-cooled reactor (*b*), which is expected to play an increasingly important role after 1980, and the liquid-metal-cooled fast-breeder reactor (*c*), which many experts believe will become the mainstay of the world's nuclear-power industry toward the end of the century. The portion of each fuel cycle from which high-grade fission-

interim from any contrary action and asked the Secretary-General to convene a meeting for the purpose of the consultations.

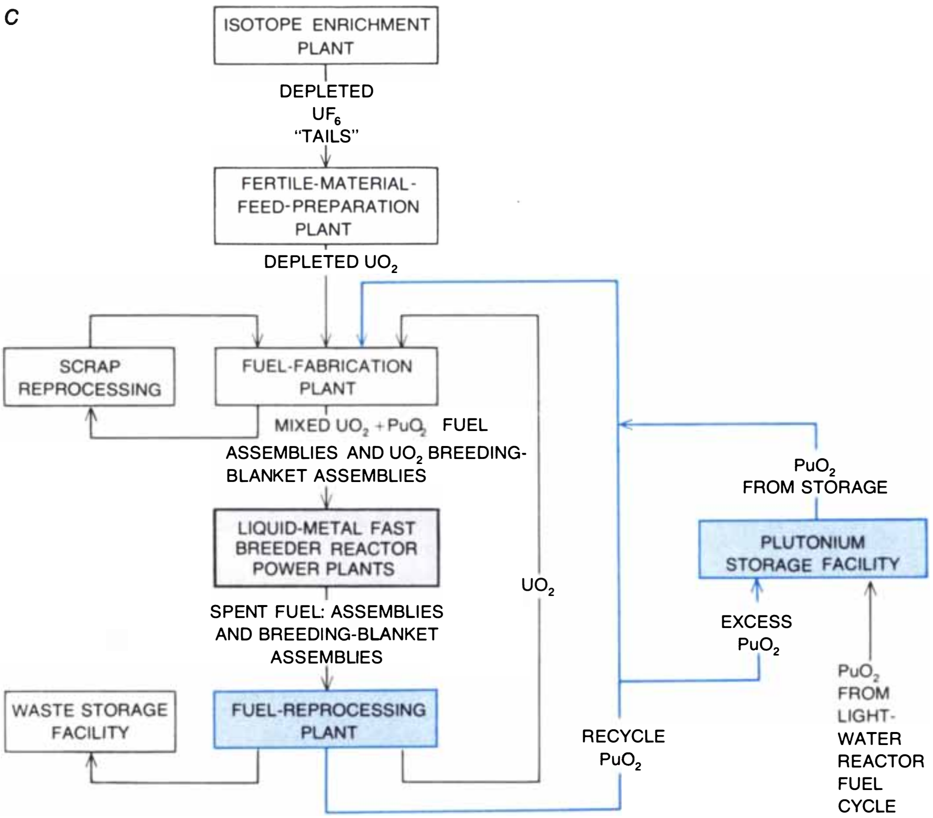
Here too, contrary to what had happened with the Treaty of Tlatelolco and the African declaration, there was no prior consultation among the parties concerned—in particular with India, which had already conducted a nuclear explosion. These circumstances again gave rise to the suspicion that the Pakistani initiative was aimed not so much at the actual creation of a nuclear-free zone as at gaining some political or propaganda advantage.

India made its position clear by proposing a resolution by which the General Assembly would state that the "initiative" for the creation of a nuclear-free zone in the appropriate region of Asia "should come from the states of the region concerned," taking into account the special features and geographical extent of the region. The Indian resolution was adopted by an overwhelming majority, including the U.S.S.R., but with the U.S., Great Britain, China, France, Israel and Pakistan abstaining. On the other hand, the Pakistani resolution was supported by the U.S. and China, with India voting against it and the U.S.S.R., Great Britain, France and Israel abstaining from the vote.

It is understandable that India might have had some resentment or reservations about not having been consulted in advance of Pakistan's initiative. Nonetheless, if India is serious in its protestations that its explosion of a nuclear device was solely for peaceful purposes and that it has no intention of developing nuclear weapons, there are sound reasons for India to react positively to Pakistan's proposal and to work for the creation of a nuclear-free zone in South Asia. It would certainly be to India's advantage if Pakistan did not go nuclear and to its disadvantage if Pakistan did. India might also find it useful to have a pledge from China not to use or threaten to use nuclear weapons against it as a member of a South Asian zone that China undertook to respect as such.

In short, India may have reacted too hastily in rejecting the Pakistani proposal. If Pakistan should go nuclear, it is unlikely that Iran and Indonesia would refrain from doing likewise, to the detriment of all four of the countries as well as to others in the area. In fact, one can make out a good case, not least in India's interest, for expanding the Pakistani proposal to include in the area of the zone both the Middle East and the Far East. Since only Iran, of the four countries mentioned, is a party to the Nonproliferation Treaty, the idea of an expanded denuclearized zone certainly merits further exploration.

C

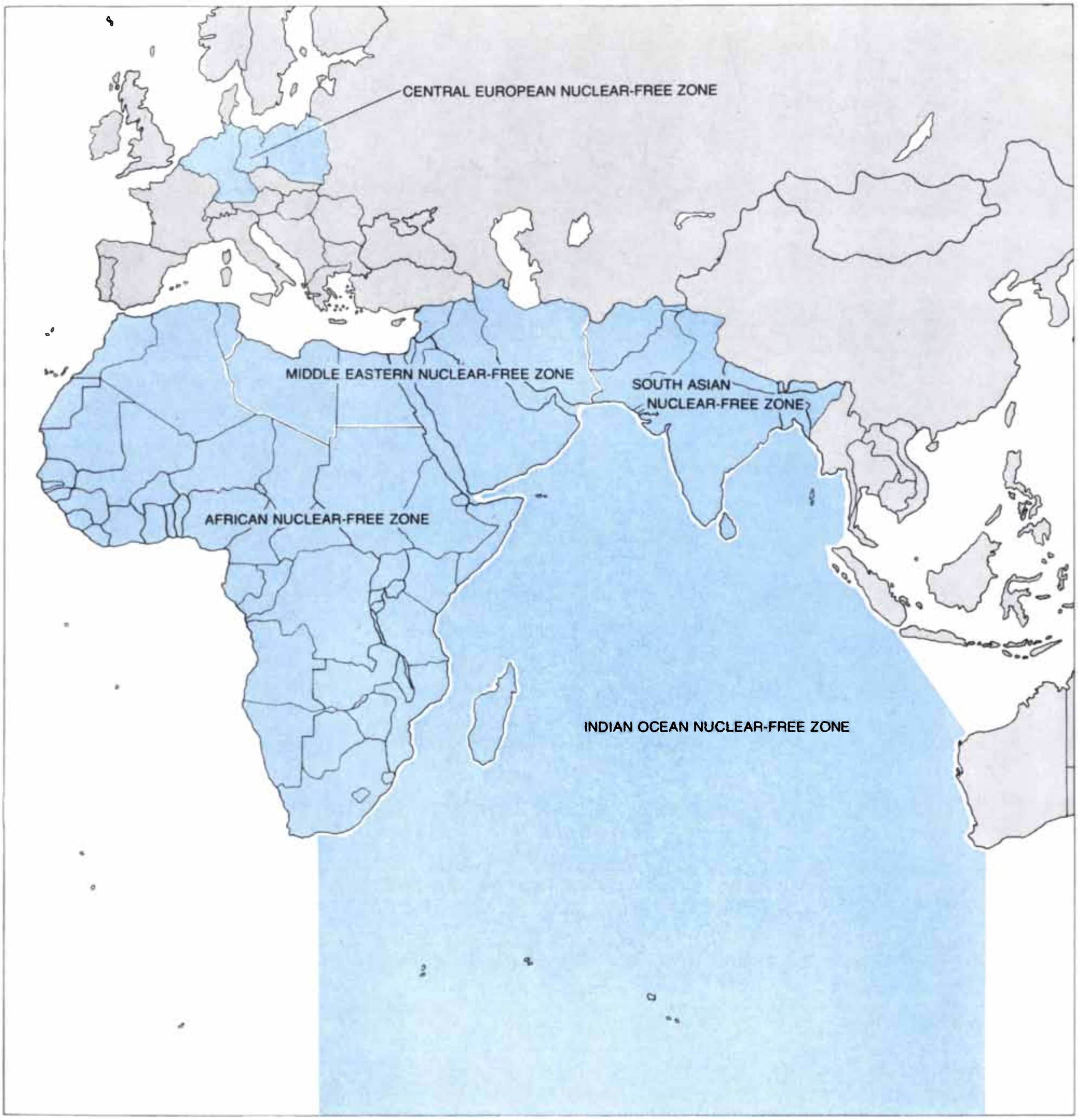


U_3O_8	URANIUM OXIDE ("YELLOWCAKE")	$PuNO_3$	PLUTONIUM NITRATE
UF_6	URANIUM HEXAFLUORIDE	ThO_2	THORIUM OXIDE
UO_2	URANIUM OXIDE	UC_2	URANIUM CARBIDE
PuO_2	PLUTONIUM OXIDE	ThC_2	THORIUM CARBIDE

able material could be diverted directly into the production of nuclear weapons is indicated in color. The diagrams are adapted from *Nuclear Theft: Risks and Safeguards*, by Mason Willrich and Theodore B. Taylor, a report to the Energy Policy Project of the Ford Foundation published in 1974. Not included is Canadian natural-uranium reactor system, the spent fuel of which is not normally reprocessed to produce plutonium (see "Natural-Uranium Heavy-Water Reactors," by Hugh C. McIntyre; SCIENTIFIC AMERICAN, October).

The concept of a "peace zone" is considerably broader than that of a nuclear-free zone, but it includes the latter concept. In 1971 the UN General Assembly, on the initiative of Ceylon, declared that the Indian Ocean was designated for all time as a zone of peace. The declaration called on the U.S. and the U.S.S.R. "to enter into consultation with the littoral states of the Indian Ocean with a view to (a) halting the further escalation and expansion of their military presence in the Indian Ocean; (b) eliminating all bases, military installations and logistical supply facilities, the disposition of nuclear weapons and weapons of mass destruction, and any manifestation of Great Power military presence in the Indian Ocean conceived in the context of Great Power rivalry." This declaration, if implemented, would amount to the denuclearization of the Indian Ocean, at least by the major nuclear powers, and indeed would almost amount to its demilitarization so far as they are concerned. China and India both voted in favor of the creation of the Indian Ocean peace zone but the U.S., the U.S.S.R., Great Britain and France abstained.

The following year a committee was established, composed of littoral states and other supporters of the peace zone, such as China and Japan, to study the implications of the declaration. In 1973



FIVE PROPOSED NUCLEAR-FREE ZONES are indicated on this map by the colored areas. The proposal for a Central European nuclear-free zone, sometimes referred to as the Rapacki Plan, was originally put before the United Nations General Assembly by Poland in 1958. The first of many proposals for such a zone in Africa was made in 1960. Nuclear-free zones in the Middle East and South Asia were advocated in 1974 by Iran and Pakistan respectively. The concept of a demilitarized "zone of peace" in the Indian Ocean goes back to 1971, when Ceylon put the question before the General Assembly. Although none of these proposals has so far made much progress, the author believes that in at least one case, that of Africa, there is a good chance that another nuclear-free zone can be established in most, if not all, of the continent (temporarily excluding, for example, Egypt and South Africa).

the General Assembly urged all states to accept the principles and objectives of the declaration, but again the four nuclear powers mentioned above withheld their support.

It should be noted that the peace zone does not seek to establish a nuclear-free zone in the territories of the littoral states of the Indian Ocean but rather is aimed at the denuclearization and further demilitarization of the Indian Ocean itself and of naval bases in the ocean. Therefore it differs fundamentally from other nuclear-free-zone proposals in that it is intended to apply to the high seas rather than to the bordering states. Although the actual establishment of such a peace zone is obviously not likely to take place in the near future, it is noteworthy that the concept is supported by practically all the littoral states of the Indian Ocean, including Australia, and that consultations among these states are proceeding and are acquiring increasing support.

The failure of the Rapacki Plan for a nuclear-free zone in Central Europe, and of other initiatives that were no less persistently pursued, was due to the fact that they were aimed at altering existing security arrangements or the existing military balance in the area. On the other hand, the Latin-American initiative succeeded because it was a genuine cooperative effort undertaken by the countries of the region in order to keep nuclear weapons out of their area. The initiative was not aimed at any particular country or security arrangement but was perceived by the parties concerned as being in their common interest.

Moreover, the Latin-American ccuntries went ahead with their project without requiring the participation of every country in the area. They were content to go forward with establishing the zone in the maximum area possible. If Cuba was not ready to join or the U.S.S.R. to give its support, the Latin-American countries were prepared to begin on a smaller scale and to work toward achieving their full goal in due course.

Since the African nuclear-free zone was also conceived as a genuine cooperative effort in the common interest of the African countries, it too has a chance of success. Here too, however, if the zone is to be created in the foreseeable future, it may be necessary for the countries of Africa to proceed without the participation of South Africa and Egypt. Even though the absence of these two countries, which are the most advanced in nuclear technology, would be a shortcoming that would reduce the effectiveness of the nuclear-free zone, there would always be the possibility of their joining later.

With respect to the other proposals for nuclear-free zones that are now being actively considered, namely for the Middle East and South Asia, the situation is entirely different. Since here the proposals seem to be more of a strategic and political move than a cooperative effort conceived and worked out in joint consultations by the main countries of the respective areas, the prognosis for them must be considered poor. The situation might, of course, change if genuine peace and minimum conditions of confidence are established between Israel and the Arab states or between India and Pakistan. In the absence of such developments it is hardly likely that a country can be politically maneuvered into accepting a nuclear-free status. No country would freely agree to become a party to any treaty, let alone one that could vitally affect its security, unless it considered that doing so was clearly in its interest, or at least not prejudicial to its interests. In fact, it is axiomatic that even if a country becomes a party to a treaty, it will not remain a party if it considers that events have changed the basic situation or have caused the treaty to be against its interests.

Nevertheless, the idea of nuclear-free zones is undoubtedly a good one. It provides a means whereby non-nuclear countries can, by their own initiative and effort, ensure their greater security. It can be an effective instrument not only to prevent non-nuclear countries in a given region from going nuclear but also to prevent the stationing or deployment of nuclear weapons in the countries of that zone by the nuclear powers. In addition it can be a means of obtaining pledges from the nuclear powers not to use or threaten to use nuclear weapons against any of the countries in the area of the zone. The U.S. and Great Britain, which refused to include such a pledge in either the Nonproliferation Treaty or the Security Council resolution for security assurances to non-nuclear states, have both agreed to a commitment of this kind by signing and ratifying Protocol II of the Treaty of Tlatelolco. China and France, which are not parties to the Nonproliferation Treaty or to the Security Council's security assurances, have likewise become parties to Protocol II. As I have indicated, it is also possible that the U.S.S.R. will become a party once relations between Cuba and the other Latin-American countries are normalized. Furthermore, nuclear-free zones would provide a sound reason and a logical basis for promoting the peaceful uses of nuclear energy by facilitating the establishment of regional nuclear-fuel-cycle centers for enriching uranium, reprocessing plutonium and handling nuclear wastes. In fact, it is not easy to visualize how regional or multinational fuel-cycle centers could be established except within some nuclear-free zone. The economic and security benefits, and the greater effectiveness of safeguards against diversion or theft of nuclear material, that would follow from having a large-scale regional center rather than a number of smaller national facilities are themselves powerful arguments in favor of the creation of such zones.

Moreover, such zones could avoid the discriminatory features of the Nonproliferation Treaty and provide the mutual security desired by countries that for one reason or another do not want to become parties to the agreement. Thus nuclear-free zones can be a very effective way of promoting and strengthening the nonproliferation regime. The creation of such zones would in no way conflict with the Nonproliferation Treaty but would provide a means for extending and reinforcing the objectives of the treaty.

Provided that the idea of a nuclear-free zone is conceived and promoted by the non-nuclear countries of a given area as a genuine effort to increase the security of all the countries affected and not just some of them, there is no reason additional nuclear-free zones should not be established in different areas of the world. Their establishment would be facilitated if the nuclear powers would demonstrate their wholehearted support for the creation of such zones not only as a means of preventing the further spread of nuclear weapons but also as a way of furthering the peaceful uses of nuclear energy and of providing security assurances to the countries in the denuclearized zone.

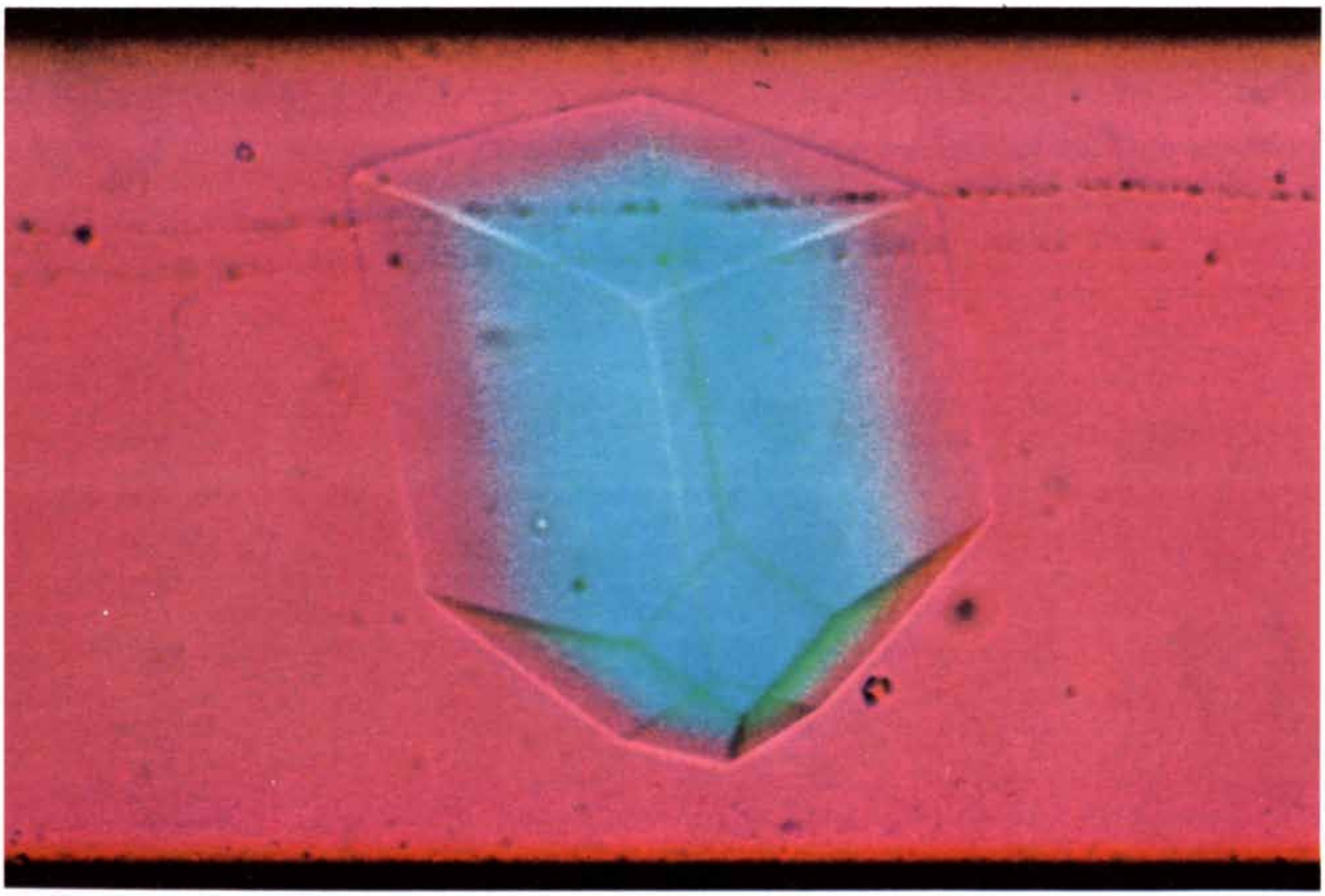
The Nonproliferation Treaty itself, although it does not explicitly encourage or facilitate the creation of nuclear-free zones, does nothing to discourage or detract from them. In fact, it gives them its blessing. The initiative, however, remains with the parties in any region. The nuclear-free zone is one of the easiest and best ways by which the non-nuclear countries can do something by and for themselves. In a number of important respects it provides a more effective means than the Nonproliferation Treaty does of maintaining and strengthening international peace and security.

THE PROTEIN SWITCH OF MUSCLE CONTRACTION

The contraction of muscle is "turned on" when calcium ions trigger changes in the conformation of two regulatory proteins: tropomyosin and troponin. The detailed nature of these changes is now emerging

by Carolyn Cohen

Biologists have always sought to understand how the form of living things—the shape of a hand, say, or of a leaf—is adapted to their function. We can now begin to see how the shapes even of molecules within cells are related to the mechanisms of biological processes. By constructing models representing the main features of the molecules and the dynamics of their principal interactions we try to describe the relation between form and function at the molecular level. The molecular structures are established by biochemical techniques or more direct methods such as electron microscopy or X-ray diffraction. This approach is well illustrated by investigations of the complex molecular architecture of muscle cells and its relation to the way muscles contract. Here I shall show how various

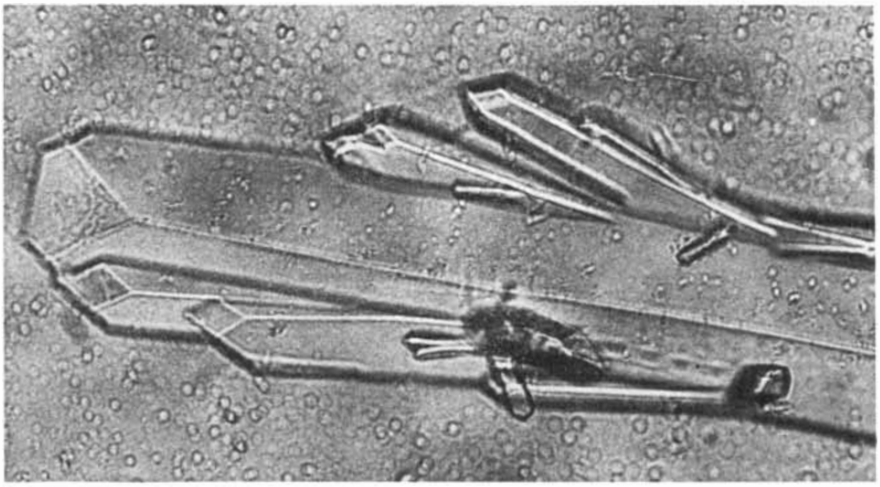


TROPOMYOSIN CRYSTAL grown in the author's laboratory was photographed in polarized light. The tropomyosin, one of two proteins involved in regulating muscle contraction, was extracted from rabbit heart muscle. The crystal (half a millimeter on a side) was grown in a capillary tube; it is so fragile (it is more than 90 percent water) that it must be kept submerged in its "mother liquor" during X-ray exposure. The pattern formed by X rays reflected from the crystal provides information on tropomyosin structure.

lines of evidence are beginning to reveal the protein assembly that turns contraction on and off.

A muscle fibril is an array of two kinds of filaments, thick ones and thin ones [*see illustration on next page*]. The thick filaments are composed almost entirely of myosin, a protein whose molecules have a long rod-shaped tail region and two globular heads. The myosin molecules are arrayed in the thick filament with the tails forming the core and the heads at the surface. The molecules are oriented in one direction on one side of the center and in the opposite direction on the other side. The thin filaments are composed primarily of another protein, actin, whose roughly spherical molecules are arrayed like a twisted double strand of beads. A muscle contracts when the projecting heads of the myosin molecules in the thick filaments reach out to make contact with the actin molecules; while the heads are attached their concerted movements pull the opposed sets of thin filaments toward each other, and the fibril shortens. The signal that initiates these events is the liberation of calcium ions from membrane-bound storage sites by the arrival at the muscle of a nerve impulse. The signal is detected and acted on by two further sets of protein molecules, tropomyosin and troponin, which are positioned along the actin strand and with it constitute the thin filament. The calcium ions bind to the troponin, which then modifies the position of the tropomyosin molecules so that the myosin heads can contact the actin molecules [see "The Cooperative Action of Muscle Proteins," by John M. Murray and Annemarie Weber; SCIENTIFIC AMERICAN, February, 1974].

At this stage in our understanding of muscle contraction there are two major areas to be investigated. One is the mechanism by which the thin filaments are pulled past the thick filaments by the myosin heads. The other area, in which our laboratory at Brandeis University has been active, is the nature of the troponin-tropomyosin switch: How do the calcium ions released by the motor-nerve impulse cause troponin to affect tropomyosin, and what changes in tropomyosin subsequently make the attachment of myosin to actin possible? To answer these questions we need to know the structure of tropomyosin and troponin in precise detail.



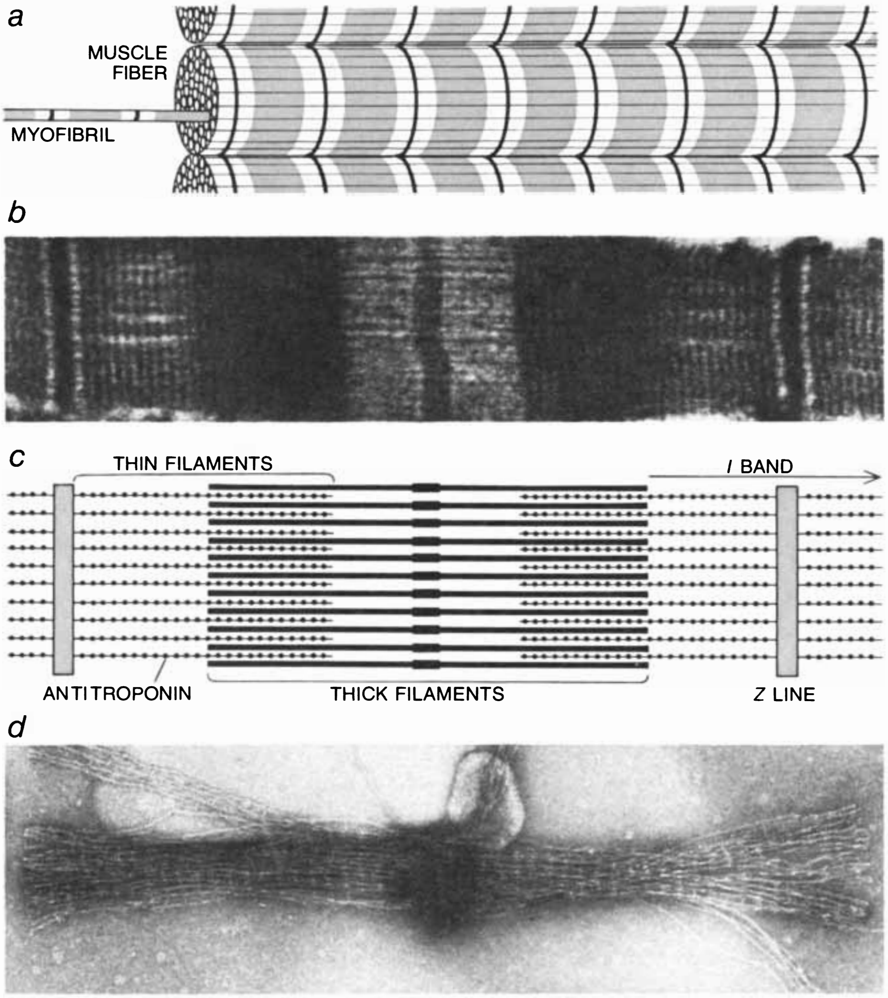
TROPONIN SUBUNIT, *TnC*, was crystallized from purified rabbit *TnC* by D. Mercola, B. Bullard and J. Priest of the University of Oxford. The crystals were photographed in unpolarized light and are enlarged 230 diameters. Troponin, the other protein in muscle regulation, is a complex of three subunits. Calcium-dependent changes in the *TnC* subunit appear to be the first in a series of structural events that take place during regulation.

The analysis of a biological system seldom proceeds in the logical way that my brief review of muscle contraction may have implied. Tropomyosin was actually discovered more than 20 years before its function was recognized. Toward the end of World War II Kenneth Bailey of the University of Cambridge isolated from ground muscle a protein with unusual properties. Its molecules were long and thin and remarkably stable, and it formed beautiful large crystals that were more than 90 percent water. W. T. Astbury of the University of Leeds, one of the pioneers in the study of biological molecules by X-ray diffraction, soon found that the X-ray pattern of Bailey's protein classified it as a member of a group he called the alpha proteins: elongated, fibrous proteins that include the keratins of mammalian skin and hair and myosin, the main protein in muscle. Like these proteins, the new muscle protein yielded an X-ray pattern indicating that its constituent polypeptide chains of amino acids were folded in some regular way that repeated every 5.1 angstroms along its axis. (An angstrom is a ten-millionth of a millimeter.) Bailey found a general similarity between the amino acid composition of his new protein and that of myosin; the similar compositions and the X-ray patterns led him to believe the new protein might be a precursor of myosin, and so he named it tropomyosin. The name stuck, even though Bailey himself later showed that tropomyosin is not a precursor of myosin; its true function remained a mystery for some time, until after its structure was demonstrated.

Astbury had deduced that the polypeptide chains in the alpha proteins were folded in some regular way. Various helical structures were proposed prior to 1951, when Linus Pauling and his colleagues at the California Institute of Technology recognized the "inner logic" of polypeptide chain folding. Through ingenious model building and crystallographic study of the precise arrangement of small groups of amino acids they discovered the alpha helix [*see top illustration on page 39*]. Pauling reasoned further that in the complex fibrous proteins the alpha helixes would tend to twist around one another and form a coiled coil. In 1953 F. H. C. Crick of the University of Cambridge independently pointed out that a coiled coil would be formed by the regular interlocking of side chains of the alpha helixes, and he calculated their detailed diffraction patterns. He proposed that the alpha proteins must be two-chain or three-chain structures, with the chains wound around one another like the strands of a cable. He noted that tropomyosin's amino acid composition would lend itself to a stable coiled coil.

In 1957 Andrew G. Szent-Györgyi and I found that tropomyosin did indeed have a very large alpha-helix content. In 1960 Kenneth Holmes joined us at the Children's Cancer Research Foundation in Boston, bringing with him a technique, in which he had been trained by Rosalind Franklin at Birkbeck College in London, for focusing X rays so that clear diffraction patterns could be obtained from specimens (such as muscles) that include large amounts of water. We began to look at muscle tissue, in particular at the "catch" muscle of mollusks, which contains large amounts of paramyosin, an alpha protein that is structurally similar to tropomyosin. The high degree of orientation of the molecules in the native muscle meant that it was in effect a paracrystalline array of alpha proteins. Its X-ray pattern revealed for the first time the X-ray reflection that is characteristic of the distance between turns of the alpha-helical coiled coil. From the position of the reflection we could deduce that paramyosin (and therefore probably tropomyosin and the rod portion of myosin) were two-chain alpha-helical coiled coils [*see middle illustration on opposite page*]. The X-ray evidence fitted nicely with physicochemical measurements that had been made by Alfred M. Holtzer of Washington University and Susan Lowey of our group. And so by 1963 we pretty well knew the basic features of tropomyosin's structure.

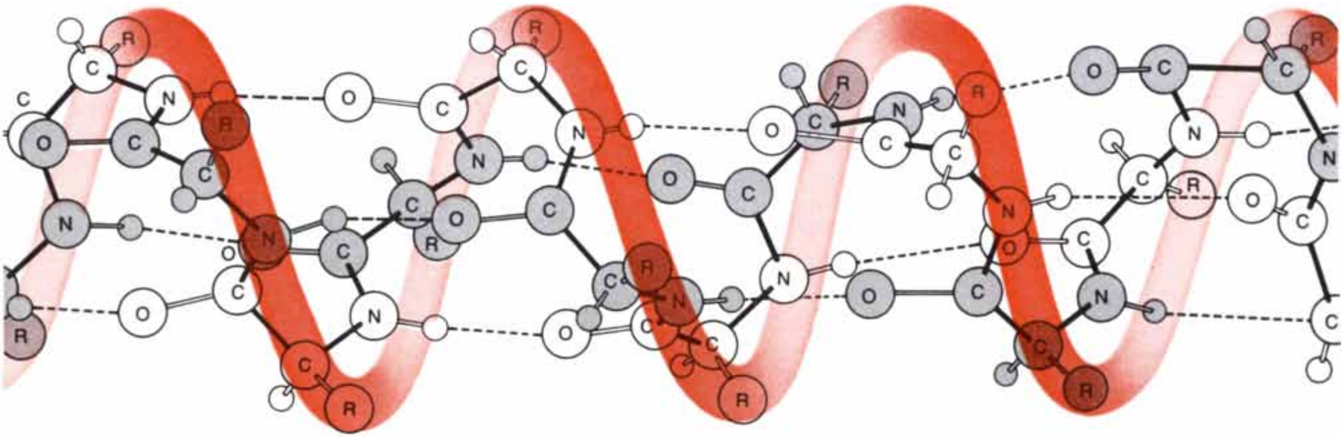
The first clue to tropomyosin's function was its location in muscle. Several lines of evidence pointed to its association specifically with the *I* band of the myofibril, the region that contains the thin filaments. H. E. Huxley and Jean Hanson of the Medical Research Council units in Cambridge and at King's College in London learned from interference microscopy and biochemical measurement that there was more material in the *I* band than could be accounted for by actin alone; Samuel V. Perry and A. Corsi found that tropomyosin and actin appeared to be obtained together when proteins were leached out of the fibrils. In a classic paper in 1963 Hanson and Jack Lowy showed the two-strand structure of the actin filaments in electron micrographs and suggested that the long tropomyosin molecules might lie in the two grooves of the actin helix.



MYOFIBRIL, the functional unit of a vertebrate skeletal muscle fiber (*a*), is seen in an electron micrograph (*b*) of a thin section made by Iwao Ohtsuki of the University of Tokyo. As the diagram (*c*) shows, the myofibril is composed of interpenetrating arrays of thick and thin filaments. A muscle contracts when projecting heads on the myosin molecules of thick filaments pull sets of thin filaments toward each other. The thin filament is composed of three proteins: actin, tropomyosin and troponin. A bundle of thin filaments separated with their *Z* line (*center*) from glycerinated chicken muscle is negatively stained in another micrograph made by Ohtsuki (*d*). Both muscle specimens had been treated with an antibody against troponin. The antibody binds to troponin, showing that the troponin complexes are positioned about every 400 angstroms along the one-micrometer-long thin filaments.

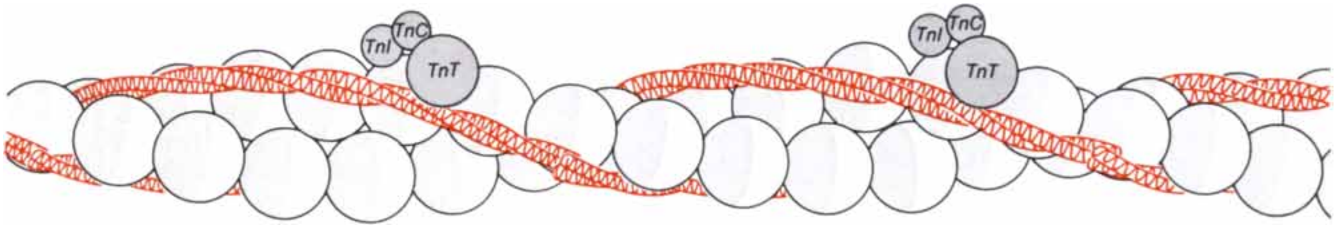
By this time the sliding-filament theory of muscle contraction, proposed by Huxley and Hanson and independently by Andrew F. Huxley and R. Niedergerke of the University of Cambridge, had become well established [see "The Mechanism of Muscular Contraction," by H. E. Huxley; SCIENTIFIC AMERICAN, December, 1965]. Information on the control of contraction was also being developed. Calcium ions had been shown to be involved as early as 1947. In 1952 B. Marsh, a student in Bailey's laboratory, isolated from homogenized muscle a factor that promoted relaxation in muscle fibers and whose effect was abolished by calcium ions. Annemarie Weber, who was then working at Columbia University, made the penetrating suggestion in 1959 that the relaxing factor acted by binding calcium. Setsuro Ebashi of the University of Tokyo showed that the relaxing factor consisted of bits of the sarcoplasmic reticulum, a specialized membrane system that surrounds the muscle fibrils. Following the arrival of a nerve impulse a signal is carried to this system, which then releases calcium to the fibril, causing it to contract; at the end of the signal calcium is again sequestered, causing the fibril to relax.

Just what does the calcium do to the fibril's proteins? Following a related but widely unrecognized finding made in 1956 by Perry and T. Grey, Weber made a crucial observation bearing on the question in 1961. She noticed that certain preparations of actin and myosin were sensitive to the effect of calcium and other preparations were not. This sensitivity appeared to depend on some property of the actin. She discussed her observations with Ebashi, who did a dramatic experiment. He prepared natural actomyosin from muscle; when he added calcium in the presence of ATP (the molecule whose splitting by myosin was known to provide the energy for contraction), the result was "superprecipitation," the test-tube equivalent of contraction. Without calcium there was no reaction. When he made actomyosin by combining purified actin and purified myosin, however, the proteins were no longer affected by calcium: they superprecipitated with ATP regardless of the calcium level. Ebashi inferred correctly that the purified system lacked some critical component. He added to the purified system a protein fraction similar to Bailey's tropomyosin and found that calcium sensitivity was restored. Highly purified tropomyosin, on the other hand, did not restore calcium sensitivity. It turned out that the "native" (unpurified)



ALPHA HELIX, a configuration that underlies the structure of many proteins, including tropomyosin, is a coiled chain of amino acid subunits. Each subunit consists of a carbon atom with a characteristic radical, or side group (*R*), flanked by CO and NH groups. In this drawing successive subunits are alternately gray and white. The helix is braced by hydrogen bonds (*broken lines*) linking the hydrogen (*small circle*) of each NH group with the oxygen of the fourth subunit along the chain. In a typical alpha helix there are 3.6 subunits in each complete turn of the helix. In this drawing the colored tube connects the radicals, tracing the turns of the helix.

COILED COIL is formed when two alpha helixes wind around each other, side chains interlocking to form a stable structure. There are about 36 turns of an alpha helix in one turn of the larger coiled coil, but this "axial repeat" varies with small changes in the helix.



COILED COILED COIL is formed by the filament of tropomyosin molecules (coiled coils) wound in each of the two grooves of the actin helix in a thin filament of muscle. The actin molecules (*gray and white*) are polar; they all point in the same direction in a double helical array. The tropomyosin filaments (*color*) consist of polar tropomyosin molecules, bonded head to tail, that lie near the grooves, each molecule spanning seven actin monomers. A troponin complex (*dark gray*) is about a third of the way from one end of each tropomyosin molecule. In this schematic diagram only the troponin complexes on one side of the actin helix are shown.

tropomyosin contained another protein, which Ebashi and his colleague Ayako Kodama separated and named troponin. Neither troponin nor tropomyosin was effective alone; it was the tropomyosin-troponin complex that made muscle sensitive to calcium. Some 20 years after tropomyosin was discovered its regulatory role in muscle contraction was beginning to emerge.

The regulatory system is still understood only in a general way, with details of the mechanism becoming known as the structures of the components are revealed. Ebashi and his colleagues soon learned that troponin is a complex of globular proteins; in muscle it is positioned at intervals of about 400 angstroms along the thin filament (thus accounting for a 400-angstrom periodicity that had been noted in micrographs of the *I* band). The thin filament, however, was a double helix of identical globular actin molecules, each about 55 angstroms in diameter, that made half a turn every 360 to 370 angstroms. What specified the attachment of the troponin complexes to the actin helix at 400-angstrom intervals? The linkage proved to be through the tropomyosin molecule, but that became clear only when the precise dimensions of the proteins and their aggregates had been established. Measuring those dimensions now took on a new significance.

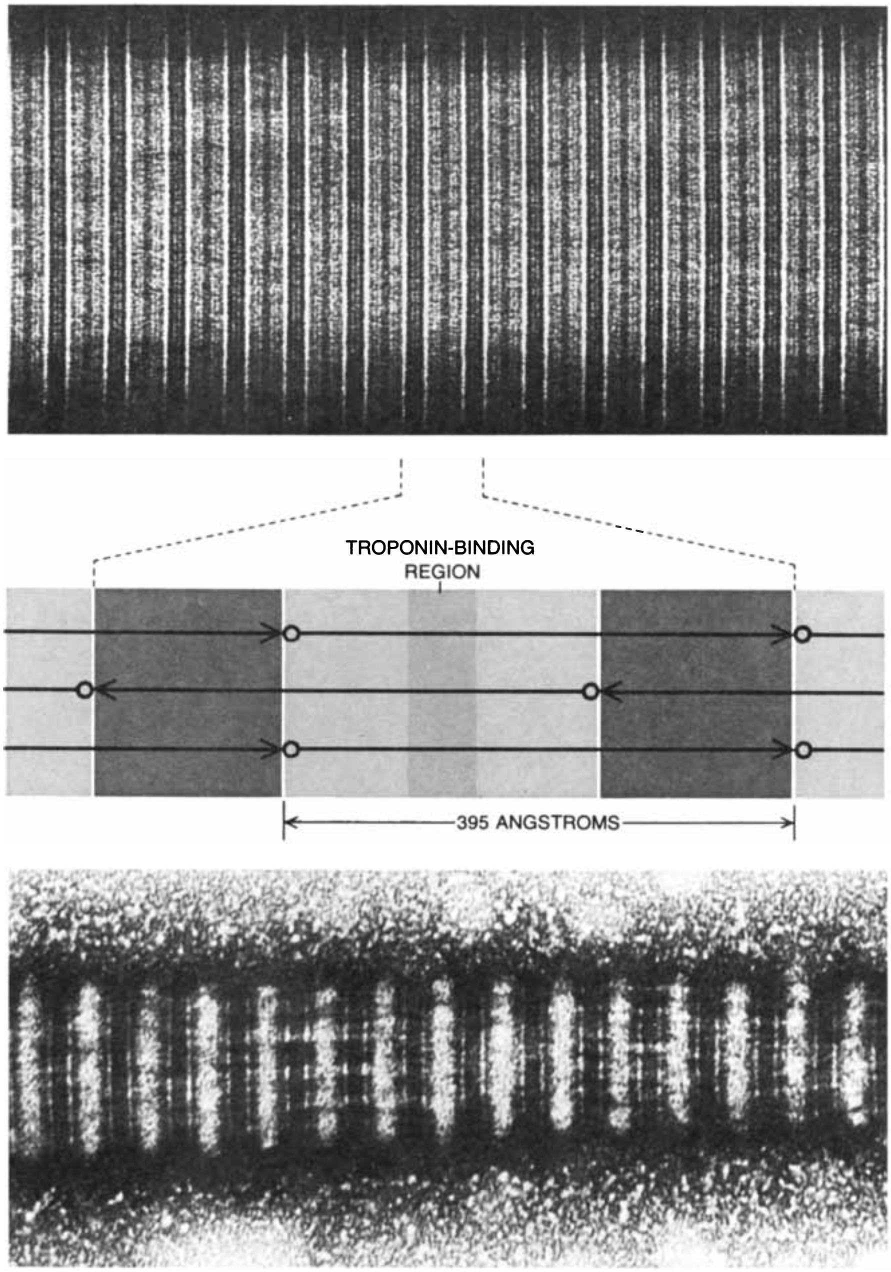
In 1965 Donald L. D. Caspar, William Longley and I began to study tropomyosin crystals by X-ray diffraction. We wanted to see how tropomyosin was affected by ions such as calcium, and we worked with low calcium concentrations, in the range that had been shown to be physiologically significant. One day a large excess of calcium was accidentally added to the solution; the result was an unusual, highly ordered fibrous form of tropomyosin. Repeating the procedure, we found that high concentrations of divalent positive ions such as calcium or magnesium caused the precipitation of tropomyosin in the form of needle-shaped paracrystals that were readily visible in a light microscope. Electron micrographs revealed a detailed banding pattern with a period of 395 angstroms—very close to what was then estimated to be the length of the tropomyosin molecule. When Longley and I first reported this form of tropomyosin early in 1966, we related the period of the paracrystal to the 400-angstrom period seen in the *I* band of muscle, but we were not aware of the findings on troponin then being made in Japan.

Ebashi, Makato Endo and Iwao Ohtsuki finally combined all the evidence in 1969 and proposed that the tropomyosin molecules lie end to end in the two grooves of the thin filament, positioning a troponin complex every 400 angstroms. At the same time they outlined the basic functional scheme of regulation. They viewed troponin as a calcium-dependent switch; the signal detected by troponin was transmitted to the actin-myosin system by tropomyosin, which thus served as "the mediator of the information from troponin." The regulation of muscle contraction could be thought of as being analogous to the regulation of genes, with troponin-tropomyosin as the repressor system and calcium ions as the derepressors.

Evidence for a movement of tropomyosin when muscle is activated first came from X-ray-diffraction studies of living muscle. When a muscle contracts, its X-ray pattern changes dramatically. Most of the changes can be ascribed to the movement or disordering of the heads of the myosin molecules during contraction. Examining the patterns more closely, both Huxley and Lowy's group at the University of Aarhus noticed a subtler but distinct shift in the intensity of spacings associated not with myosin but with the thin filaments.

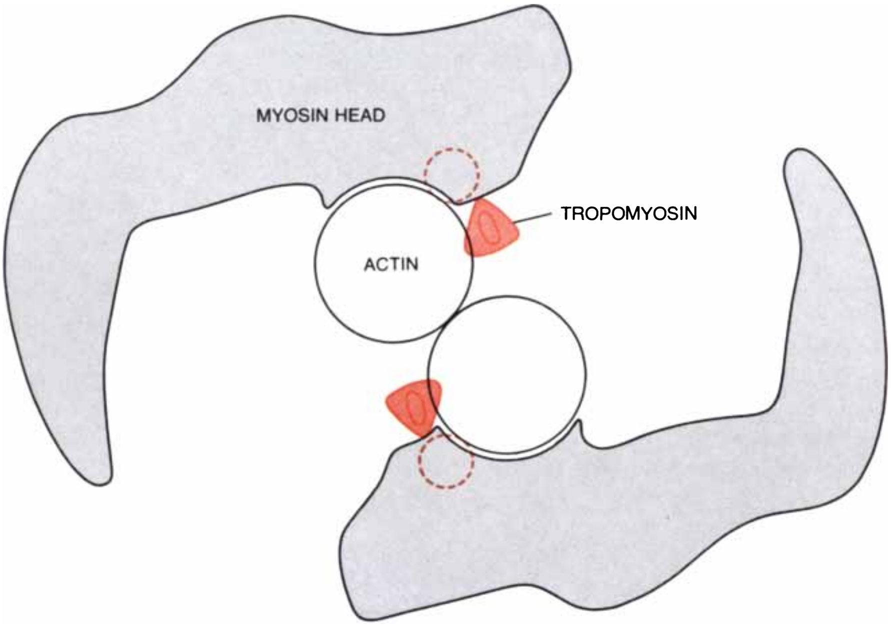
The X-ray patterns were in fact showing that the tropomyosin molecules are positioned differently in the groove of the actin helix when the muscle is at rest than they are when the muscle is contracting. This finding led to a simple steric, or three-dimensional, model for the action of the regulatory proteins, which was suggested independently in 1972 by Huxley and John Haselgrove and in 1973 by David A. D. Parry and John Squire, who were then at the University of Oxford. For a muscle to contract, as I have indicated, the myosin heads must interact with actin, splitting ATP. In the relaxed state (at low calcium levels, as Ebashi had shown) troponin somehow represses the interaction of myosin with actin, working through the tropomyosin molecule. At a critical calcium level that repression is removed, the proteins interact, ATP is split and the muscle contracts. The three-dimensional model specifies the nature of the derepression. The X-ray patterns can be interpreted to show that in the relaxed state the long tropomyosin molecule lies along the actin molecules near the outer edge of the groove [*see illustration on opposite page*]. When a critical number of calcium ions have become bound to troponin, the configuration of the troponin changes and tropomyosin moves deeper into the groove, exposing a specific site on the actin. Now the myosin heads can bind to actin and the muscle is switched on. In other words, the tropomyosin literally and physically blocks the actin site in the relaxed state, and then moves out of the way and uncovers the site to make contraction possible.

While this steric model of the regulation mechanism was being proposed and confirmed, the effort to specify the molecular structure and interactions of its elements was continuing. When we



PARACRYSTAL OF TROPOMYOSIN is made by precipitation with calcium or magnesium ions. An electron micrograph of a paracrystal, negatively stained with uranyl acetate (*top*), shows a pattern of light and dark regions bounded by thin white bands. The pattern is interpreted in the drawing (*middle*). The thin white bands mark the ends of molecules, perhaps indicating a small overlap. The symmetrical, nonpolar pattern derived from polar tropomyosin molecules means there must be sets of molecules (*black arrows*) bonded head to tail to form filaments that are oppositely directed in the paracrystal. The filaments are shown as being straight but are probably bent, since the repeat period here is 395 angstroms whereas the effective repeat length of the molecule in the filaments of crystals is about 410 angstroms. If troponin is added, it forms a bright new band (*bottom*). These micrographs and most others illustrating this article were made by Paul Norton in author's laboratory at Brandeis University and Marjorie Kasac of the Children's Cancer Research Foundation.

began examining tropomyosin crystals by means of X-ray diffraction in 1965, we found that the highly hydrated crystals were rather disordered and did not yield high-resolution patterns. Moreover, it appeared unlikely that even those limited X-ray diagrams could be deciphered without the application of rigorous crystallographic procedures. Electron microscopy provided a shortcut. In 1963 Huxley had published electron micrographs of fragmented crystals of tropomyosin made by the negative-staining technique, in which the protein molecules were outlined against a dense metallic background. The micrographs showed a beautiful kite-shaped mesh of cross-connected strands. The repeat distance along the strands was about 400 angstroms; the strands were wavy and in some places could be seen to be double, composed of two filaments. Caspar recognized that Huxley's image of the crystal lattice corresponded perfectly with our X-ray pattern of the tropomyosin crystal oriented in a particular direction. Indeed, the X-ray pattern was just what would be given by the crossed helical arrays of filaments that make up the mesh [*see illustrations on next page*].



STERIC MODEL of tropomyosin's action is based on X-ray diffraction and a technique, developed by Aaron Klug and David DeRosier, in which a three-dimensional image is obtained by combining information from a number of electron-micrographic views. This composite end-on reconstruction by H. E. Huxley shows the thin filament in the "on" position, in which the myosin heads bind to actin. In the "off" position the tropomyosin molecules (*broken colored circles*) block a specific site on the actin monomer. Under the influence of calcium, mediated by troponin (not shown), tropomyosin moves toward the groove, uncovering the site so that myosin can bind to it, split ATP and cause the muscle to contract.

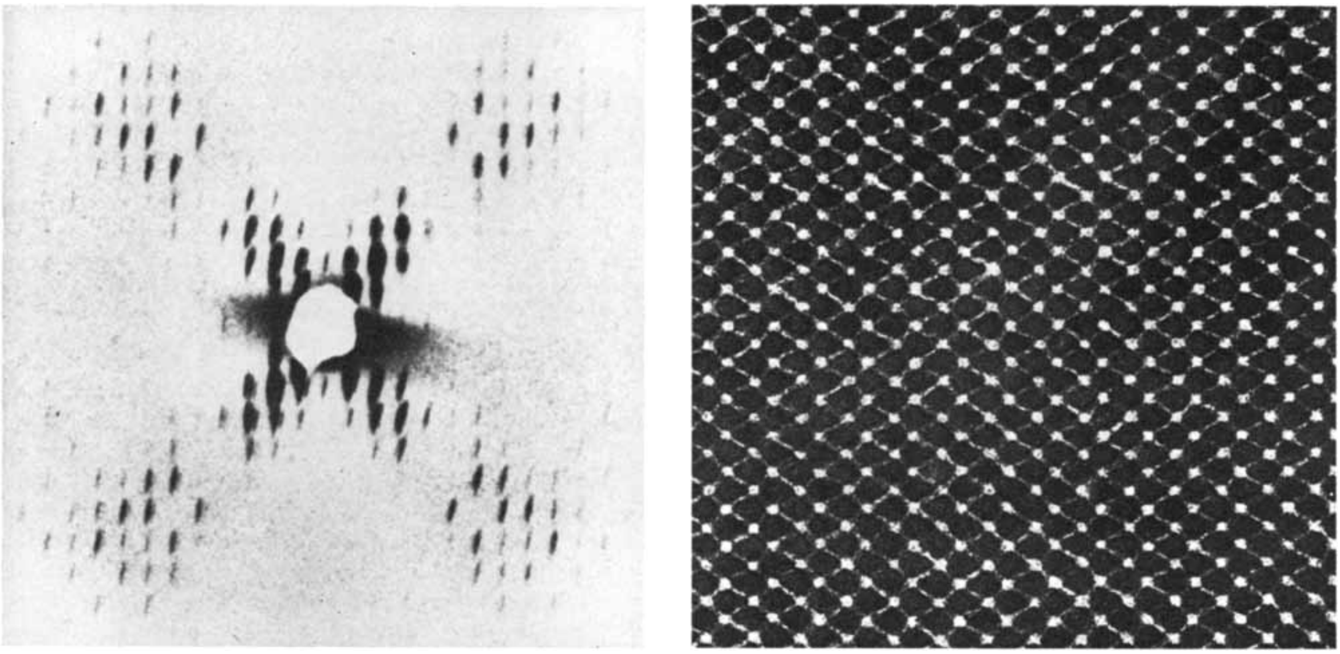
Now we could put together the information contained in the X-ray pattern and the electron micrograph; each approach clarified the other. The unit cell of the crystal (the smallest volume that when repeated gives rise to the entire crystal) has a body diagonal 402 angstroms long. The molecules run along the diagonal, and so 402 angstroms is the minimum length of the molecule. The X-ray patterns also showed us that the tropomyosin molecules are polar, that is, one end of the molecule is different from the other. The filaments are made up of molecules about 400 angstroms long bonded head to tail. The waviness of the strands is the result of cross-connections in the lattice, which produce periodic bending. (If the filaments were straight, the repeat length of a tropomyosin molecule would be about 410 angstroms.) Since each filament is made up of molecules that are themselves two-strand alpha-helical coiled coils, the regular perturbations caused by the cross-connections bend the filaments into a further supercoil: a coiled coiled coil.

A tropomyosin crystal, then, is made up of filaments cross-connected into a particular lattice. We find that under different conditions the tropomyosin filaments form various open nets or compact fibrous structures. All of them have a repeating unit of about 400 angstroms, close to the length of the molecule. Various nets have wavy or straight filaments depending on the particular interactions in the lattice; similarly, the fibrous aggregates, such as the paracrystal I have described, display a variety of band patterns, both polar and nonpolar, depending on the orientation and relative displacement of the polar filaments. The basic structure that builds all these varied forms is the polar tropomyosin filament made up of 400-angstrom molecules bonded end to end.

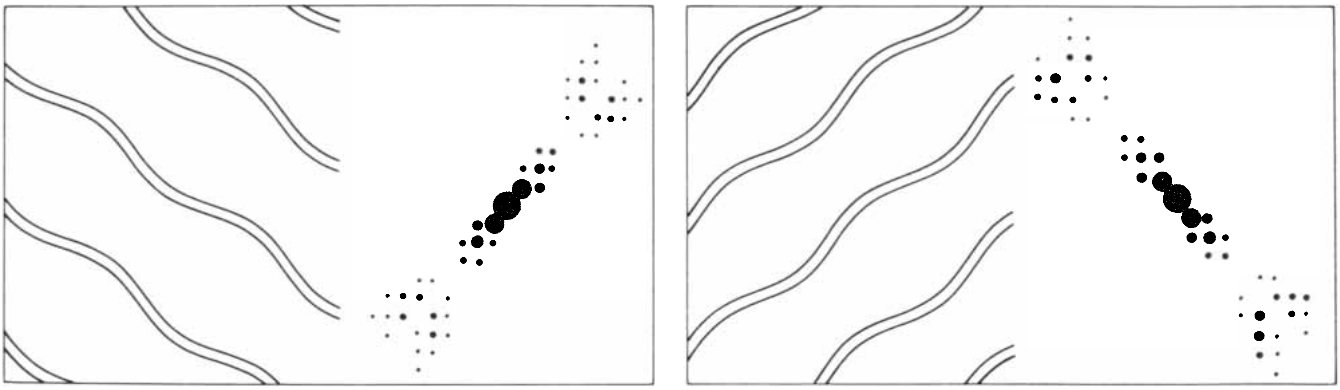
One might expect the tropomyosin crystal to be a static system, but the crystal structure actually changes in ways that parallel the movements of tropomyosin in muscle. For example, the entire troponin complex can be bound within the open mesh of the tropomyosin crystal. In a micrograph of a negatively stained crystal the bound troponin appears as a lightly staining node on the long arms of the mesh. Diffraction patterns suggest that the binding of a small amount of troponin in the lattice also changes the unit-cell dimensions. Yet whatever those dimensions are, the length of the body diagonal is constant. The cross-connected molecules can pivot and slip at crossover points, but the end-to-end connections between molecules are invariant; the molecules may be bent in various ways, but only certain linkages can vary.

The structure and dynamics of the crystals are closely related to the structure and movements of tropomyosin in muscle. In the thin filament the polar tropomyosin filaments are wound along each of the two grooves of the actin helix. They thereby assume the form of a coiled coiled coil; the bending of the filaments as they follow the helical grooves is comparable to their bending in the crystal lattice. X-ray diffraction of whole muscle yields a reflection at a spacing of 385 angstroms, which corresponds to the axial spacing of the troponin complex. Tropomyosin filaments with a 410-angstrom molecular repeat, wound along the actin groove at a radius of about 30 angstroms, would give an axial repeat of 395 angstroms; some additional kinking or bending of the tropomyosin could account for the shortening of the spacing to the 385 angstroms we observe. The important point is that the attachment of troponin to the supercoiled tropomyosin filament is specifically at one point on each tropomyosin molecule and corresponds almost exactly to the length of seven actin molecules. When the tropomyosin filament rolls toward the center of the groove under the influence of calcium (relayed by troponin), the 385-angstrom reflection due to the spacing of the troponin complexes remains constant. In other words, we can distinguish two kinds of linkage, as we can in the crystal lattice. The end-to-end bonding of the tropomyosin remains invariant; the short-range links to actin, on the other hand, which are analogous to the cross-connections between molecules in the crystal lattice, are variable. The movement of tropomyosin on the thin filament appears to be a reversible, sidewise rolling without axial displacement along the filament. The tropomyosin crystal is remarkable in that the connections and the movements of its fibrous molecules reflect the connections and the movements of those molecules in muscle.

Further information about the structure of tropomyosin is emerging from studies of amino acid sequences. Peter Cummins and Perry showed that vertebrate tropomyosins are heterogeneous, that is, they can be separated into two populations of slightly different chains of amino acids. Lawrence B. Smillie of the University of Alberta and his colleagues have determined the complete amino acid sequence of one of the chains, designated the alpha chain, comprising 284 amino acid subunits. Crick had long ago predicted that in the coiled coil about two of every seven amino acids would be found to be nonpolar (electrically uncharged), so that they would tend to point toward the inside and thus help to stabilize the structure. Smillie's analysis bore out the prediction: nonpolar amino acids are spaced regularly along the chain, separated by an average interval of 3.5 subunits. The two chains can therefore be meshed by an interlocking of the nonpolar subunits; a "knobs into holes" packing arrangement, which Crick foresaw, is quite likely. The next step is to construct a detailed model showing just how the chains are aligned with each other. Smillie and his co-workers originally suggested that the two chains might be offset by a small multiple of seven residues. More recent



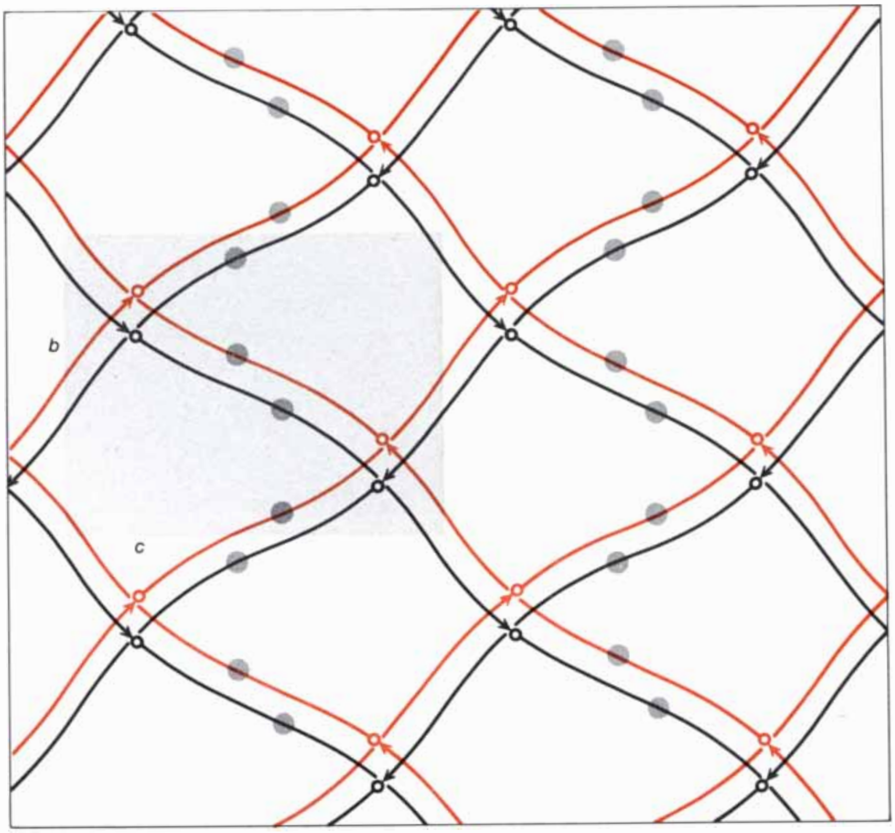
X-RAY PATTERN AND ELECTRON MICROGRAPH of tropomyosin crystals were interpreted together to decipher the structure and interactions of the molecule. The X-ray-diffraction pattern (*left*) could not be readily interpreted by itself but could be understood with the aid of an electron micrograph (*right*) of a negatively stained tropomyosin crystal. The micrograph is a projection in two dimensions of a three-dimensional lattice of cross-connected molecular strands enlarged 200,000 diameters. The strands are wavy (resembling a sine wave, the projection of a helix), so that their mesh forms kite-shaped regions. The repeat along a strand (a long arm plus a short arm) is about 400 angstroms. In parts of the micrograph one can see that strands are made up of two filaments.



X-RAY PATTERN is related to the crystal lattice as shown here. One spike of reflections is contributed by an array of wavy double filaments oriented in one direction (*left*). The other spike is contributed by an array oriented transversely (*right*). Diffraction patterns like those shown here were made by optical diffraction from such arrays and agree closely with the X-ray-diffraction pattern.

results favor a model in which the two chains are in register.

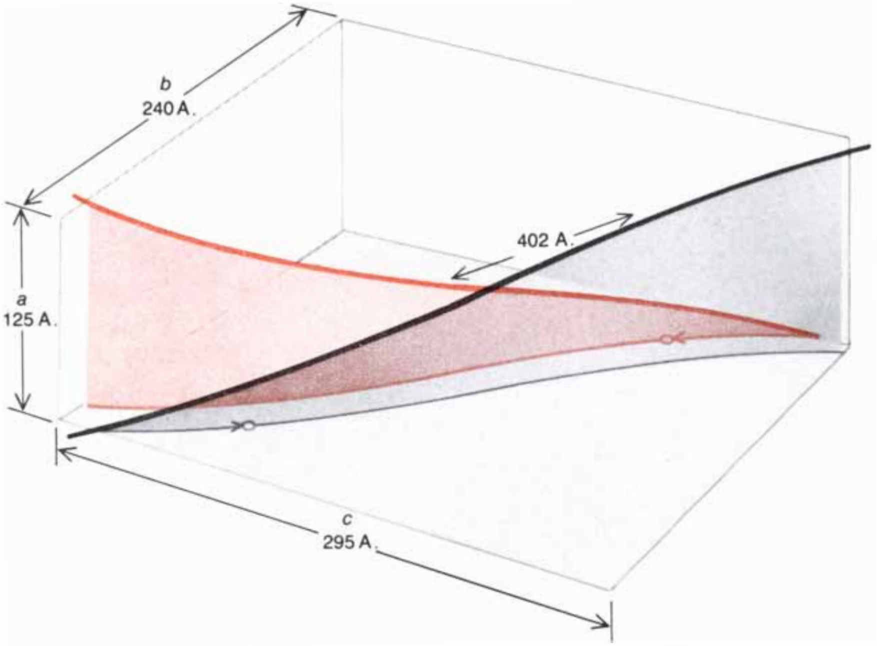
X-ray crystallography can provide further structural information and may help us to decide between the two arrangements. Crystals with a higher degree of order have now been grown in our laboratory by Cummins and Kien-yin Lee from the homogeneous tropomyosin (containing two alpha chains) of rabbit heart muscle. Diffraction from such crystals yields patterns extending to a higher resolution. Moreover, each alpha chain has one amino acid, cysteine, that carries a sulfhydryl group (SH), to which it is possible to attach a heavy-metal atom. Such a dense atom acts as a high-contrast marker that can be used to measure any offset between the two chains. With Eaton Lattman in our laboratory I am now trying to detect and measure the distance between mercury atoms attached to the tropomyosin molecule. Knowledge of this distance, together with the amino acid sequence, should make it possible to see the tropomyosin filaments in greater detail. The offset (if any) between the two chains will indicate how the molecules are bonded end to end in the tropomyosin filament. If there is an overlap at the junction between molecules, it could account for the difference between the length (423 angstroms) implied by the 284 amino acids in a supercoiled alpha helix and the effective repeat length (410 angstroms) of tropomyosin molecules in filaments. The position of the mercury markers will also make it possible to locate the ends of the molecules in the crystal lattice.



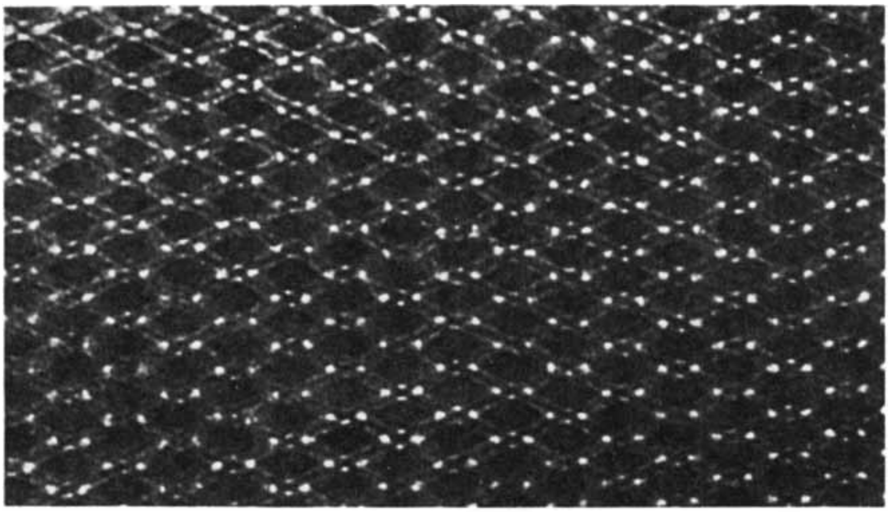
MOLECULAR CONFIGURATION in the tropomyosin crystal deduced from X-ray patterns is shown in two dimensions (the *bc* plane of the crystal, a view corresponding to that of the micrograph on the opposite page). Each filament is a polymer of tropomyosin molecules (*color and black*) bonded head to tail. The position of the ends of the molecules is not established but is here assumed to be near crossover points. The periodic bending of the molecular filaments (the coiling of the coiled coil) is caused by cross-connections in the lattice. If troponin is incorporated, it binds a third of the way from the COOH terminus and is seen in electron micrographs about midway on the long arm of the mesh (*gray disks*).

Knowledge of the sequence and the three-dimensional structure should also help to explain how tropomyosin is bonded to troponin and actin. The troponin-binding site is now known to be about a third of the way from the COOH terminus of the tropomyosin molecule. The tropomyosin, which spans almost exactly seven actins, may be attached in a similar way to each actin because there appears to be a fourteenfold repeat in the sequence along the tropomyosin molecule. The fourteenfold repeat has been noted in studies of the sequence by Parry, by Smillie and his colleagues and more recently by Andrew McLachlan and Murray Stewart of the University of Cambridge. Although there is no pattern of identical repeating regions along the tropomyosin molecule, there appear to be seven pairs of "quasi-equivalent" regions at which the tropomyosin may be linked to actin.

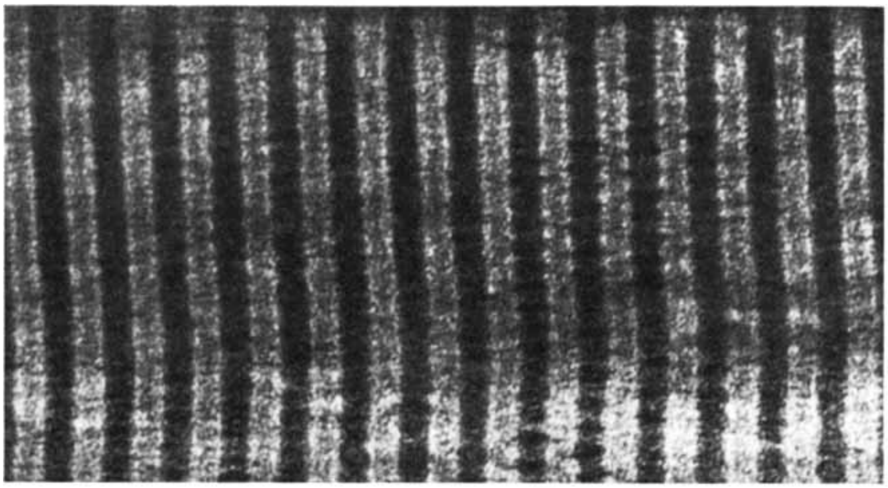
The linkage to troponin depends not



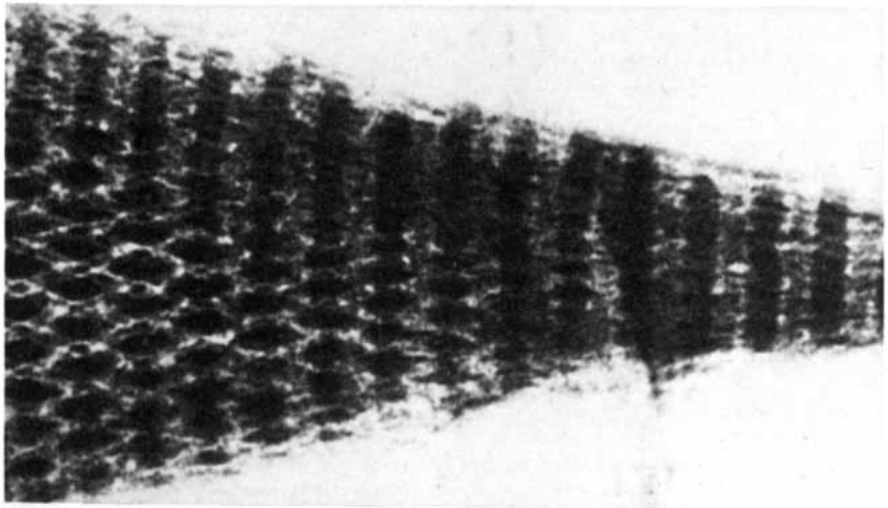
PART OF UNIT CELL OF TROPOMYOSIN CRYSTAL (the smallest volume that, when repeated, reproduces the entire crystal) is represented in three dimensions. Two helical filaments of tropomyosin molecules (*color and black*) run in opposite directions along body diagonals of the cell; their parallel sine-wave projections are shown schematically on the base. The dimensions of the cell were determined from X-ray data; calculation gives the 402-angstrom length for the body diagonal, which must be the minimum length of the molecule.



DOUBLE-DIAMOND MESH is produced when the *TnT* subunit of troponin is crystallized with tropomyosin. The period along the strands is still 400 angstroms but the crossovers are separated to produce a mesh of large and small diamonds. Tropomyosin molecules cross at or near position of *TnT* (*bright nodes*), which apparently generates new cross-connections.



DOUBLE-DIAMOND PARACRYSTAL formed when *TnT* and tropomyosin are precipitated with magnesium has a 400-angstrom repeat with a pair of bright bands: *TnT* subunits.



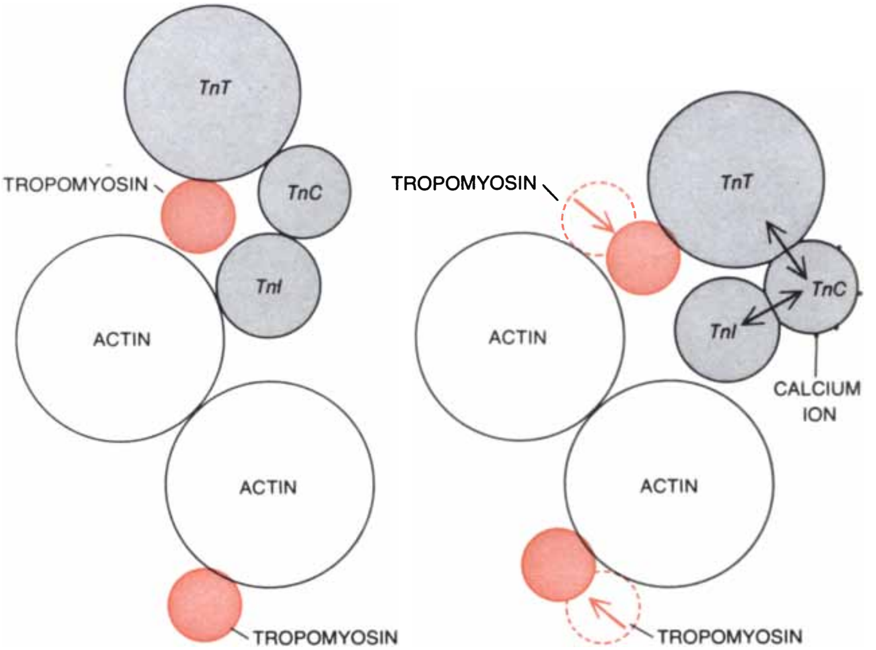
COLLAPSE OF DOUBLE-DIAMOND MESH to form a paracrystal was first observed in an electron micrograph made by Sugie Higashi-Fujime and Tatsuo Ooi in 1969. The extreme collapsibility of the double-diamond lattice may be due to the *TnT* swivels at the joints.

only on the structure of tropomyosin and troponin but also on how the complex changes its shape under the influence of calcium. The troponin complex consists of three subunits with markedly different properties. The largest subunit, designated *TnT*, binds strongly to tropomyosin. A smaller subunit, *TnI*, binds to actin and can inhibit the actin-myosin interaction. The smallest subunit, which binds calcium, is called *TnC*. The complex has not yet been crystallized and so its structure is quite unknown, but several different kinds of evidence make it possible to imagine how the three-dimensional model might work.

Sarkis S. Margossian and I found that the interaction of tropomyosin and *TnT* can be observed if a mixture of the two is crystallized under the same conditions as those of Bailey's original crystallization. A new kind of lattice appears. The spacing along the strands is still 400 angstroms, but the cross-connections are different, giving rise to a "double diamond" mesh [*see top illustration at left*]. The node at the acute vertex of each diamond is distinctly larger, in a negatively stained preparation, than the node at the obtuse vertex. We think that the larger node marks the location of the *TnT* component, which appears to cross-connect the tropomyosin filaments. If, instead of forming crystals, we precipitate the tropomyosin-*TnT* mixture with divalent positive ions such as magnesium (the conditions under which we first produced a banded fibrous paracrystal), the double-diamond paracrystal that is formed has a distinctive pattern, with two bright bands in the repeating unit [*see middle illustration at left*]. A remarkable micrograph made by Sugie Higashi-Fujime and Tatsuo Ooi in Japan recorded the transition between the crystalline and the paracrystalline form of tropomyosin-*TnT*. It shows the mesh collapsing to form the banded fiber [*see bottom illustration at left*].

If we add *TnC* to the tropomyosin-*TnT* mixture but keep the calcium concentration low, little *TnC* is bound and the double-diamond paracrystal is formed. Above a certain concentration of calcium, however, more *TnC* is bound, the double-diamond paracrystal is disrupted and a banded paracrystalline array is formed that is similar to the one that was formed by incorporation of the entire troponin complex. This experiment demonstrates a calcium-dependent linkage between *TnT* and *TnC*. Strong calcium-dependent binding has also been demonstrated between *TnI* and *TnC*. Moreover, Sarah Hitchcock, Huxley and Szent-Györgyi have shown that the binding of the *TnC-TnI* complex to actin associated with tropomyosin has a distinct inverse dependence on calcium: increasing the calcium level decreases the interaction of the troponin subunits and actin.

These various findings can be related by a simple model in which *TnI* attaches the troponin complex to actin [*see illustration at right*]. At low calcium levels *TnI* binds strongly to actin; tropomyosin is in the blocking position and prevents the attachment of the myosin head to actin. Saturation with calcium strengthens the linkage of *TnC* to the troponin subunits and weakens the *TnI*-actin linkage. *TnI* therefore may move away from the actin (or to a different position on the actin), allowing tropomyosin to move from the blocking position to one nearer the groove. This is the "on" position: the myosin heads bind to actin, ATP is split and the muscle contracts. The binding of calcium ions to *TnC* can be pictured as leading to a "tightening" of the bonds among the subunits of the troponin complex and a weakening of the interaction of the complex with actin; lowering the calcium level "loosens" the complex and makes the *TnI* subunit bind more strongly to actin. Tropomyosin amplifies the *TnI* blocking effect, extending it over seven actins.



SUBUNITS OF TRONONIN may function as suggested in this tentative model of the regulation of muscle contraction. The thin filament is seen schematically, end on. In the resting state (*left*) the *TnT* subunit binds to tropomyosin and the inhibitory subunit *TnI* binds to actin; the calcium-ion concentration is low and the links between troponin subunits are relatively loose. In the active state (*right*), above a critical calcium level, linkages between troponin subunits are tightened and the link between *TnI* and actin is weakened; tropomyosin moves deeper into the actin groove, exposing the site at which myosin can bind.

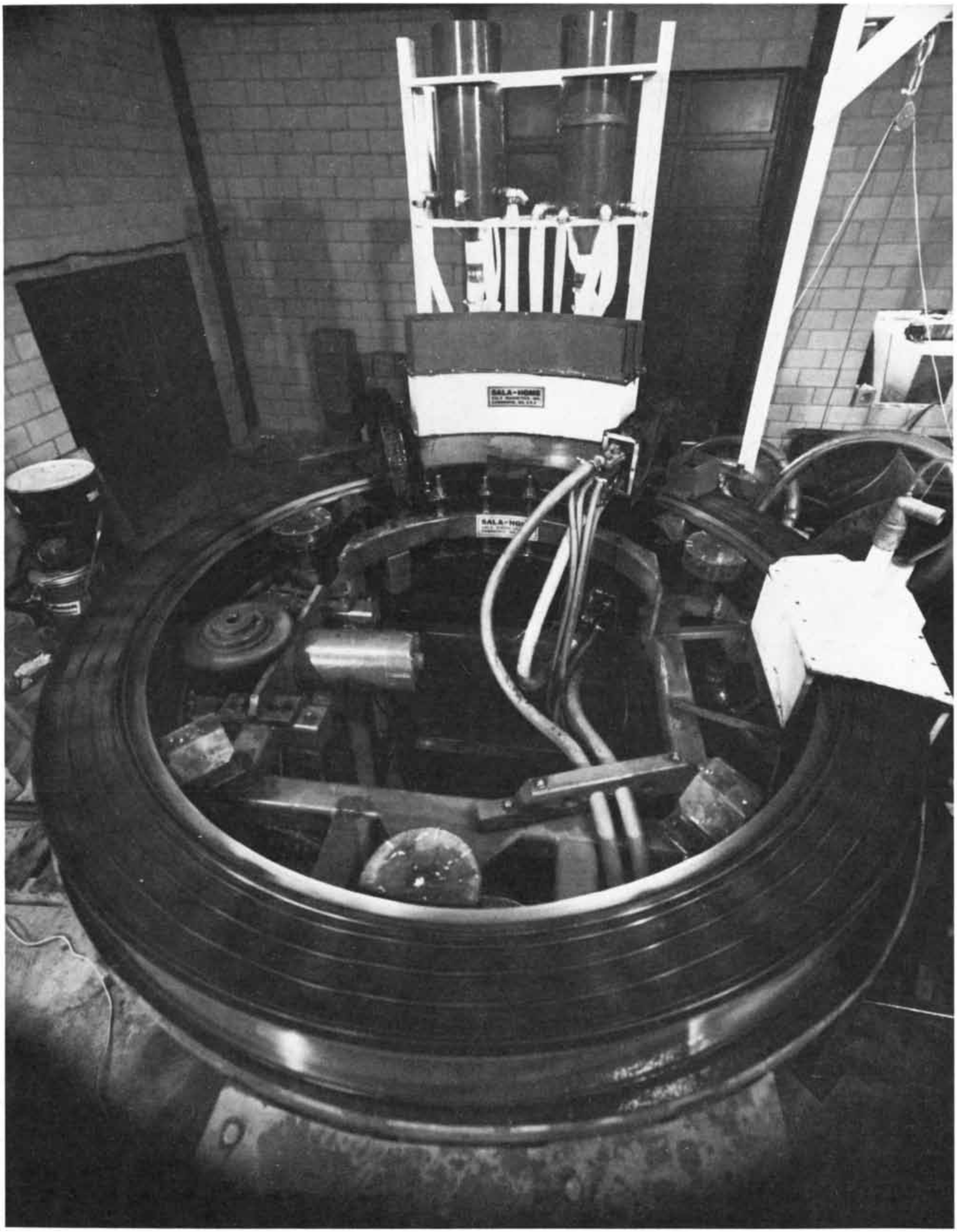
A model such as the one I have described is simply a synthesis of data; the structural states it invokes have to be looked for—if possible by direct demonstration in clear visual images. It has not yet been possible to obtain such images for the troponin complex by electron microscopy or X-ray diffraction. Some idea of how and where the calcium is bound is beginning to emerge, however, from studies of the amino acid sequences in two other calcium-binding proteins.

One of them is parvalbumin, a small protein found in fish muscle (but not in the contractile fibril itself). In 1970 Robert Kretsinger of the University of Virginia undertook an X-ray study of parvalbumin, hoping simply to learn something about the folding and cross-linking of its protein chains. When it was discovered that parvalbumin has a high affinity for calcium, Kretsinger set out to find its calcium-binding sites. He soon showed that there are two sites formed by two homologous stretches of the chain (regions of similar amino acid sequence); at each site the chain forms a loop around a calcium ion. In 1973, when John Collins and his co-workers at the Boston Biomedical Research Institute determined the sequence of the *TnC* of rabbit muscle, they recognized four homologous regions corresponding to the calcium-binding sites. Further information should soon be forthcoming: quite recently a group at the University of Oxford has crystallized rabbit *TnC*, and X-ray studies are now in progress that should reveal the structure of the molecule and perhaps also its conformational changes under the influence of calcium.

Attempts to track down the various activities of calcium ions in muscle have led to still greater surprises. Calcium is now known also to regulate the key muscle protein, myosin. For some time after Ebashi's discovery of troponin it was assumed that troponin mediated calcium control in all types of muscle, but Szent-Györgyi and his colleagues found in 1970 that there is no troponin in the muscles of mollusks, although tropomyosin is present; calcium triggers contraction by acting on the myosin directly. (They have since found that most animal species have both myosin-linked and actin-linked regulation. Vertebrates and mollusks are apparently unusual in having only a single system, the troponin-actin regulatory system in the former and the myosin system in the latter.)

The mechanism of myosin-linked regulation is still unknown, but the action of calcium is associated with the presence of a specific light chain (or small subunit) on the head of the myosin molecule. Light chains are present on the myosins of all muscles, and a remarkable homology has recently been shown by the fact that one of the several light chains in rabbit myosin can substitute effectively for the specific regulatory light chain of molluscan myosin; another has been shown to have similarities of amino acid sequence to both parvalbumin and *TnC*. These findings suggest that the ability to bind calcium, established by a particular type of amino acid sequence, has been acquired by several different proteins, perhaps through a process of gene duplication. Some more subtle distinctions among these several proteins determine which of them exercises the predominant role in the regulation of contraction in a given type of muscle.

These homologies underlying the various involvements of calcium in muscle suggest how great diversity of function may arise from small differences in the pattern of connections between similar polypeptide chains. The tropomyosin molecule illustrates too how a relatively simple structure is so designed that it can participate in diverse interactions. Although much remains to be learned about the regulatory proteins, we are beginning to see the structural basis of some critical linkages between components and to understand how the structure of a molecular assembly is adapted to a complex physiological function.



HIGH-GRADIENT SEPARATOR at Sala Magnetics, Inc., in Cambridge, Mass., has a six-foot ring divided into compartments filled with steel wool. The ring is rotated through the magnet structure at the top, where a finely ground slurry of iron ore is introduced. The magnetic particles stick to the magnetized steel wool and the nonmagnetic particles are flushed out. After the ring has emerged from the magnet structure, it passes through the structure at the right, where the magnetic particles are flushed out and formed into pellets that serve as a blast-furnace feed. This machine is the Mark II carousel separator, designed by Peter Marston and his associates. It can handle up to five tons of feedstock per hour. In this photograph the ring of separator is blurred because it is in operation.

HIGH-GRADIENT MAGNETIC SEPARATION

A recent advance in the generation of strong magnetic fields opens the way to removing very weakly magnetic particles from mixtures. One novel application of the new process is purifying wastewater

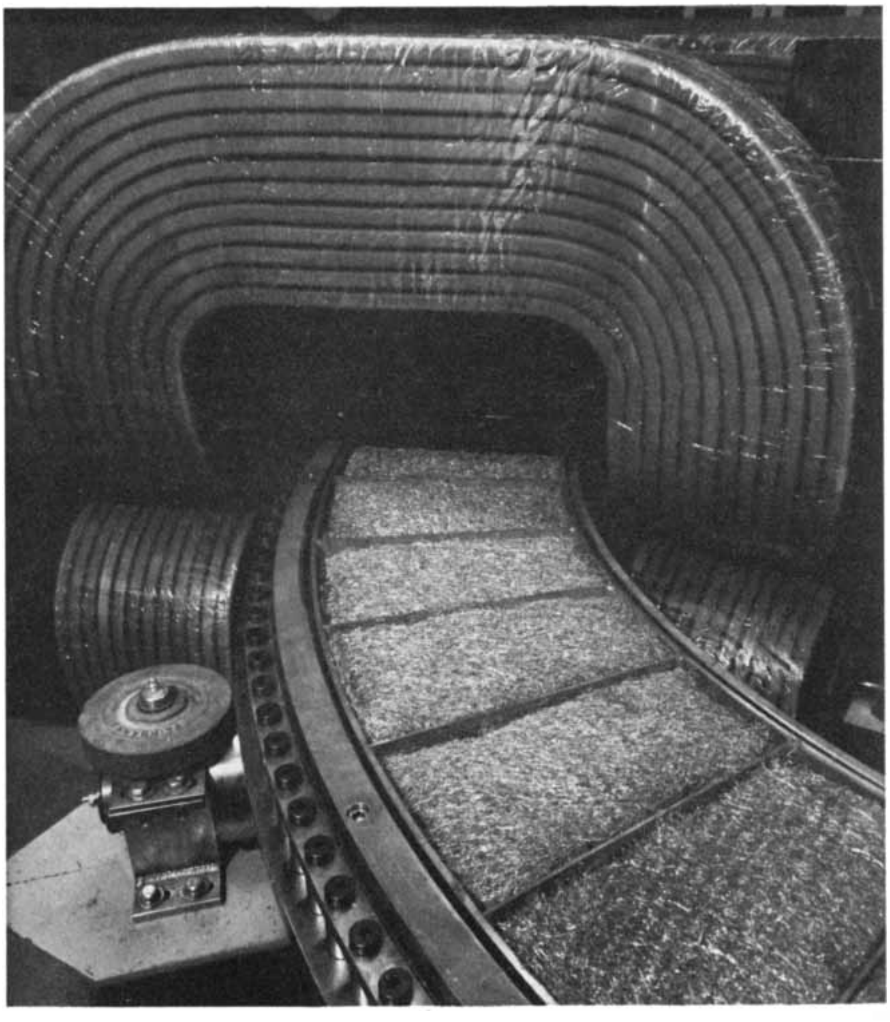
by Henry Kolm, John Oberteuffer and David Kelland

A major activity of modern industry is separating mixtures of fine particles; the desired particles are kept and the unwanted ones are discarded. Many different techniques of separation, including filtration, flotation, precipitation and magnetic separation, have proved economically practical. A recent advance in the technique of magnetic separation may prove to be revolutionary. It opens the way to dealing economically with a broad range of separation problems that have appeared unsolvable, involving fine particles that are only weakly magnetic and even some particles that are not magnetic at all. The new technique is known as high-gradient magnetic separation.

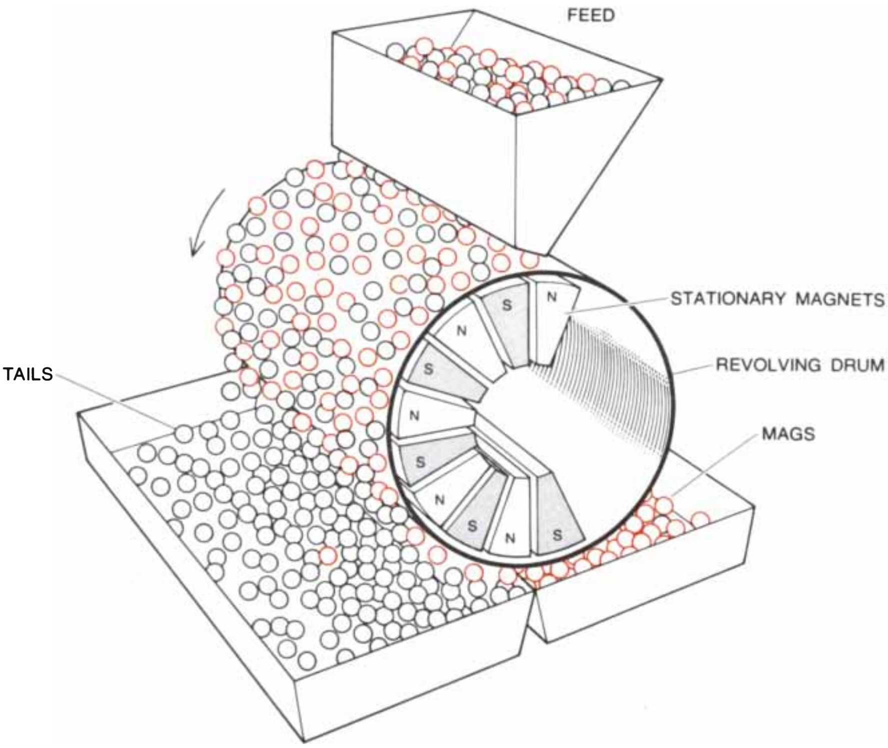
Reference to the periodic table of the elements indicates how far the high-gradient technique has advanced the art of magnetic separation. Until now magnetic separation has been confined to manipulating mixtures that contained one or more of three strongly magnetic elements: iron, nickel and cobalt. The high-gradient technique is potentially applicable to any one of a far greater number of mixtures: at a minimum those that include one or more of 56 weakly magnetic elements. Among other new possibilities is access to more than half of the world's iron resources, which consist of a weakly magnetic oxide of iron. Perhaps of even greater immediate significance is the possibility, after the appropriate "seeding" with magnetic particles, of manipulating substances that are entirely nonmagnetic, such as various pollutants in water. In brief, high-gradient magnetic separation promises advances not only in the mineral and chemical industries but also in environmental protection and water purification.

The earth is full with mixtures of finely divided matter. On a geological scale the margins of continents are constantly being enlarged by the fine solids that rivers carry to the sea. On a lesser scale a large fraction of the mineral industry is concerned with the separation of finely divided matter; indeed, the vast bulk of the earth's mineral resources remains unexploited because the materials are too finely divided to be separated economically. The chemical industry also utilizes many processes in which mixtures of finely divided solids must be purified by removing unwanted components from powders or slurries.

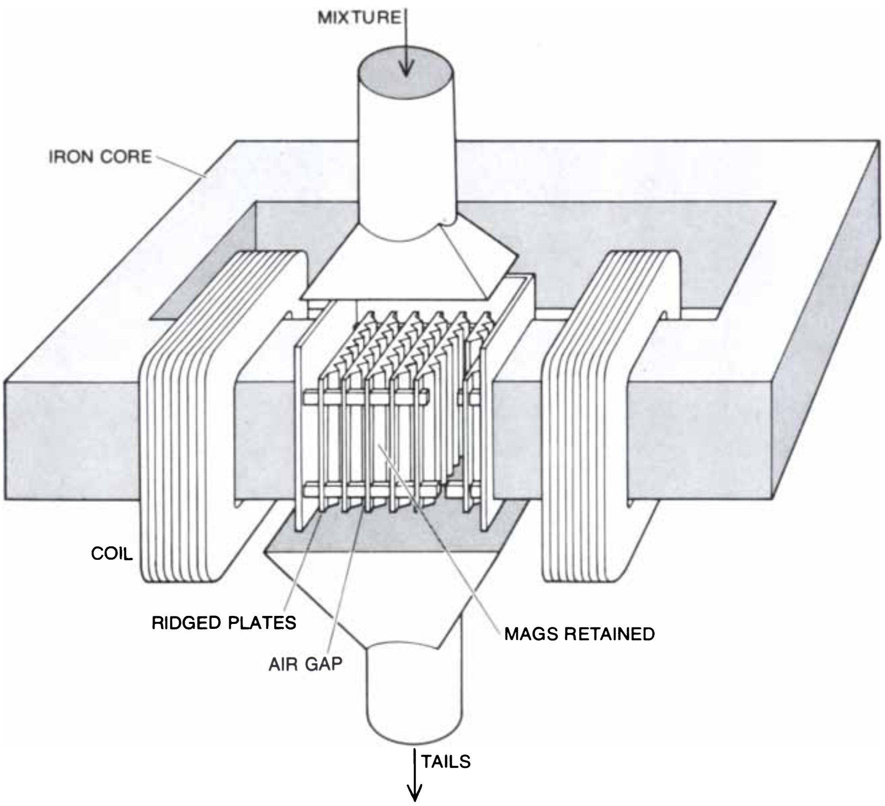
When the particles involved in these



COMPARTMENTS OF STEEL WOOL are shown at the point where they enter the magnet of the Mark III carousel separator, also designed by Marston and his associates. Surrounding ring are high-performance copper coils of magnet, which are hollow for water cooling.



CONVENTIONAL SEPARATION of magnetic and nonmagnetic particles in a mixture is exemplified by the drum separator, a device that can be applied only to strongly magnetic materials. A stationary half cylinder of magnets is fixed inside a revolving metal drum. As the mixture strikes the drum, the magnetic particles ("*Mags*") adhere and the nonmagnetic particles ("*Tails*") fall off. In this schematic drawing size of particles is greatly exaggerated.



HIGH-GRADIENT SEPARATION is applicable to mixtures containing weakly magnetic particles. Shown here is the first practical separator of this kind: George Jones's salient-pole-plate separator, designed early in the 1960's. The ridges on each plate in the stack of plates within the air gap act effectively to trap fine particles of weakly magnetic materials in a mixture. Jones machine is used to divide low-grade iron ores into hematite and waste.

processes are very small (with a diameter of a tenth of a micrometer or less), they fall within the domain of the colloids. A colloid is a state of matter quite distinct from the solid, liquid and gaseous states. Particles of colloidal dimensions will not settle when they are suspended in a liquid. The electrostatic surface charge on each individual particle keeps the particles from coalescing into larger aggregates. At the same time the thermal energy of each small particle more than counterbalances the particle's gravitational potential energy, or tendency to settle.

The reverse of the colloidal state is the mechanism that builds river deltas and makes continental margins grow. Contact with the ions in seawater neutralizes the surface charge of the suspended particles, promoting the process of aggregation. Once enlarged by aggregation the particles yield to gravity and sink. In dealing with matter consisting of particles of a size lying within or near the boundaries of the colloidal domain, high-gradient magnetic separation has advantages over other separation techniques.

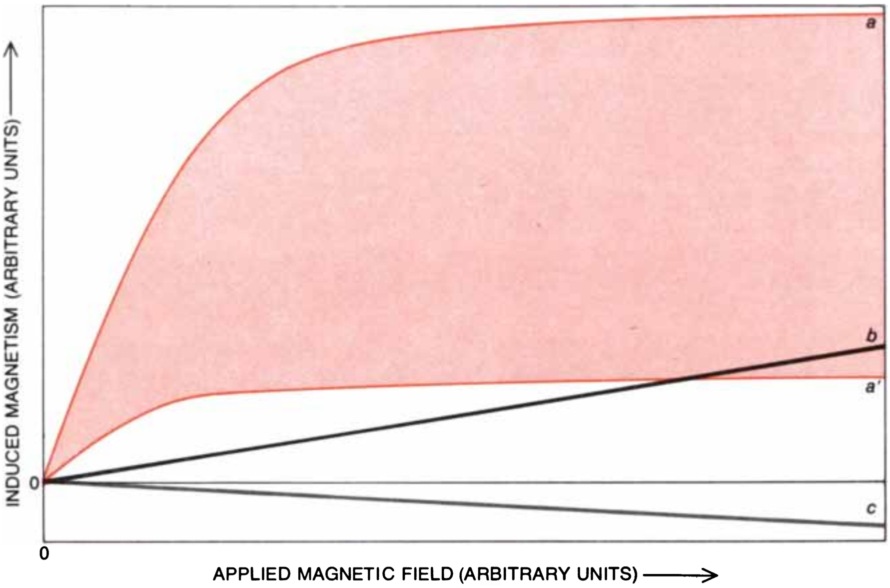
The physical phenomena governing magnetic separation fall under two headings. How do various substances behave when they are exposed to a magnetic field? How are the magnetic forces exerted? On the first heading, consider a graph that plots the response of various classes of substances to increasingly strong magnetization. Such a graph displays three different kinds of magnetic behavior [*see top illustration on opposite page*]. Strongly magnetic materials, members of the ferromagnetic group, are easily magnetized by a relatively weak magnetic field, and so the slope of their magnetization curve is steep at the beginning. (That is why a bar magnet, which does not have a very strong magnetic field, can attract ferromagnetic materials.) As the strength of the magnetic field increases, all the individual domains—regions with paired north and south magnetic poles—in a ferromagnetic material become aligned; magnetization "saturates" the material. Thereafter the slope of the curve remains relatively flat regardless of any further increase in the strength of the magnetic field. The saturation level, that is, the field strength beyond which no further magnetization takes place, depends on the iron content of the material. For example, pure iron is saturated at a magnetization of some 220 electromagnetic units per cubic centimeter in an applied field of several hundred gauss.

Weakly magnetic materials, members of the paramagnetic group, are far less susceptible to an applied magnetic field than ferromagnetic materials. At low field strengths their magnetization curve on the graph remains well below the ferromagnetic curve and its slope is much shallower. A paramagnetic material, however, rarely becomes saturated, and so its degree of magnetization continues to increase as the applied field gets stronger. This means that even though a bar magnet will not attract a paramagnetic material, such materials may become more highly magnetized in a sufficiently strong field than dilute ferromagnetic materials.

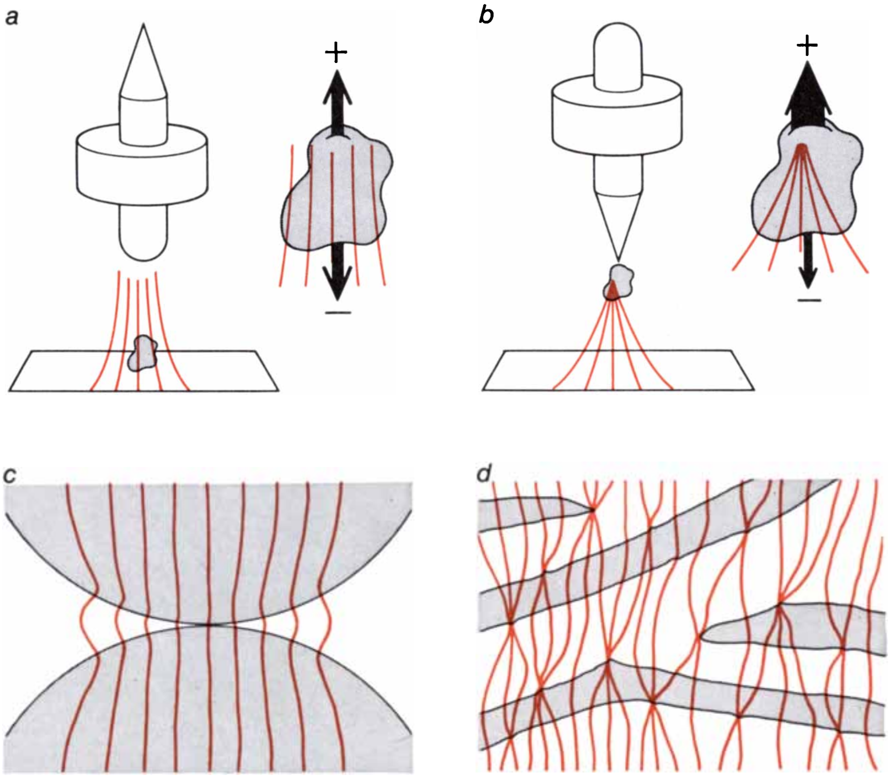
A third type of behavior in a magnetic field is displayed by materials that become magnetized in a direction opposite to that of the applied field. These are diamagnetic materials, and their curve on the graph is shallow and negative. For the purposes of this discussion diamagnetism is a small effect with no practical importance.

On the second heading—how magnetic forces are exerted on materials in an applied field—it is useful to think of each magnetized particle as acting temporarily as if it were itself a small bar magnet, with a north pole at one end and a south pole at the other. In magnetically "hard" materials the parallel alignment of the dipoles is in fact not temporary but permanent, and a magnet made of such a material is called a permanent magnet. The alignment in magnetically "soft" materials is impermanent; it is induced only while a magnetic field is applied, and it becomes random when the field is absent.

When a uniform magnetic field is applied to a magnetized particle, the forces acting on the two poles of the particle will be equal and opposite. The forces therefore cancel each other, and the resulting net force is zero. Only if the applied field differs in intensity at the two extremities of the particle will a net magnetic force act on the particle. This is to say that the applied field must have a gradient, a spatial variation that is appreciable in terms of the dimensions of the magnetized particle. Anyone who has tried to remove iron filings from a horseshoe magnet is familiar with this effect. The lines of force extending from the poles of the magnet diverge in such a way that the density of the magnetic flux (the intensity of the field) increases with nearness to either pole and the maximum spatial variation in the flux is at the sharp edges of the pole. It is the higher gradient of the magnetic field at the sharp edges that makes the iron



THREE RESPONSES are evident when the magnetization of a material is plotted as a function of the applied magnetic field. Ferromagnetic materials (*a*) show an immediate steep curve; some respond more strongly than others (*colored area*) but all eventually "saturate," as the flattening curves show. Paramagnetic materials (*b*) show a much shallower response, but they rarely saturate; in a strong field their magnetization can exceed that of a weakly ferromagnetic material. Diamagnetic materials (*c*) are of no industrial importance. On exposure to a field such a material shows a slight but opposite magnetization.



MAGNETIC-FIELD GRADIENT is needed if an appreciable force is to be exerted on a fine particle. Flux lines (*color*) around the blunt pole of a magnet (*a*) are essentially uniform with respect to the particle below the pole; the magnetic forces acting on the two poles of the particle (*enlarged at right*), being equal and opposite, cancel out. Flux lines around the sharp pole (*b*) are divergent with respect to the particle; the resulting high gradient exerts a strong force on the particle (*enlarged at right*). Stacked iron balls (*c*) represent one extreme in making a high-gradient matrix. They conduct the flux well (*color*), but their field is almost uniform. Steel wool (*d*) represents the other extreme. As the pinched flux lines show, gradients are high, but steel wool, being mostly void, is very hard to magnetize.

filings preferentially collect there. In short, the net force exerted on a magnetized particle by a magnetic field is proportional to three quantities: the intensity of the magnetization the field has induced in the particle, the volume of the particle and the gradient of the field, that is, the difference between the intensity of the field at one end of the particle and the intensity at the other.

With these basic facts in mind many experimenters have attempted to apply improved techniques of magnetic separation to two kinds of mixture that are particularly hard to deal with. These are mixtures containing particles that are strongly magnetic but very small and mixtures containing weakly magnetic particles of any size. The improved separators were generally built around a scaled-up version of a horseshoe magnet: an iron yoke, or frame, provided a strong magnetic flux across the air gap between the two poles of a permanent magnet. The passage of an electric current through a pair of copper coils surrounding the yoke further increased the strength of the magnet. The mixture that was to be separated into its magnetic and nonmagnetic components was fed through a filter placed in the air gap. The principle was simple: the filter would complete the magnetic circuit and, being strongly magnetized itself, would selectively entrap any magnetic particles in the mixture while letting the nonmagnetic particles flow through. The practical problem the experimenters faced was that the two functions the filter structure was supposed to serve are incompatible.

On the one hand, a filter structure must be a good conductor of magnetic flux; this is necessary not only to close the gap in the magnetic circuit but also to keep the flux lines from straying out of the gap. On the other hand, in order to provide both an adequate surface area for the entrapment of magnetic particles and a high-gradient local magnetic field, the conducting filter material must present a combination of many sharp edges and abundant void spaces.

In an attempt to overcome these incompatibilities experimenters made filter matrixes out of stacked iron balls, iron mesh, steel wool and even carpet tacks. The stacked balls represent one extreme: they furnish a good flux path but a low field gradient. Steel wool represents the opposite extreme: when it is magnetized to saturation, it has a large number of points with high field gradients and it has ample surface area, but it is very difficult to magnetize to saturation. The reason steel wool resists magnetization is that however tightly it is compressed it remains a spongy material that is 90 percent empty space. It is only slightly magnetized when inserted into the air gap of a conventional magnetic circuit. If the experimenter turns to alternative ways of applying a more powerful magnetic field to steel wool, he encounters a second difficulty that arises from the geometry of the stuff. The steel ingot from which steel wool is made would be relatively easy to magnetize because its long axis can be made parallel to the direction of the applied field, but in the steel wool the long axes of a large number of the filaments are inevitably perpendicular to the direction of the applied field. The applied field required to magnetize the randomly oriented mass of filaments is more than 10 times stronger than the field that will saturate the ingot from which the filaments were made.

Two experimenters have attempted to solve the filter-matrix dilemma. As early as the 1940's the inventor Samuel Frantz recognized both the advantages of steel wool and similar matrixes and the difficulty of magnetizing them in the air gap of a conventional magnetic circuit. Frantz began by eliminating the magnet's iron yoke and inserting a canister, filled with the matrix of his choice, in the hole through the center of the coil previously occupied by one leg of the yoke. The coil now operated as an electromagnet of the solenoid type.

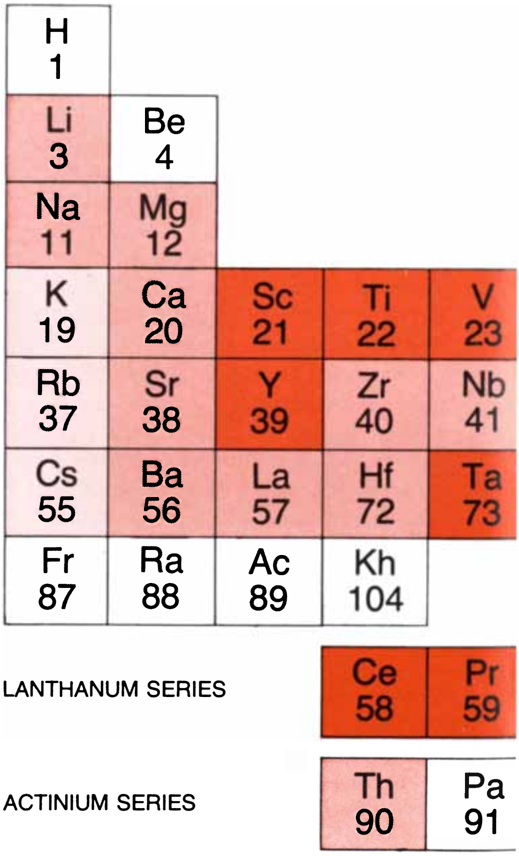
Now, if a solenoid is to generate a reasonably uniform field, it must be longer than it is wide. Unfortunately for Frantz, this is an unfavorable shape for separation purposes. If the matrix is to work efficiently, it must be shorter than it is wide.

Frantz also found that the strength of his magnetic field was limited almost entirely to what the solenoid itself could provide. Adding iron poles to the ends of the coil did not strengthen the field appreciably, nor did surrounding it with an iron sleeve. The Frantz Ferrofilter, incorporating a relatively coarse matrix, is an effective separator of some strongly magnetic particles. It could not, however, economically achieve the field intensities needed to magnetically saturate steel wool or to magnetize weakly magnetic materials. When it was equipped with a steel-wool matrix, it was suitable only for laboratory-scale work.

In the early 1960's George Jones, a British engineer, worked out a compromise between the two extremes of stacked balls and steel wool. Jones preserved the yoke-and-electromagnet structure and completed the magnetic circuit by placing a different kind of matrix in the air gap of the magnet. His matrix consisted of a stack of steel plates, flat on one face but with many parallel sharp ridges on the other. The plates were held out of contact with one another by means of nonmagnetic spacers at each corner. The use of spacers that differed slightly in thickness enabled Jones to make fine adjustments in the width of the air gaps between adjacent plates.

When the stacked plates are magnetized, the field gradients in the vicinity of the sharp ridges are high because the flux lines passing through each ridge diverge widely toward the flat back of the adjacent plate. By adjusting the air gap between the plates, Jones was able to balance the degree of gradient against the degree of flux conduction to achieve optimum conditions for each mixture he undertook to separate.

The Jones separator leaves something to be desired. For example, a dispropor-



PERIODIC TABLE of the elements is marked to indicate which elements are ferromagnetic (*gray*) and which have paramagnetic properties (*shades of color*). Only three elements are ferromagnetic: iron, cobalt and nickel. Compounds that include tionately large volume of steel is needed to magnetize a relatively small matrix. Furthermore, the geometry of the ridged plates is such that they can present field gradients only about 1 percent as high as saturated steel wool does. The Jones machines were nonetheless the first high-gradient separators to achieve commercial success. They are effective separators of paramagnetic materials as long as the particles in the mixture are not too small.

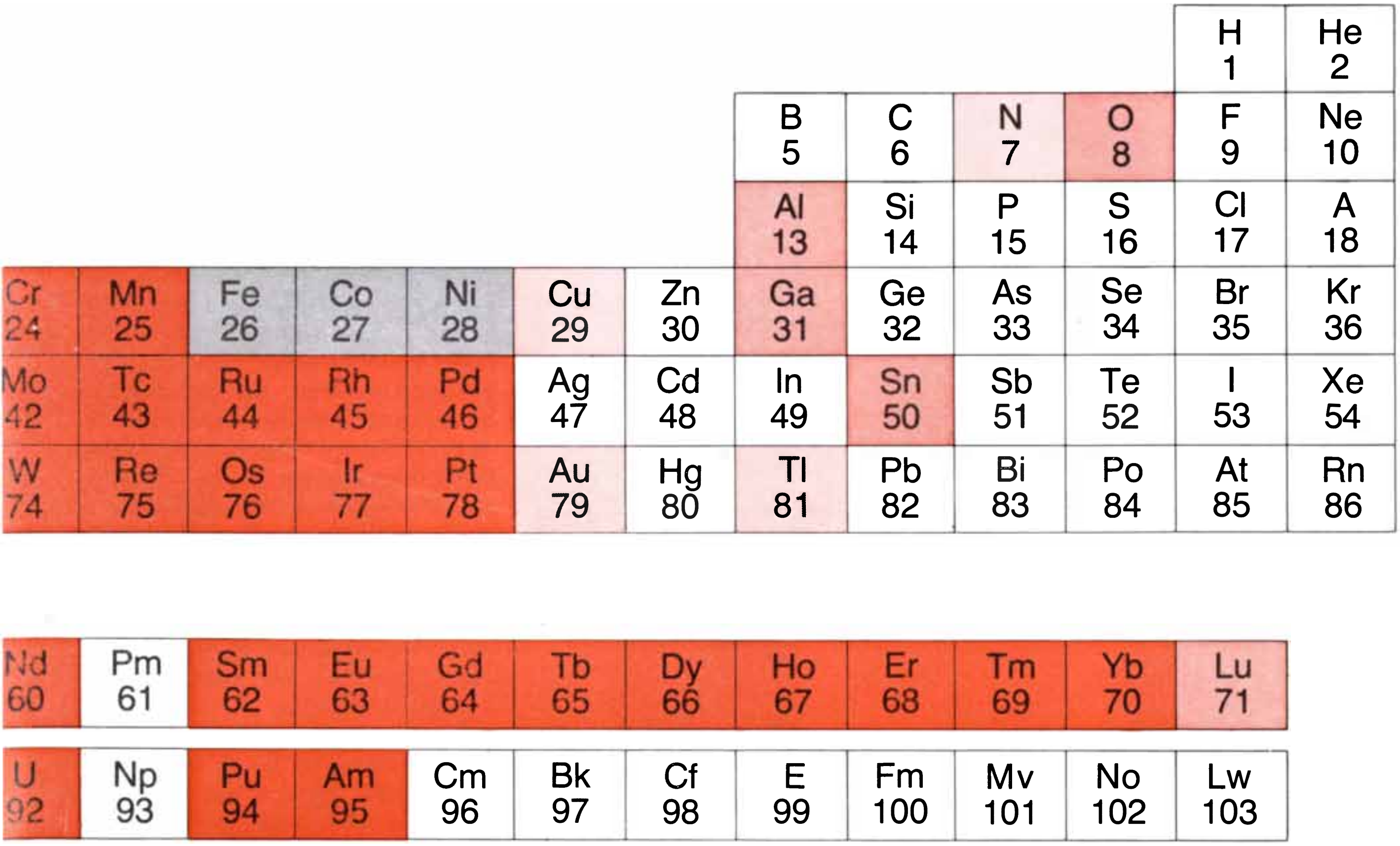
In the mid-1960's we became involved in these matters when our laboratory, the Francis Bitter National Magnet Laboratory at the Massachusetts Institute of Technology, was asked to assist in solving a practical problem in magnetic separation. Today the available supply of kaolin, the white clay used as a coating for paper, usually contains certain micrometer-size impurities. The principal impurity, titanium dioxide, is accompanied by particles of iron oxide that tend to discolor the finished product. As it happens, kaolin is very weakly paramagnetic. The impurities are more strongly paramagnetic, although they are only a ten-thousandth as magnetic as iron. Thus it was possible to separate them from the kaolin magnetically, although at that time no device existed that could do so on an industrial scale.

Some years earlier we had investigated very intense magnetic fields and had become aware of the very high gradients that could be attained with magnetically saturated steel wool and similar matrixes. Approaching the kaolin problem with this background, we soon found that saturated steel wool would almost totally remove the impurity particles. There remained the practical problem of building a magnet that could saturate a large volume of steel wool. The magnet also had to meet two other specifications: its cost of construction had to be acceptably low and so did its consumption of power.

All three specifications were met by Peter Marston and his associates, a group in Cambridge, Mass., specializing in the design and construction of complex magnets for both scientific and industrial purposes. (The Marston group is now associated with a Swedish manufacturer of mineral-processing equipment, Sala.) Like Frantz, Marston based his design on a solenoid surrounding an empty space where a matrix could be placed. Unlike Frantz's solenoid, however, Marston's incorporated an iron pole structure and was wider than it was long. The solenoid coil was also surrounded by an iron shell that facilitated the return of flux.

The practical impact of Marston's innovation has been considerable: his magnets can generate uniform fields of up to 20,000 gauss in large empty spaces. The throughput rate of some of the high-gradient separators incorporating his magnets is 60 tons of kaolin clay per hour. Moreover, compared with conventional separators, Marston's iron-bound solenoids require only a tenth of the capital investment and consume only a tenth of the power.

Kaolin purification is an example of separatory processes where a small un-



one or more of the three elements may be strongly or weakly ferromagnetic or may only be paramagnetic. Fifty-five elements have paramagnetic properties. Of these, 32 (*dark color*) form compounds that are paramagnetic. Another 16 elements (*lighter color*) are paramagnetic in pure form, but the compounds that incorporate them are diamagnetic. The remaining seven elements (*lightest color*) become paramagnetic when one or more are present in a compound, although two of them, nitrogen (*N*) and copper (*Cu*), are slightly diamagnetic in pure form. Of the 46 remaining elements, all but 16 have diamagnetic properties. Since the paramagnetic compounds greatly outnumber the ferromagnetic compounds, high-gradient magnetic separation has many potential applications.

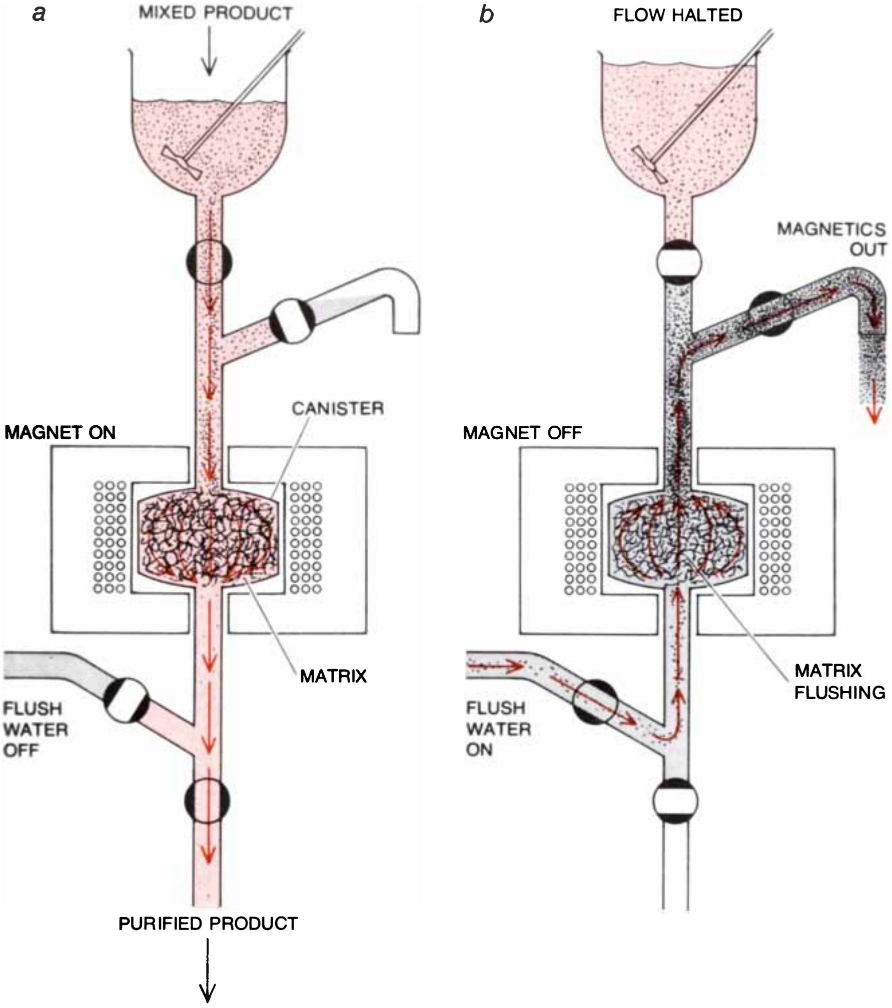
wanted magnetic fraction is removed from the mixture fed into the separator. Other examples of such processes are the purification of certain pigments and glass sands. There are numerous potential applications of high-gradient separation where the small magnetic fraction is the desired one. A case in point is the material discarded in the extraction of molybdenum and tungsten by nonmagnetic methods such as the removal of sulfides from the ores by flotation. The tailings that are discarded after flotation still contain appreciable amounts of these metals in the form of oxides that are paramagnetic. It is estimated that magnetic methods could extract from the tailings an average of about a cup of molybdenum or tungsten oxide per barrel of material processed.

There is, of course, a substantial economic difference between magnetic purification and magnetic recovery. With kaolin the investment in extraction yields barrels of the final product in the same length of time it takes to get a few cups of molybdenum or tungsten oxide. Many applications of oxide extraction so far lack the economic incentive that accompanies shortages and will have to await the time when today's tailings have become tomorrow's ores.

The most important long-term application of high-gradient separation to minerals may prove to be the beneficiation of low-grade iron ores, an area where conventional magnetic separation is already playing a leading role. The reason is that reserves of low-grade taconite ore include far more of the weakly magnetic iron mineral hematite than of the strongly magnetic mineral magnetite. For example, in the Mesabi Range of Minnesota, the nation's chief source of iron, the ores with an iron content of 65 percent or better (the only ores that can go directly into the blast furnace) were exhausted more than a decade ago. The place of the high-grade ores was then taken by taconite ores containing iron in the form of finely disseminated magnetite. The ores are pulverized and subjected to conventional magnetic separation. The concentrate is then formed into hard pellets, making it suitable for feeding into the blast furnace. The iron content of much of the Mesabi taconite and of most of the world's low-grade ores, however, is in the form of hematite rather than magnetite. High-gradient magnetic separation ought to do for hematitic ores what conventional magnetic separation and pelletization have done for magnetite. The new process may even do so economically before the world's supply of magnetite is exhausted; its feasibility has been demonstrated in pilot-plant tests supported by the division of the National Science Foundation concerned with research applied to national needs (RANN).

Another potential application of high-gradient separation is in the desulfurization of coal. Much of the sulfur in coal is in the form of the mineral pyrite. Pure pyrite is very weakly magnetic, but it usually contains a certain amount of pyrrhotite, a rather more magnetic sulfide of iron. We have found that when pulverized coal (the form in which coal is fed to most steam-turbine electric-power plants) is passed through a high-gradient apparatus, the pyritic sulfur in the mixture can be economically removed. Emanuel Maxwell of our laboratory and his colleague Sergio C. Trindade conducted these studies. Application of the technique to the removal of sulfur from liquefied coal products is now under investigation. Achieving a practical process for this purpose will require a much higher degree of technical sophistication than was needed for the purification of kaolin.

A key potential application of high-gradient separation lies in the purification of water. Water is best purified by percolation through the ground or evaporation into the air, but the demands of modern communities have long since outstripped the capacity of these natural cycles. The primary criteria for urban water supplies have been that the incoming water be reasonably sterile and the outgoing water be something less than poisonous. The principal means of



BATCH SEPARATION utilizes a canister filled with a magnetic matrix and surrounded by a solenoid electromagnet enclosed in iron. A slurry of the materials to be separated flows through the strongly magnetized high-gradient matrix, where the paramagnetic particles in the slurry are retained (*a*). At intervals the feed is halted and the electromagnet is turned off so that the particles can be flushed out (*b*). Such a system can process 50 tons per hour and is equally suited to removing small amounts of magnetic impurities from materials such as kaolin clay, glass sands and pigments or to collecting small amounts of such metals as molybdenum and tungsten from low-grade ores or materials discarded by older processes.

achieving sterility has been chlorination, a process with a potential for toxic and carcinogenic by-products that is only now being recognized.

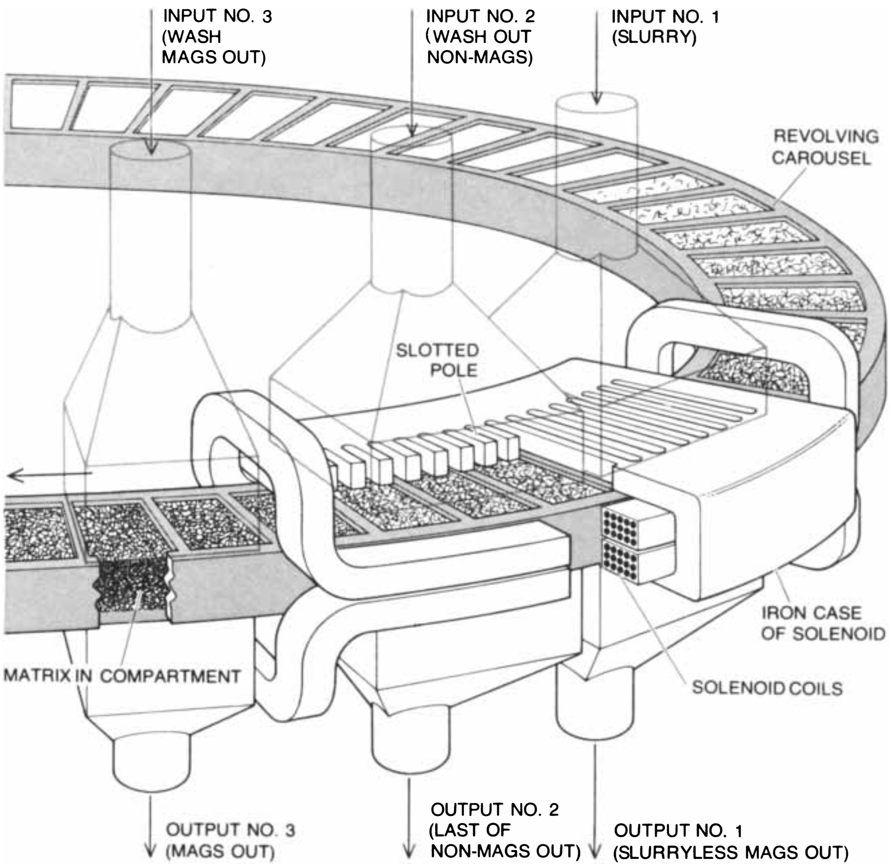
In fact, no form of water purification currently in use is acceptable in the long run. Most effluent treatments do little to diminish the eutrophic degradation of rivers and lakes into which the nutrient-laden wastewaters are discharged. Neither do most treatments prevent the cumulative loading of groundwater and other reservoirs with nutrients, particulate matter and heavy metals. A current example is the contamination of the water supply of Duluth and surrounding Lake Superior communities by the discharge of asbestos fibers into the lake. No conventional filtration methods will remove these carcinogenic particles.

Since high-gradient separation is an efficient means of collecting fine particles, it can be applied to water purification in a variety of ways. For example, the particulate contaminants that are present in steel-mill process water and wastewater are paramagnetic. Other contaminants can be made to associate themselves with magnetic particles seeded in the water; such associations can be brought about by surface adsorption, mechanical entrapment, coagulation or coprecipitation. For example, it has long been known that coliform bacteria, one of the commonest contaminants, tend to adhere to the surface of fine particles of iron oxide. Soon after Marston developed his ironbound solenoid he seeded water samples from the Charles River Basin in Boston with small amounts of iron oxide and subjected the samples to high-gradient separation at the remarkable rate of 150 gallons per minute per square foot of the magnetic matrix's surface. The Charles River drains parts of Suffolk, Norfolk and Middlesex counties, receiving a broad spectrum of effluents in the process, but it does not discharge directly into the sea. It is kept at an artificially high level by a dam near its mouth, and so it is virtually stagnant and has been heavily polluted with coliform bacteria and other contaminants for years. Marston found that a single high-velocity pass through the separator purified water from the river almost to the standards of drinking water. The treatment not only removed most of the coliform bacteria but also reduced the turbidity and the color of the water and lowered the number of suspended solid particles.

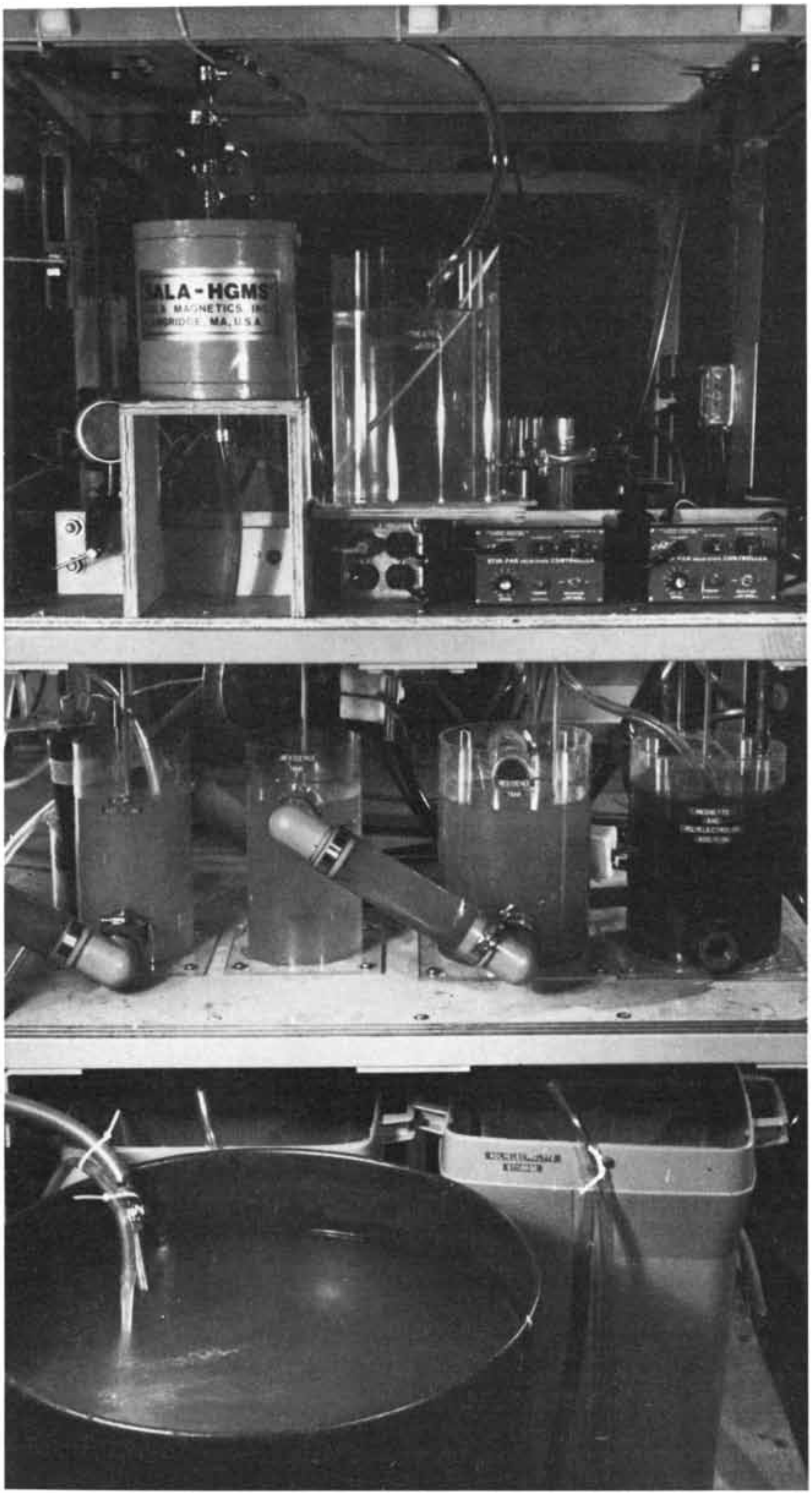
Christopher DeLatour, a graduate student working in our laboratory, followed up Marston's test with a series of separation experiments conducted in collaboration with the Metropolitan District Commission of Boston, the agency responsible for sewage and water treatment. DeLatour found that the process was almost as effective in purifying effluent from Boston's Deer Island sewage plant as it was in purifying river water. His results were sufficiently encouraging for the commission to order a full-scale engineering feasibility study to determine whether or not a high-gradient separator mounted on a barge could economically restore the waters of the Charles River Basin to recreational standards of purity. The conclusion of the study was affirmative. Swimming in the Charles may one day be a pleasure instead of a risky venture by an occasional rash teenager.

The removal of dissolved nutrients in wastewater, notably phosphate and nitrate, is mandatory if the lakes and streams that receive waste discharge are to escape eutrophication. In conventional water recycling the process that removes nutrients, referred to as tertiary treatment, requires filtration and chemical treatment that are elaborate and costly. It is therefore of particular interest that in preliminary tests DeLatour has been able to reduce the phosphate content of Charles River water and Deer Island effluent significantly by seeding the samples with bentonite clay, magnetite and aluminum sulfate. With these additives acting as coagulating agents, the nonmagnetic phosphate was trapped as the water passed through the high-gradient separator.

Probably the most dangerous contaminants of water are viruses. Their ability to pass through the finest filters is definitive, so that they not only are not removed by conventional water-recycling methods but also are not detected. The possibility of infection by, say, the hepatitis virus prevents the utilization of waste nutrients for growing food. It is therefore encouraging that Ralph Mitchell of Harvard University and Theodore G. Metcalf of the Uni-



CONTINUOUS SEPARATION utilizes a rotating ring divided into compartments filled with magnetic matrix. Such a system is embodied in the separators shown in the photographs on pages 46 and 47. As each compartment enters an elongated solenoid (*right*) a slurry is fed into it through slots in the magnet. While the compartment is still within the magnet it is washed (*center*) to rid the matrix of nonmagnetic particles that might be adhering to it. As compartments leave magnet (*left*) a second wash removes magnetic particles.



WATER IS PURIFIED in this laboratory high-gradient system at Sala Magnetics. The high-gradient magnetic separator itself is the small cylinder at the upper left labeled SALA-HGMS. At the bottom is a tank filled with dirty river water. Water drawn from it passes through a series of vessels, some of which are visible on the middle shelf. In them the *p*H is adjusted, alum is added as a coagulant and finally a polyelectrolyte and finely divided iron oxide are added. Most impurities adhere to the particles of iron oxide and are removed in the magnetic separator. The output of the process is the clear water in the vessel to the right of the separator. Overall the system removes coliform bacteria, viruses, suspended solids and some dissolved nutrients. It substantially reduces turbidity, color and odor. In one pass the coliform bacteria count is reduced from 16,000 per 100 milliliters to less than five.

versity of New Hampshire have found that certain viruses can be scavenged by iron oxide and other magnetic seeding materials and removed from water by the high-gradient separation process. Preliminary tests on a laboratory scale also show that asbestos particles of the kind and size entering Lake Superior as an industrial effluent can be removed from water by high-gradient separation. The development of a practical large-scale system for this purpose, however, would call for a substantial engineering effort.

Could the magnetic-separation method be applied to water recycling on a nationwide scale? Preliminary estimates indicate that its costs are competitive with conventional treatment systems and its space requirements (an important consideration in many urban areas) are substantially smaller. For very large systems, designed to treat billions of gallons per day, the energy requirement would be a serious limiting factor. Here superconducting magnets are likely to be used. In superconducting magnets the temperature of the coil is reduced to a few degrees above absolute zero; the coil becomes superconducting and current flows through it without resistance (and heating). One of the world's largest superconducting magnets, the ironbound solenoid used in a bubble chamber at the Argonne National Laboratory, generates a 20,000-gauss field at an energy cost no greater than that of illuminating the room where the magnet is housed. To generate an equivalent field with an ordinary copper solenoid would require about 12 megawatts, which is roughly the amount of power that is consumed by a small city.

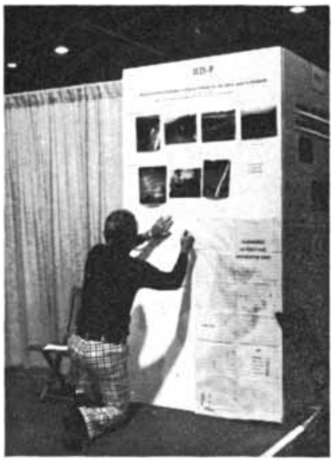
The Argonne superconducting magnet was built at a cost of about $3 million. If a magnet of similar size, generating a less intense field, were equipped with a steel-wool matrix, it could purify more than 100 million gallons of water per day at an extremely modest cost in energy. (For purposes of comparison, the five boroughs of New York City produce about 1.4 billion gallons of raw effluent per day.) For the present, economic factors and the level of the industrial commitment to magnetic separation favor the continued use of ordinary copper-coil magnets for any high-gradient system. High-gradient separation may, however, become the first major practical application of superconductivity. Such a development would make it possible to close the water cycle, to decontaminate lakes and rivers and to exploit many of the earth's currently unrecoverable minerals.

The new way to "give a paper"

Program chairpersons and committee members of the world, lend us your attention.

There is restlessness among the masses who are expected to sit quietly in the gloom fighting off Morpheus while a succession of fellow scientists, terrified or cocky as the case may be, mumble or declaim. The only questions which need to be asked by those familiar with the subject at hand will—likely as not—be turned off by the chair for lack of time, and the speaker will escape to the dark recesses of a concurrent session.

Here, from a recent meeting of a biological society, we show an alternate way to communicate by word of mouth with one's peers, formally and yet, for both giver and seeker of truth, comfortably. A "poster session" is an adult version of the school science-fair format. The presentation is boiled down to what can be put up in readily legible words and graphics on a board of, say 8 x 10 feet. It remains up for some hours. The peers drift by and look it over. When discussion ensues, the presenter need feel no embarrassment to learn more from the questioner than vice versa. The occasion is less of a performance and more of an exchange at a mutually agreeable level of discourse among parties fascinated by the topic.



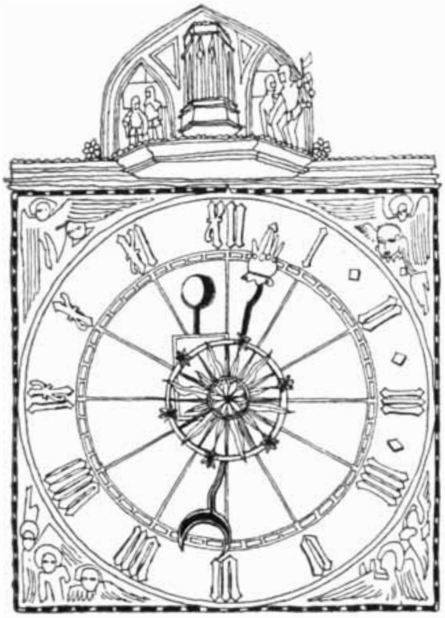
Photos taken at 1975 annual meeting of American Institute of Biological Sciences, Corvallis, Oregon.

The passerby need not come ready equipped with that fascination. It can be engendered on the spot, right then and there.

The presenter who would rather have people stop than not stop uses photography to fascinate. A few well-done color enlargements probably work better and easier under the circumstances than projected slides or movies. If table space can be provided for actual specimens and equipment, so much the better. But it's the photographs that attract.

Potential poster presenters who are not photographically inclined personally will generally find that their institutions employ people who are so inclined and are there to render service.

Kodak



SCIENCE AND THE CITIZEN

Rich Lands, Poor Lands

Where once the world seemed to be divided into Communist and non-Communist countries, it is now increasingly polarized into rich and poor countries. There are indications that confrontations between the developed and the underdeveloped nations will increasingly dominate international relations and specifically the business of the United Nations. The conflict was illustrated at the recent special session of the UN General Assembly on development and international economic cooperation. The session was marked by the most effective interchange yet attained between the rich and the poor nations, but its concrete actions were few. And the rich countries, led by the U.S., continued to resist the underdeveloped world's demand for a "new international economic order" that would, in the underdeveloped nations' view, redress old injustices and begin to redistribute the world's wealth.

Apart from the petroleum producers the underdeveloped countries have been getting poorer: per capita incomes have declined in recent years. The Third World's share of world trade has been reduced. Its exports, largely raw materials, are subjected to severe market fluctuations, which can be particularly devastating for one-crop countries. As underdeveloped nations seek to build up infant industries they find their manufactures confronted by tariff and nontariff barriers that often exceed those the developed nations negotiate with one another. Meanwhile inflation in the industrial countries and the sharp rise in the cost of fuel have escalated the cost to poor nations of essential imports. For many underdeveloped nations the terms of trade (price index of exports divided by price index of imports), after improving for a few years, have been deteriorating. The poor countries' annual trade deficits have increased, and those deficits have in turn increased the poor countries' total burden of debt.

The debate at the special session dealt not only with specific actions but also with sensitive conceptual issues. The U.S. in particular resisted such phrases as "new international economic order" and any imputation to the rich nations of blame for economic injustices. The session ended with unanimous agreement on a resolution, aimed at "redressing the economic imbalance between developed and developing countries," that compromised the rhetorical arguments. Most of the concrete actions considered at the session were referred to UN agencies and study groups. The poor countries, for example, sought "indexation," which would tie the price of their export commodities to the price of the industrial goods they import. The developed countries rejected indexation, and the idea was referred for further study. The poor countries asked the rich countries to agree to an annual level of official development assistance equal to .7 percent of the industrial countries' gross national product. (The average has been about .3 percent lately; the U.S. contribution, one of the lowest among the industrial nations, has been less.) In the final resolution the target of .7 percent was mentioned, but without any pledge of acceptance.

The resolution also recommended a number of actions, most of which will depend on voluntary agreement by individual nations, to stabilize the purchasing power of commodity-exporting underdeveloped countries, to improve the transfer of capital and of industrial and technological knowledge and to lower the barriers that inhibit underdeveloped countries' exports. The most significant long-term result of the session may be a restructuring of the economic and social machinery of the UN itself to enable it to deal more effectively with issues of economics and development. One change that has been proposed would be the creation of a new director-general for development and international economic cooperation, who would be second in rank only to the UN Secretary-General, to coordinate all activities in those sectors.

Engines for the Eighties

A comprehensive study of alternative automobile power plants conducted by the Jet Propulsion Laboratory of the California Institute of Technology vigorously urges that a $1-billion program be launched immediately to develop a new engine for introduction by 1985 or sooner. After observing that "the automobile will maintain its dominant role in personal transportation through the foreseeable future," the study concludes that two engines, the gas turbine and the Stirling-cycle engine, promise significantly greater fuel economies than such widely discussed alternatives as the diesel engine, the Rankine (steam) engine, the all-electric car, hybrid configurations (part heat engine, part electric) or any conceivable improvement in the present Otto-cycle engine. A fully developed gas-turbine engine in a fleet of cars comparable in size and performance to those built in the 1975-model year should provide about 22 percent more miles per gallon than an equivalent fleet powered with a "mature" (maximally improved) Otto-cycle engine. A comparable fleet powered with a Stirling engine should show an even greater mileage improvement: about 35 percent. If all cars were converted to one of the new engines, the U.S. fuel saving in the 1990's should amount to about $10 billion per year (assuming that the price of crude oil at that time would be $11 per barrel).

The fleet of cars built in the 1975-model year averaged 15.6 miles per gallon on the Environmental Protection Agency's composite city-highway driving cycle. The Federal Government has asked the industry to achieve an average of 19.6 m.p.g. in the 1980-model year, representing a 40 percent improvement over the 14-m.p.g. fleet average in 1974. The 19.6-m.p.g. target assumes that there will be not only improvements in vehicles and engines but also a strong market shift to smaller cars. (The EPA has recently announced that the 1976-model fleet averages 17.6 m.p.g. on the composite cycle, and that a fifth of the improvement is attributable to the larger fraction of small cars in the assortment of cars on the market.) The Jet Propulsion Laboratory study estimates that if the assortment of cars on the market had remained unchanged, the best that could have been achieved by improvements in vehicle design and in the Otto engine would be an average fuel economy of 18.6 m.p.g. With a mature gas-turbine engine a comparable fleet of cars should achieve 22.7 m.p.g., and with a mature Stirling engine they should achieve 25.2 m.p.g. If a comparable fleet were equipped with a mature diesel engine, it would get only 19.5 m.p.g., and if it were provided with a mature steam engine, it would get only 15.6 m.p.g. (The diesel figure uses an energy-equivalent "gallon," since diesel fuel contains about 11 percent more energy than gasoline on a volume basis.) The JPL study sees little promise in electric vehicles without a "major breakthrough in battery technology." It also finds hybrid vehicles "singularly unattractive" because of the weight and cost of energy-storage systems.

The study points out that the gas turbine and the Stirling engine have an overwhelming advantage over all forms of Otto and diesel power plants for achieving and even surpassing the statutory standards for exhaust emissions. The reason is that in the gas turbine and the Stirling engine the combustion process is physically divorced from the work-producing process, hence combustion can be optimized to minimize pollutants. In the Stirling engine the heat of combustion is transferred to a separate closed system containing a small quantity of high-pressure gas as a working fluid. In producing power the working fluid is transferred between cylinders in such a way that it is alternately cooled during compression and heated during expansion. Both the Stirling engine and the gas turbine have "ideal" thermodynamic efficiencies in the vicinity of 66 percent. Comparable ideal values are 55 percent for the diesel engine, 45 percent for the Otto engine and 33 percent for the steam engine.

The JPL study emphasizes that a major effort will be required to achieve "mass-producible" versions of the gas-turbine and Stirling power plants. "Given a halfhearted effort, they may never be ready for introduction." Thus the report urges that "government should provide incentives and/or share in the funding to ensure that this program *will* be accomplished." The study was financed by a $500,000 grant from the Ford Motor Company, which, according to the JPL, "meticulously avoided any action that might influence the conduct of the study or the conclusions."

Picture of a Protein

A protein molecule is an elaborately coiled and folded structure in which form largely determines function. Over the past 20 years the structures of a number of proteins have been elucidated by studying the diffraction of X rays passing through a crystallized specimen of the protein. A quite different method of structural analysis has now been devised by Nigel Unwin and Richard Henderson of the Medical Research Council Laboratory of Molecular Biology in Cambridge, England. Instead of irradiating the protein with X rays, Unwin and Henderson make electron micrographs of it; in addition to a diffraction pattern, their technique yields an actual image of the molecule, albeit one that cannot be seen without intensive enhancement.

The protein studied by Unwin and Henderson is found in the cell membrane of *Halobacterium halobium,* a photosynthetic bacterium adapted to an environment of concentrated brine. The protein is confined to certain differentiated regions of the membrane, which because of their pigmentation are called purple patches. Within the patches the protein makes up 75 percent of the mass of the membrane, and the protein molecules are packed together in an orderly, periodic array one molecule thick; in other words, they form a two-dimensional crystal. Because the crystal is so thin X-ray studies of single crystals are impossible; there is not sufficient depth to generate a distinct diffraction pattern. Unwin and Henderson therefore turned to electron microscopy, applying principles developed by David DeRosier and Aaron Klug, also at the Medical Research Council Laboratory in Cambridge.

The transmission electron microscope has a potential resolution of about three angstroms, more than enough to reveal the internal features of a single large molecule. When biological molecules are investigated, however, that resolution cannot be achieved because the intense flux of electrons required would instantly destroy the specimen and with it the structural features being studied. Unwin and Henderson avoided this problem by employing relatively low magnification and reducing exposure to extremely low levels; the equivalent procedure in light microscopy would be to observe at low power with a dim light source. They also omitted the staining of the specimen, which is ordinarily employed to improve contrast. The electron micrographs made in this way are unreadably faint and have negligible contrast; nevertheless, Unwin and Henderson extract enough information from them to reconstruct a picture of the protein molecule with a resolution of seven angstroms.

At the low dosage employed only about 25 electrons pass through an area of membrane seven angstroms square. The maximum contrast—the difference in illumination between the brightest areas and the darkest ones—is less than 1 percent. As a result noise, or random fluctuation in the number of electrons transmitted, is of far greater magnitude than the desired signal. The structural information is extracted from this background by averaging the images of thousands of protein molecules; when that is done, the random fluctuations tend to cancel, whereas the fluctuations produced by the specimen tend to reinforce one another. The averaging is accomplished by Fourier transformation, in which features of the image that repeat at various frequencies are combined; the Fourier transformation is carried out by a computer after the micrographs have been encoded in digital form. Finally, the entire process must be repeated several times with the specimen tilted to provide projections from various angles. The ultimate result is a three-dimensional reconstruction of the density distribution within the protein molecule.

Writing in *Nature,* Unwin and Henderson show that the purple-membrane protein is composed of seven helical segments, all aligned roughly perpendicular to the plane of the membrane and extending all the way through it. The helixes must be linked at their ends by shorter segments, but those have not yet been resolved. In the membrane the protein molecules appear to be organized in groups of three, and some of the 21 helical segments in each group seem to form a superhelix. Interspersed among the protein molecules are molecules of the lipid bilayer that forms the basic structure of all membranes, and molecules of the pigment retinal, which gives the purple patches their color.

The purple patches of *H. halobium* are of interest because they offer a relatively simple system in which to investigate the interaction of light with living matter. In the bacterium they are involved in photosynthesis, and Walther Stoeckenius and his colleagues at the University of California at San Francisco School of Medicine have shown that the purple patches function as an ion pump; in the presence of light they transport hydrogen ions (protons) across the cell membrane. The membrane protein is presumably the agent of that transfer, and knowledge of the structure of the protein may provide the key to how it operates.

A Technology Assessed

Three years ago Congress set up an Office of Technology Assessment in response to suggestions that many problems of society result from the introduction of technologies without regard for their impact on the society as a whole. The aim of technology assessment is to predict such impacts and if possible to propose measures by which they can be controlled. An example of this concept at work is provided in *Oil Transportation by Tankers: An Analysis of Marine Pollution and Safety Measures,* a study published by the Office of Technology Assessment. The agency undertook the study at the request of the Senate Committee on Commerce, which has announced its intention to hold hearings on the effects of the Ports and Waterways Safety Act of 1972 and other legislation affecting tankers.

The study examines the evolution of tankers and the pollution they cause, presents approaches for reducing pollution and improving the safety of operation of tankers and reviews the international and domestic regulatory authority affecting tanker operations. One finding is that tankers spill about 1.5 million tons of oil into the seas every year: a million tons from such normal operations as cleaning tanks and dumping ballast, 200,000 tons as a result of accidents and 250,000 tons associated with drydocking. This spillage accounts for nearly a third of all the oil pollution of the world's oceans.

A number of other findings relate to means of reducing the pollution caused by tankers and of improving the safety of operation of tankers. One finding is that fitting tankers with double bottoms or double hulls offers "a significant degree of protection from oil pollution" in the event of a tanker accident. A double bottom also provides space for segregating the seawater ballast from the oil tanks, so that the ballast does not have

Programmability for $79.95
The new Sinclair Scientific Programmable.

Sinclair's new Scientific Programmable is no ordinary calculator. It has only 19 keys—and a very low price—but its problem-solving capability far exceeds that of any ordinary scientific calculator. Sinclair's big plus is programmability—the ability to remember a calculation sequence of up to 24 steps entered directly from the keyboard. Once stored in the program memory, a calculation sequence can be recalled at the touch of a single key, and applied to new numbers to produce new results. So if you carry out repetitive calculations, the Scientific Programmable will save you key strokes and lots of time. It will also save you from errors, since the calculation sequence will be recalled exactly as you entered it, over and over again.

Programming Entering a calculation sequence into the program memory couldn't be easier. Press the B E key to tell the calculator to remember a sequence. Then key in the calculation almost exactly as you normally would. Use VAR at the points where you'll want the program to stop, so that you can enter new numbers or display partial results. During entry, the number of steps is displayed, so you won't exceed the program memory's capacity. When you've finished, press B E again to tell the calculator that the sequence is complete. The program is now available until you overwrite it with another program or until you switch the calculator off. Just enter new numbers, press EXEC, and get new results. Of course, the Scientific Programmable can be used as an ordinary scientific calculator. Even half-way through the execution of a program, you can stop, carry out a calculation from the keyboard, then press EXEC to continue execution of the program when you're ready.

Sinclair Program Library You don't have to be a programmer to use the Scientific Programmable's full problem-solving capability. Each calculator comes with a fully documented library of hundreds of programs to solve standard problems, and complete instructions on how to use them.

Applications The Scientific Programmable has many applications in areas of computation other than repetitive calculations. It can be used for the analysis of experimental data, the evaluation of integer functions, and with a variety of methods for the iterative solution of equations. They are dealt with in detail in the instruction book and Program Library. And, although the Scientific Programmable is no toy, it does play games—and wins!

Display 5 digit mantissa, 2 digit exponent. Number entry is floating decimal point and/or scientific notation; results in scientific notation.

Logic Reverse Polish Notation.

Functions Sine, cosine, arctangent (radians), log, alog (base 10), square root, reciprocal, change sign, clear/clear entry.

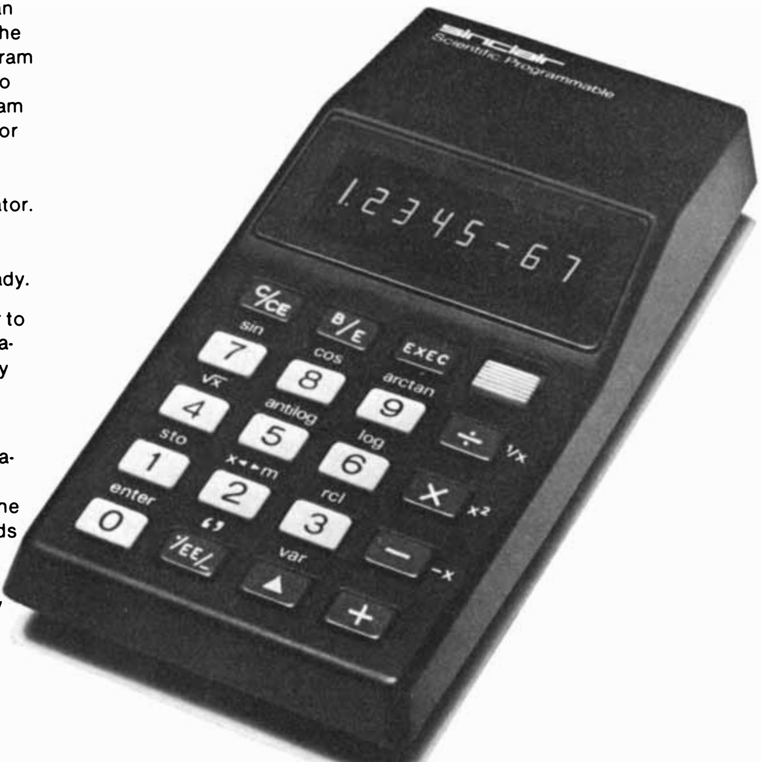
Memory Store, recall, exchange.

The Scientific Programmable is small enough to hold in your hand and big enough to use on your desk. It measures 6" x 2-7/8" x 1-1/4", has non-slip rubber feet and a big green display. It uses a small, inexpensive 9v battery or the Sinclair AC adapter.

$79.95 includes battery, AC adapter, carrying case, instruction book and the Sinclair Program Library.

10 Day Trial Offer If the Scientific Programmable doesn't meet your requirements return it to us within 10 days for a prompt refund.

Warranty The Scientific Programmable is backed by Sinclair's full one-year parts and labor warranty.



sinclair™ **THE LOGICAL CHOICE.**

Sinclair Radionics Inc., 375 Park Avenue, New York, NY 10022 (212) 688-6623

Please send me____Scientific Programmable(s) at $79.95 each
(NY residents add sales tax).

Enclosed is my check for____ Please charge my credit card account

Credit card______Account #______Expires______

Name______

Signature______

Address______

______

Mail to: Sinclair Radionics Inc., 375 Park Avenue, New York, NY 10022
Or call: (212) 688-6623

to be carried in empty oil tanks and will not contain oil when it is dumped. (A bill now pending in the Senate would require double bottoms on most U.S.-flag tankers built in the future.) Another finding is that filling the space between the top surface of the oil and the top of the tank with an inert gas "can substantially reduce risks of tank explosions." The inert gas displaces the oxygen needed to sustain an explosion. The agency also found that "improvements in the training and licensing of shipboard personnel are greatly needed" in the light of the fact that "a substantial portion of tanker accidents are caused by human error."

Wave of the Future

In 1969 Joseph Weber of the University of Maryland announced that he had detected bursts of gravitational radiation from the center of the galaxy. In the years since then more than a dozen groups of investigators all over the world have independently built gravitational-wave antennas in an attempt to confirm Weber's observations. Although some of the detectors are several times more sensitive than Weber's original apparatus, so far all the results have been negative. Nevertheless, points out Kip S. Thorne of the California Institute of Technology, gravitational-wave astronomy may well be a reality some five years from now.

Gravitational waves are predicted by the general theory of relativity. Just as an accelerating charge gives rise to oscillating electric and magnetic fields that propagate as electromagnetic radiation at a universal velocity (300,000 kilometers per second), so an accelerating mass should give rise to a gravitational field that propagates as a gravitational wave at the same universal velocity. Gravitational waves resemble electromagnetic waves in that they carry energy, momentum and information; they are also characterized by frequency and wavelength. Gravitational waves differ from electromagnetic waves, however, in that they interact with all forms of matter instead of only with electric charges or currents.

In 1918 Einstein calculated the behavior of gravitational waves that would be generated by the acceleration of a slow-moving mass with a weak gravitational field. Today, in preparation for the era of gravitational-wave astronomy, Thorne argues that physicists should understand how gravitational waves would be generated by fast-moving objects with strong gravitational fields. Thorne and his colleague Sándor J. Kovács have applied the general theory of relativity to the problem. In a recent issue of *The Astrophysical Journal* they summarize the results of the "plug in and grind" formalism they have devised.

They break down the gravitational radiation into five different types. The first type is direct radiation (the type Einstein analyzed), generated by the source when it accelerates through space. The second type is "whump" radiation, generated when the source is gravitationally stressed. The third type is transition radiation, generated by time-varying delays in the propagation of the source's ordinary nonradiative gravitational field. The fourth type is focusing radiation, which arises when one part of the source focuses the nonradiative field of another part of the source. The last type is tail radiation, emitted when the nonradiative field is scattered backward in the region of focusing.

"Although Weber's 'events' may turn out to be nongravitational in origin," Thorne and Kovács state, "second-generation detectors... with amplitude sensitivities roughly 100-fold better than today's [antennas] are now under construction, and third-generation detectors are being discussed." Such third-generation detectors, with a gravitational antenna consisting of a crystal of sapphire weighing some 10 kilograms, should be able to pick up bursts of gravitational waves generated several times a year by the explosion of supernovas in the Virgo cluster of galaxies, 50 million light-years away. Other detectors may eventually succeed in receiving gravitational waves from pulsars, from near-encounters of stars in dense star clusters and from the nuclei of galaxies and quasars.

Dispensing with the Plow

In recent years much consideration has been given to the concept that plowing the soil is not necessary in many kinds of agriculture where it is usually taken for granted. According to the results of a 12-year study conducted at the Ohio Agricultural Research and Development Center in Wooster, Ohio, growing corn on unplowed fields can significantly increase the yield of the crop on many types of soil. The no-plow system offers additional benefits: a substantial reduction in soil erosion, increased retention of rainwater, quicker planting and savings on tractor fuel.

The no-plow system calls for a special planter that can operate on fields that are covered with stalks and roots. The planter cuts a narrow furrow and simultaneously lays the seed in it. A band of fertilizer is laid along the furrow, and later insecticide is applied. Describing the results of the Ohio study in *Crops & Soils Magazine,* David M. Van Doren, Jr., Glover B. Triplett, Jr., and James E. Henry report that on well-drained silt-loam soil the yield of corn with no plowing was always equal to or greater than the yield with conventional plowing and secondary tillage.

Other workers had reported that on poorly drained soils the no-plow system generally led to reduced yields. The results of the Ohio study, however, show that the no-plow system can be applied to some types of poorly drained soil with no decrease in yield. On poorly drained silt loam and silty clay loam, for example, a two-year rotation of corn and soybeans or a three-year rotation of corn, oats and alfalfa produced the same yield of corn on both the unplowed fields and the plowed ones. Continuous seeding of corn on these soils eventually led to lower yields on the unplowed fields. On a poorly drained clay soil the yield of both continuous corn crops and rotated ones was the same on both kinds of field.

The matched test plots received the same amounts of fertilizer and lime (to reduce soil acidity). Nitrogen fertilizer was broadcast on the fields before they were seeded. The plots were thinned to the same stand each year. Vegetation on the unplowed fields was killed with herbicides before seeding. Different mixtures of herbicides were applied, depending on the type of weeds that were present. The herbicides also kept the field free of weeds during the growing season.

Because the soil had a vegetative cover throughout the year, soil erosion was reduced and the infiltration rate of rainwater was increased. One year during a drought corn on an unplowed field continued to grow while the corn on a nearby plowed field wilted. Van Doren, Triplett and Henry note that no-till farming is particularly applicable in areas where erosion is a problem because the land has steep rolling contours. There is already a trend among farmers in the corn-growing regions of the Middle West to do less plowing, which breaks down soil structure and leads to soil erosion. In many instances, the Ohio investigators conclude, the savings in tractor fuel and in the farmer's time more than offset the extra cost of the herbicides that are required for no-plow farming.

Announcing Jaguar XJC.
The Corporate Sports Car.

Beneath the polished well-disciplined manner of a corporate executive there often lurks a spirit that yearns for the freedom of a sports car. It is to this duality in a successful man's character that we dedicate our new two-door vinyl-top coupe, the Jaguar XJC.

The XJC moves like a sports car because it is fitted out like a sports car with power-assisted rack and pinion steering and independent suspension, steel-belted radial tires and power disc brakes on all four wheels.

You may have the XJC with the incredibly smooth fuel-injected Jaguar V-12 engine that delivers turbine-like acceleration throughout the entire range of driving speeds. Or, you may prefer an XJC powered by the history-making double-overhead cam six.

Inside, the XJC is all Jaguar luxury, with rich cabinet woods and deep carpeting. Thermostatically-controlled air-conditioning, eight-track AM/FM stereo sound system, electric windows and central door locking system are all standard. And now, in addition to its uncommon luxury and high performance capabilities, the Jaguar XJC provides its owners with a new warranty.

For 20,000 miles or 12 months, whichever comes first, Jaguar will replace or repair any part in the car that is defective or that simply wears out, provided only that the owner maintains the car properly. The only exceptions are the tires, which are warranted by the tire manufacturer, and spark plugs and filters, which are routine replacement items. Even then, if they are defective, Jaguar will pay to replace them.

See this magnificent Jaguar XJC soon. Become a sports car enthusiast even though you may wear a corporate executive's clothing. For the name of the dealer nearest you, call these numbers toll-free: (800) 447-4700, or, in Illinois, (800) 322-4400. British Leyland Motors Inc., Leonia, New Jersey 07605.

Jaguar


IBM

IBM announces the new 5100 Portable Computer

A compact problem-solving aid for engineers, statisticians, scientists and financial and business analysts.

Now you can have a computer right on your desk. Exactly where you need it. When you need it.

The new IBM 5100 Portable Computer incorporates the latest in semi-conductor technology. It features a typewriter-like keyboard and numeric key-pad for simplified data entry, a 1024 character display screen, an integrated magnetic tape drive, and 16K characters of memory.

Options available with the 5100 include a bi-directional 80-characters per second printer, a second magnetic tape drive, and additional memory up to a maximum of 64K characters. Also available is a communications feature which allows the 5100 to be used as a terminal.

The IBM 5100 comes with either APL or BASIC language or both.

Over 100 often-used analytical routines in mathematical, statistical and financial calculations are available for such functions as forecasting, modeling, matrix arithmetic, engineering and design calculations, regression and correlation analysis, return on investment and cash flow analysis.

In addition, the 5100 features a self-study training package that makes it easy to learn and easy to use without taking any classes or relying on specially trained experts.

If you'd like to find out more about IBM's new 5100 Portable Computer and arrange for a demonstration right at your desk, call your IBM General Systems Division office or fill out this coupon.

SA

IBM®

IBM General Systems Division
P.O. Box 2068, Atlanta, Georgia 30301

☐ I would like more information about IBM's new 5100.
☐ I would like a demonstration of IBM's new 5100.

My major area of interest is:
☐ Engineering/Scientific ☐ Statistical Analysis
☐ Business/Financial Analysis

Name________________________________
Title________________________________
Company______________________________
Address______________________________
City____________State__________Zip______
Phone_______________________________

THE CANCER PROBLEM

Almost all cancers appear to be caused by exposure to factors in the environment. The most promising approach to the control of the disease is to identify those factors and eliminate them

by John Cairns

During the past 150 years the Western world has virtually eliminated infectious diseases as a significant cause of death. A child born in the U.S. today can look forward to a life untroubled by fear of diseases such as scarlet fever, diphtheria, tuberculosis, typhoid fever and dysentery, which were major causes of death three or four generations ago. Life expectancy has been increasing steadily since the middle of the 19th century. The longer average life span is a result mainly of improvements in public health; the more spectacular fruits of scientific research, such as the introduction of vaccines and antibiotics, merely completed the process.

Death has now been confined mostly to old age and can therefore be attributed to diseases that are either peculiar to old age or lethal only in old people. Although innumerable changes in the body that accompany advancing age could be classified as diseases, two particular conditions commonly arising in old age are often a direct cause of death: arterial disease (atheroma and arteriosclerosis) and cancer. Arterial disease is lethal when it affects the arteries supplying the heart or the brain; it now accounts for about 50 percent of all deaths in the U.S. Cancers are lethal when they spread from their site of origin; they now account for almost 20 percent of all U.S. deaths.

People have very different attitudes toward these two diseases. It is not just that death from arterial disease is often rapid, whereas death from cancer can be painfully drawn out. For some reason heart attacks and strokes tend to be thought of as natural hazards of age, and either a normal end to a satisfactorily long life or, when they occur in middle-aged men, the wages of overeating and lack of exercise. In contrast, cancer is thought of as an unpredictable disease that strikes indiscriminately at rich and poor, fat and thin, old and middle-aged, as if it usually owed nothing to external causes. If that were true, our only hope of overcoming cancer would be to improve the treatment of the disease. One object of this article is to show that most of the common kinds of cancer seem to be caused in large part by environmental factors; because we can act to alter the environment, those cancers are potentially avoidable.

Incidence

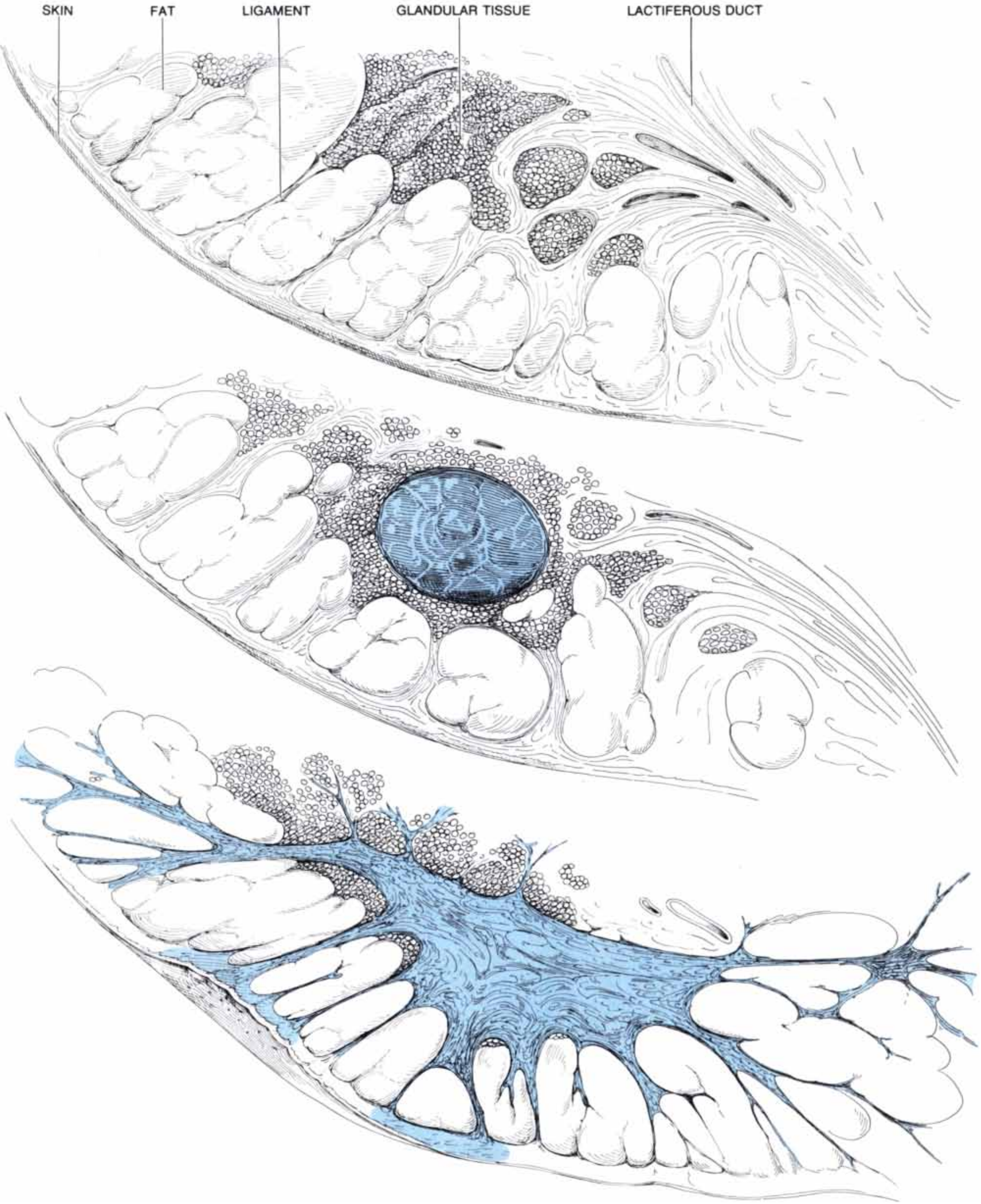
Groups of abnormally proliferating cells can arise in any part of the body. Those that cannot invade the surrounding tissues and so remain strictly local growths are called benign tumors. Those that spread from their site of origin and can therefore reach the bloodstream and the lymphatic system are called malignant tumors, or cancers.

The cancers are divided into three broad groups. The carcinomas arise in the epithelia, the sheets of cells covering the surface of the body and lining the various glands. The much rarer sarcomas arise in supporting structures such as fibrous tissue and blood vessels. The leukemias and lymphomas arise in the blood-forming cells of the bone marrow and lymph nodes. These three words—carcinoma, sarcoma and leukemia—are so entrenched in everyday usage that they must be mentioned, but I do not mean to imply by their use that there are three basically different forms of carcinogenesis or that the three kinds of cancer have different prospects for prevention and cure. That kind of information can be obtained only through a finer system of classification.

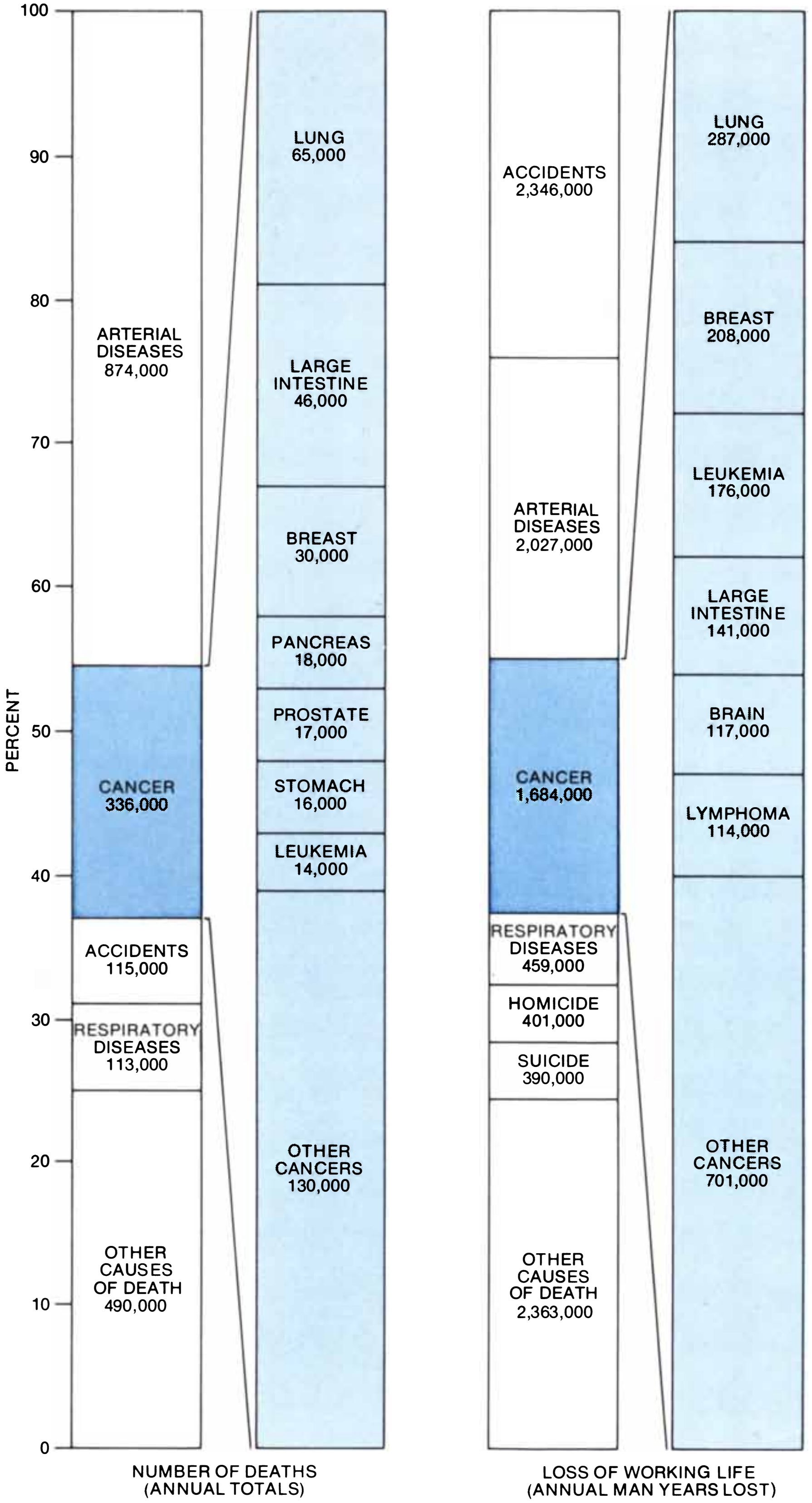
Cancers are classified mainly by the organ in which they originate and by the kind of cell involved. When they are considered in this way, there are 100 or so distinct varieties of the disease. Such an elaborate classification would be of no general interest were it not that the different varieties plainly have different causes, since the incidence of each one changes independently when the environment is altered. Most of the 100 varieties are rare, and so we can account for most cancer mortality by considering a fairly short list of diseases.

Roughly half of all cancer deaths are caused by cancers of three organs: the lung, the large intestine and the breast [*see illustration on page 66*]. There can therefore be no major inroad on overall cancer mortality until some means are found for curing or preventing these three kinds of cancer. Each of them can be considered a discrete entity because the frequency of each varies independently when factors in the environment are changed.

It could reasonably be argued that we are not interested in total numbers of deaths as much as in loss of life span. The death of a 90-year-old man from cancer of the prostate is less of a tragedy than the death of a young man from leukemia. In determining our priorities we should therefore take into consideration the age distribution of the victims of each cause of death. There are various ways of doing this. For example, it is possible to calculate how much each major cause of death or each kind of cancer diminishes the average life expectancy or, in particular, how much each reduces our working life up to the age of 65. The main effect of such a weighting procedure is to increase the relative importance of accidents among the general causes of death, and of the leukemias and lymphomas among the cancers. Can-



GROWTH OF A TUMOR in the breast ordinarily threatens life only when the tumor can spread to distant parts of the body. The normal breast (*top*) is organized into glandular tissue, fat and other structures. Tumors arise almost exclusively in the glandular tissue; they are composed of cells in which the normal restraints on growth and reproduction have been removed. A benign tumor (*middle*) can grow rapidly and become quite large, but it cannot escape the tissue in which it develops. A cancerous tumor (*bottom*) can spread throughout the glandular tissue, can often involve ligaments and skin and can sometimes penetrate the muscle underlying the breast. In addition cancers can in some cases migrate through the blood or the lymphatic system to establish new colonies of cells in distant, unrelated organs. This is the process known as metastasis. It is the metastatic spread of cancers that is responsible for their lethality.



DEATHS FROM CANCER now make up almost a fifth of all deaths in the U.S. Of the deaths that are attributed to cancer, more than 60 percent are caused by a few common forms of the disease. The remainder are distributed among more than 100 other cancers. The impact of each kind of cancer can also be judged from the loss of life span it causes, measured here in lost working years, with working life assumed to extend from age 20 to age 65. Among all causes of death, the principal effect of considering loss of life span instead of simple mortality is to increase the importance of accidents; among cancers, it emphasizes relatively high incidence of leukemias in young people. Lung cancer, however, is still predominant.

cer of the lung, however, still remains at the top of the list.

All these statistics refer to mortality, and that is the most accessible and reliable measure of the impact of cancer. Estimating the incidence of the different cancers is not as easy. To begin with, one might consider the patients who present themselves to a physician and are found to have cancer. Most of them will sooner or later die of their cancer; this single depressing statistic reinforces the point that science has had little impact on diseases that mainly affect the middle-aged and the old. An alternative definition of incidence would include not only those cancers that have begun to cause symptoms but also those that can be detected by a deliberate search. Many kinds of cancer have been sought in this way by routine surveys, and it has become plain that far more small, symptomless cancers can be detected than might have been expected. It follows that most of those minute collections of invasive cells must either regress before they become very large or, more likely, grow so slowly that they do not give rise to symptoms during the patient's lifetime. For example, the conventional estimate of the incidence of cancer of the prostate in 70-year-old men is about 200 cases per 100,000 men per year, or .2 percent per year; routine autopsies of 70-year-old men who had died of other causes, however, have shown microscopic invasive cancers of the prostate in from 15 to 20 percent of them. The incidence measured by this method is thus 100 times as great.

When the site of the cancer is accessible to direct examination, surveys are much easier to carry out; they give the same result. For example, a recent survey in a rural district of Tennessee showed that about 4 percent of the adult population have skin cancer. Indeed, it seems likely that if we could extend such a detailed examination to the entire body we would find that by middle age each of us has acquired several nests of proliferating, invasive cells that might reasonably be classified as cancers. Without knowing much more about the natural history of the disease we cannot predict which of these cancers will spread and which will not. Before discussing the natural history of cancer, however, I should like to consider what can be deduced about the causes of the disease by studying its epidemiology.

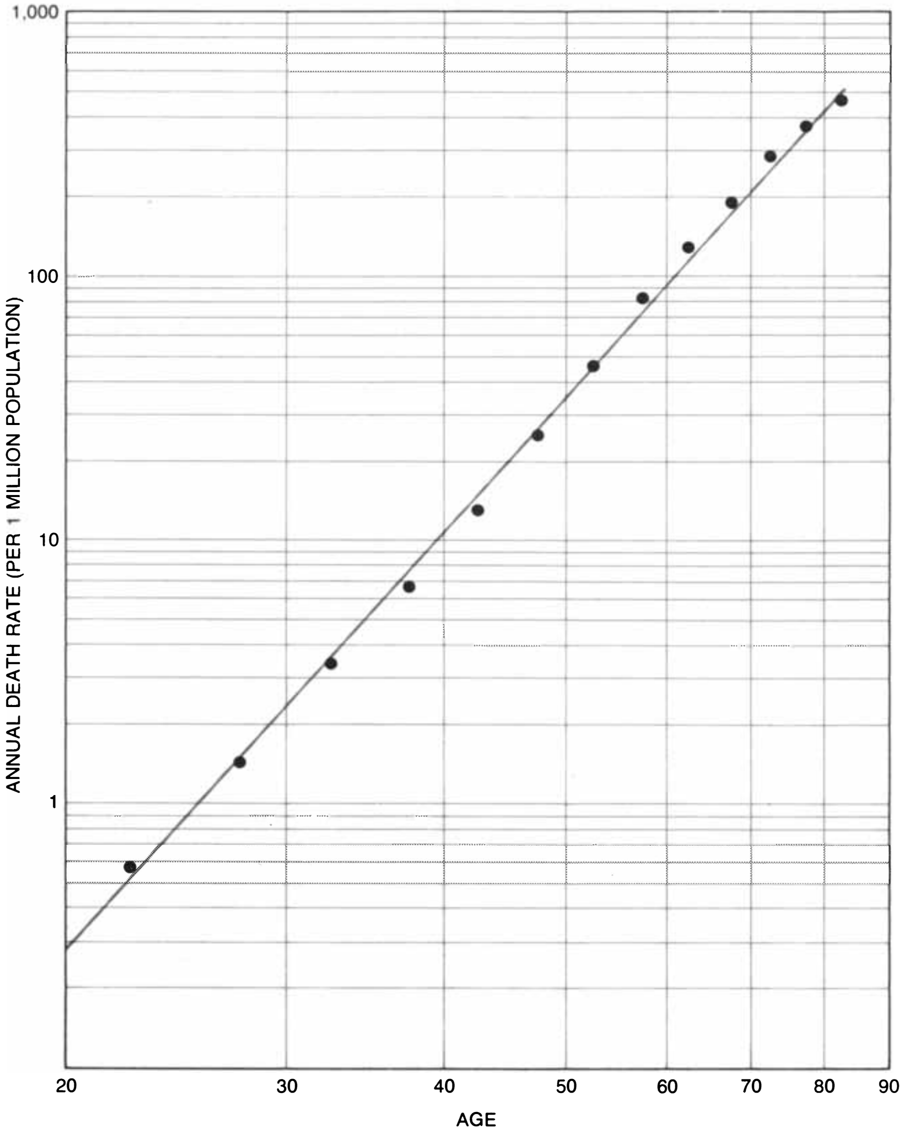
The first step in finding out what causes any particular variety of cancer is to determine which groups of people show the highest incidence and what distinguishes them from other people.

Those most conspicuously at risk are of course the old. Almost all kinds of cancer are much commoner in old people, and the incidence rises steeply with age. To take a typical example, the death rate from cancer of the large intestine increases about a thousandfold between the ages of 20 and 80, and most of the increase comes after age 60 [*see illustration at right*].

Model of Carcinogenesis

Various models have been proposed to account for the clustering of cancer in old age. One of the most reasonable models postulates that each cell has several genes that independently restrain it from forming a cancer, so that it will not form one until each of those genes has been inactivated by mutation. Because mutations can be introduced at any time in the life of a cell or of its ancestors, the probability that any one of our cells has a mutation in a particular gene increases in direct proportion to our age. The probability that the cell has mutations in all *n* of its *n* restraining genes (and is thus cancerous) therefore rises as the *n*th power of our age. The risk of having cancer should therefore increase as the *n*th power of our age; expressed another way, the logarithm of cancer incidence should be linearly related to the logarithm of our age. The theory is often in excellent agreement with the observed age distribution of cancer. Assuming that the model is correct, one can calculate from the slope of the linear logarithmic relation the number of mutations needed to create a cancer. In a typical cancer, that of the large intestine, the number appears to be about five.

That interpretation of the relation between age and incidence has one important implication. Each cancer is considered to be the end result of several mutational steps that may have taken place at any time in the patient's life. The total incubation period of any cancer therefore dates back to the moment when the first step took place and so must often extend over much of the patient's lifetime. In some instances this is demonstrably true. For example, the incidence of lung cancer in each of several countries is directly proportional not to the number of cigarettes its inhabitants are smoking today but to the number they smoked about 20 years ago. Similarly, occupational cancers induced by exposure to certain industrial chemicals may not appear until from 10 to 20 years after a person has retired from work. One particularly good but rare example of a long incubation period is provided by cancer of the penis. The cancer is seen only in old men, but it is certainly caused by factors operating in youth because it is prevented by circumcision in the first few days of life but not if circumcision is postponed for a few years. Finally, as we shall see, various studies of migrant populations show that the incidence of many common cancers is partly determined by our environment in youth. It follows that when we ask what has caused a particular cancer, we must not confine our attention to the patient's recent past. Conversely (and this is most important), if we inadvertently start to expose a population to some carcinogenic agent, it may be many years before the first cancers awaken us to the danger, and by then it may be too late to prevent the wave of cancer cases that is about to come.



OLD PEOPLE compose the subpopulation that is most conspicuously at risk in the development of cancer. The incidence of almost all forms of cancer increases dramatically with advancing age. Here the U.S. death rate from a representative cancer, that of the large intestine, is plotted against age. It can be seen that the logarithm of the death rate is linearly related to the logarithm of age. The relation can be explained by the hypothesis that several mutations are required to generate a cancer, and that the probability of each mutation is proportional to age. The slope of the line suggests that the number of mutations required is five.

Environmental Factors

Knowledge of the relation between age and death rate does not tell us what causes a cancer, only that the steps leading up to it are probably accumulated over many years. What we want to know is whether or not the main causative factors are environmental (and therefore potentially avoidable). For example, if the steps in forming a cancer are mutations, we want to know whether they are induced by environmental mutagens or arise as spontaneous errors during the replication of DNA. The distinction can be made by observing what happens to cancer incidence when people migrate from one country to another. Many populations have been studied; they all show that environment plays a decisive part. For example, cancer of the stomach is much commoner in Japan than it is in the U.S., but cancer of the large intestine, the breast and the prostate are much less common. When Japanese emigrate to the U.S., these differences are lost within a generation or two [*see top illustration on page* 77]. Because the Japanese immigrants and their children tend to marry within the group, the change in incidence must be caused by the changed environment rather than by genetic factors; moreover, since the incidence of the various cancers takes more than one generation to reach levels typical of the U.S., some of the causative agents must be factors such as diet, which tend to persist as part of a cultural heritage, rather than factors such as air pollution, which tend to be the same for everyone in a given place. Similarly, Jews who migrate to Israel from Europe or the U.S. have an incidence of cancer that is typical of their country of origin, but their children born in Israel have a much lower incidence of almost all kinds of cancer. In this respect they have become more like the indigenous Jewish and Arab populations and the Jewish immigrants from Asia and Africa [*see bottom illustration on page* 77].

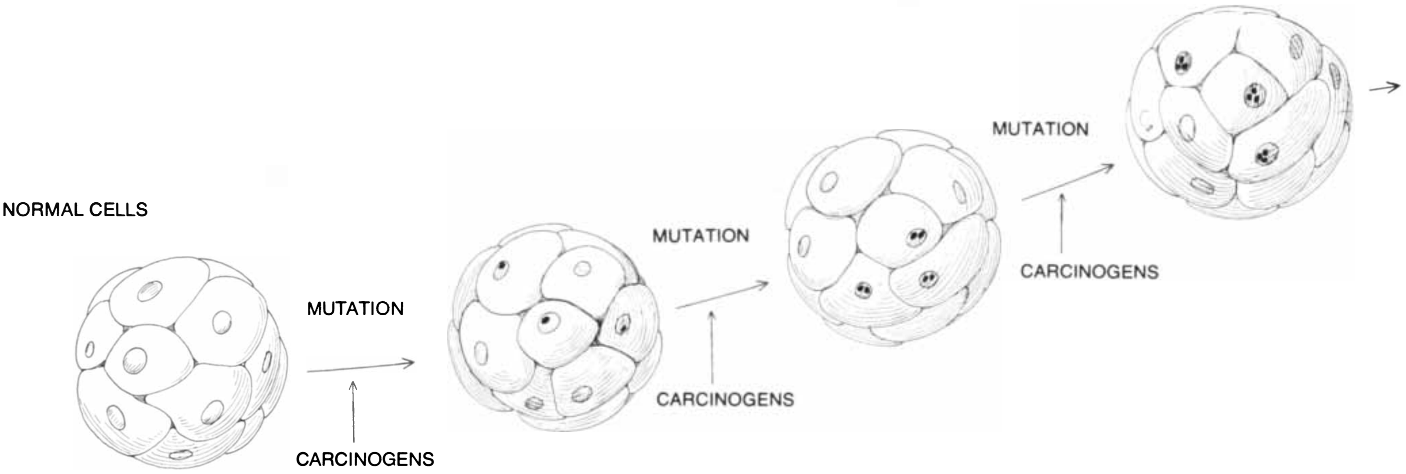
Even within a single country it is possible to detect the influence of local variation in environment and circumstances. In the U.S. the death rate from most of the common cancers is much lower in college graduates than in nongraduates; the exceptions are cancer of the breast and the prostate, which are commoner in graduates. If you are fair-skinned, your chance of dying of skin cancer is greater if you live in one of the Southern states; if you live in a Rocky Mountain state, your chance of dying of any form of cancer is much less than the national average. These differences in death rates within the U.S. are generally less than twofold, but even that limited range implies that a substantial proportion of cancer deaths could be prevented by controlling appropriate constituents of the environment.

Incidence and Causation

By demonstrating the influence of environment we have not, of course, excluded the possibility of a genetic contribution to carcinogenesis. We should like to know if genetic factors are important in the formation of any of the common cancers, not least because screening programs to achieve early diagnosis would be much less expensive if a particularly susceptible subsection of the population could be identified in advance. Certain rare inherited diseases are known to be associated with a greatly increased risk of some kinds of cancer; for example, an inherited defect in the enzymes that repair DNA damaged by ultraviolet light, called xeroderma pigmentosum, leads to multiple skin cancers. Some rare cancers in children seem to be caused by inherited mutations. Even collectively, however, these and other obviously familial types of cancer are too rare to contribute much to the overall cancer problem.

It is not easy to determine whether or not the frequency of the common cancers is significantly influenced by genetic factors. Close relatives are likely to share the same environment and therefore should show some tendency toward the same kinds of cancer even in the absence of a genetic contribution. Of the common cancers the one with the strongest propensity to run in families is cancer of the breast, which is found at about twice the normal frequency in the close relatives of patients with breast cancer. Such effects are less obvious for the other common cancers. The best evidence for the inheritance of traits that influence cancer would be the demonstration that identical twins were more likely to have the same kind of cancer than nonidentical twins of the same sex. But, to take a single example, a registry of the twins born in Denmark since 1870 has not sufficed to establish even that simple fact. Genetic constitution is therefore probably not an important variable, at least when considering the cancer problem on a national scale.

The ultimate object of epidemiological studies is the prevention of cancer by identifying its causes and removing them. It is an exercise that runs into a great variety of problems, as is evident when we consider what has been deduced about the causes of four types of cancer: carcinoma of the lung, the stom-



THEORY OF CARCINOGENESIS states that cells are restrained from becoming cancerous by several independent genes and that tumors develop only when mutations accumulate in all those genes within a single line of cells. The mutations (*black dots within*

ach and the large intestine and the various forms of leukemia and lymphoma.

Cancer of the lung is a disease of the 20th century [*see illustration on page 72*]. At first it involved only men, but recently it is being seen in women as well. In the U.S. it accounts for about a third of all cancer deaths in men and in England for roughly a half. From the start the most likely cause was thought to be cigarette smoking; after all, that was the new form of atmospheric pollution to which men were exposed and women initially were not. This explanation, however, encountered difficulties. In particular the incidence of lung cancer in different countries was not simply related to their per capita consumption of cigarettes. Most of the difficulties were resolved once it was realized that the incubation period was very long. There are still many unresolved questions, but the basic fact is no longer in doubt. If you smoke cigarettes, you increase your risk of dying of lung cancer tenfold to fiftyfold, the exact value depending on how much you smoke, the country you live in and various other factors. If many members of a group give up smoking, the mortality from lung cancer for the group as a whole will decline. There is every reason to believe that the abolition of cigarette smoking would largely eliminate lung cancer, the commonest of all forms of death from cancer. So far, however, there is no sign that smoking will be abolished. The professional classes smoke less than they once did, but the poor smoke more. It could even be argued that few Western societies could afford to abolish a habit that creates a large secondary industry, generates considerable revenue and kills mostly the older members of the population, who otherwise would draw on government welfare and social security benefits.

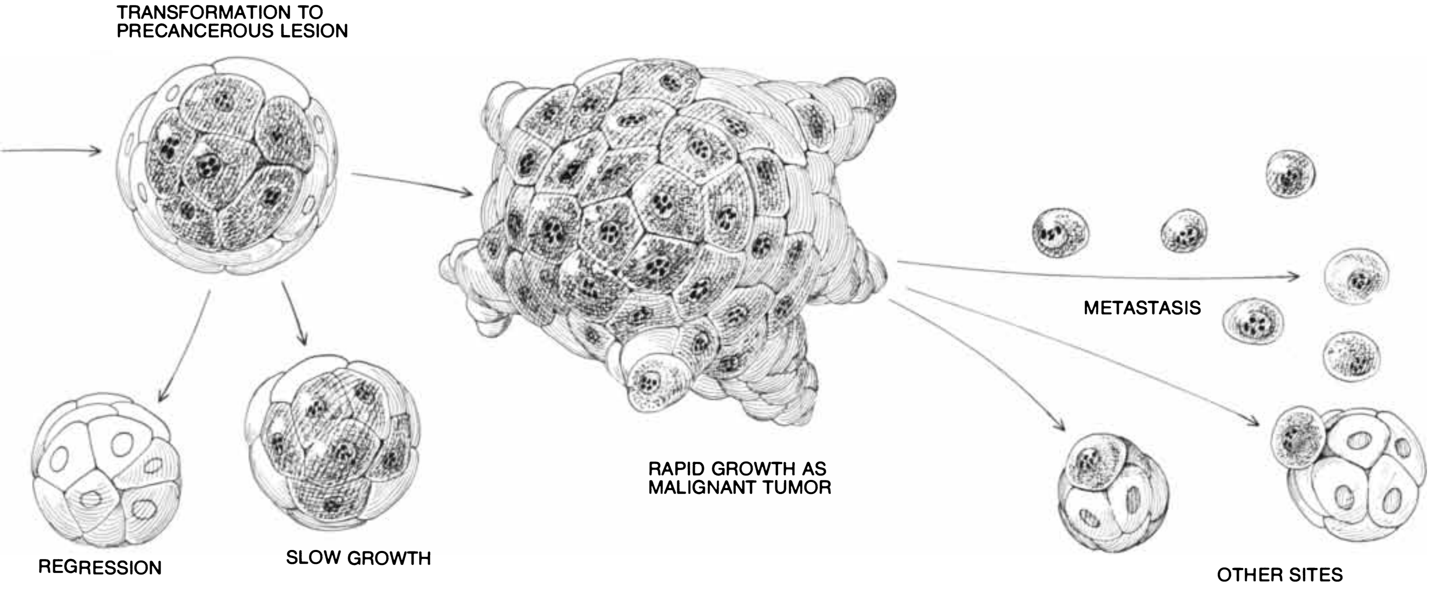
The incidence of cancer of the stomach has changed almost as markedly as that of cancer of the lung, but in the opposite direction. In the U.S. deaths from cancer of the stomach have decreased almost eightfold in the past 50 years. From this we can conclude that a single factor, or a group of closely related factors, must have been responsible for most of the cases observed in the past, simply because if there had been many unrelated causes, they would hardly all have declined at the same time. Stomach cancer has thus actually been prevented, albeit by chance.

Cancer of the large intestine has shown no great change in incidence with time, but it does vary greatly from one country to another, and that variation should give us some clue to its cause. Generally the richer the country the higher the incidence. In seeking the cause it has seemed natural to examine the diet, just as in explaining lung cancer it seemed reasonable to look for an inhaled carcinogen. The most likely causative agent is a high level of meat in the diet or alternatively a low intake of cereals [*see illustration on page 78*]. One proposed mechanism for the development of cancer of the large intestine suggests that normal intestinal bacteria convert various components of bile into carcinogens. The conversion might be much more extensive with low-residue diets, which are known to retard the transit of the intestinal contents.

Cancers of the lung, stomach and large intestine show how it is possible to deduce something about the cause of a cancer from the way it varies in incidence from one year to the next or from one country to another. A fourth and last example consists of a heterogeneous group of cancers involving the various cells of the immune system; they include the several varieties of leukemia and certain cancers of the lymph nodes, of which Hodgkin's disease is one of the commonest. Individually they are fairly rare, but they gain in importance because they often affect children and young adults. Recently they have received much attention because it is thought that they may be caused by viruses.

Viral Theory

The idea that viruses may be involved in carcinogenesis springs from the fact that certain leukemia-like diseases of chickens, cats and inbred strains of mice can be induced in young animals by inoculating them with viruses isolated from leukemic animals. The thought that human leukemia, and perhaps some other human cancers, may similarly be caused by viruses is attractive for two reasons. First, the induction of cancer by a virus is an event more amenable to investigation by the techniques of molecular biology than carcinogenesis by chemical mutagens. Second, there is the hope that once a virus is established as the cause of some human cancer it might be



***cell nuclei*) are seldom spontaneous but are apparently caused by carcinogenic factors in the environment. Once a precancerous lesion has formed, it must in many cases regress or grow very slowly; only a few lesions progress to an invasive, metastasizing tumor.**

a short step to developing a vaccine and so preventing the cancer. Unfortunately the relation of most of the animal leukemia viruses to their hosts is exceedingly complex. Many of the viruses are transmitted "vertically" by inheritance rather than "horizontally" from one animal to another, as the familiar pathogenic viruses are. Furthermore, even in the presence of the virus the development of leukemia in the adult animal requires some separate precipitating event. For example, the virus associated with leukemia in mice is inherited by many wild mice and is present in many inbred strains of mice, but leukemia is probably very rare in the wild, and it can be prevented in the inbred strains by a slight restriction in the diet; thus it seems to be precipitated as much by dietary factors as by the presence of the virus. In those cases where the virus is acquired by horizontal transmission it is only the rare animal that develops leukemia; for example, the leukemia virus of cats spreads horizontally, but apparently it causes leukemia only if the cat is infected with an unusually large dose of virus and then fails to have the normal immune response. The precipitating cause may therefore be some other event that depresses the immune system.

The quest for human cancer viruses has been conducted at several levels. Investigators have searched for viruses regularly associated with particular cancers; they have looked for familial clustering in diseases such as childhood leukemia, which might indicate the inheritance of a leukemia virus, and they have looked for spatial and temporal clustering of cases of leukemia or other similar diseases that might indicate the horizontal transmission of an infective agent. As yet there is no unambiguous evidence that any class of human cancers is regularly caused by a virus. Some cancers are often associated with elevated levels of antibody to certain viruses, some cancer cells have been shown to contain viral nucleic acid and certain cancers of lymph nodes (Hodgkin's disease and a rare cancer of children in tropical Africa called Burkitt's lymphoma) occasionally arise as clusters of cases. In each instance, however, some complicating factor makes interpretation difficult. We must assume that a genuine human cancer virus will eventually be found, simply because cancer viruses are known to exist in animals. It is important to remember, however, that with few exceptions viruses produce cancers in animals only if they are administered to very young animals and in particular combinations, whereas any animal of any age will produce a cancer if given the right chemical carcinogen by almost any route.

This brief review of the epidemiology of human cancer is intended to show what we know and what we hope to find out about the causes of cancer. To emphasize that the object is the practical one of preventing the loss of life from cancer, the examples were chosen from among the commonest cancers in Western society, and that choice has tended to emphasize the importance of diet and of habits such as smoking. The populations of the world each suffer their own group of cancers. If by the appropriate public-health measures the incidence of each kind of cancer could be reduced to the lowest level observed anywhere in the world, the overall incidence of cancer would be reduced at least tenfold. That is roughly equivalent to the reduction in mortality from infectious diseases that has been achieved in the past 50 years.

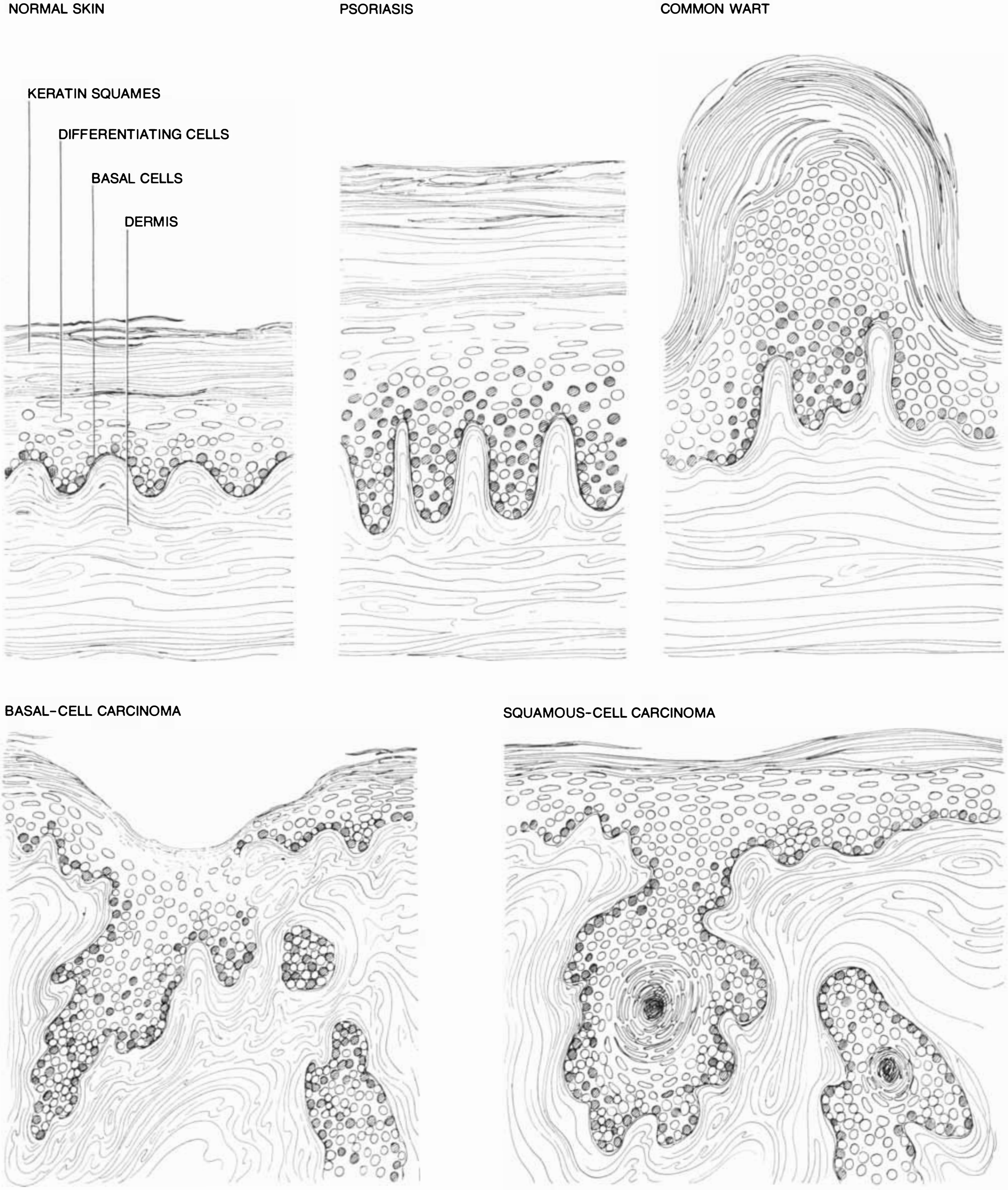
Mechanism of Carcinogenesis

Epidemiology bears on the question of causes and therefore on the prospects for prevention. One can also approach the cancer problem through the question of mechanism and the prospects for a cure. At the moment there is no effective general cure and no sign that one is about to be discovered. A few cancers (including Hodgkin's disease and Burkitt's lymphoma) can often be cured by a combination of cytotoxic drugs, and the growth of certain cancers of organs sensitive to sex hormones can be slowed by administration of hormones. For the vast majority of cancers, however, there is no specific drug and so they are treated by whatever combination of surgery, radiation and cytotoxic drugs has been found empirically to give the best results. The results are not very good. Fewer than half of all cancer patients survive five years from the time cancer is first diagnosed. Death is almost always caused by metastasis: the spread of the cancer to distant sites. If it were not for such spreading, few cancers would be beyond the reach of modern surgery; indeed, many benign tumors grow rapidly to great size and yet are rarely fatal. When considered as a cause of mortality, therefore, cancer is not so much an abnormality of growth control per se as it is a defect in the mechanism that normally sets the territorial limits of the cell. Because cancer is predominantly a disease of the epithelia, the biology of cancer can best be discussed in terms of the controlling systems that determine the form and establish the territorial limits of epithelial cells. For this purpose the skin serves as a convenient organ in which to compare the behavior of normal cells and cancer cells.

The epithelium of the skin is called the epidermis; it forms a sheet, usually from five to 10 cells deep, overlying a loosely knit layer of supporting cells, the dermis. The entire epidermis continuously replaces itself through the division of the cells in its deepest, or basal, layer, next to the dermis. As a result of this constant cell division cells are continuously squeezed out of the basal layer into more superficial layers. There they begin to differentiate according to an established program: they become flattened, begin to synthesize the insoluble protein keratin and lose their nucleus. Finally they fuse into the flakes called squames, which are eventually shed from the surface. The result of this program of development is that we are separated from our immediate environment by a relatively impenetrable layer of insoluble keratin that is continuously shed and replaced.

From the behavior of the epidermal cells we can deduce that they must be subject to several kinds of control. First, the fact that the only cells that divide are those in contact with the underlying dermis suggests that some short-range signals pass between the dermis and the basal cells; in the absence of these signals an epidermal cell stops multiplying and starts differentiating. Second, in order to prevent the multiplying basal cells from invading the dermis some mechanism must establish and enforce the boundary between the two layers. Third, some system of lateral signals must regulate the spacing of epidermal structures such as hair follicles and sweat glands.

In addition to local regulation of epidermal growth various overriding systems of control can be perceived in the behavior of the cells. Although surface characteristics such as fingerprints are expressed by the epidermis, they are determined by the dermis; if epidermal tissue is removed from the thigh, for example, and grafted onto the palm, it will thicken and take on the pattern of lines characteristic of the palm. If an area of skin is subjected to increased wear, the program of differentiation is somehow modified to increase the depth of the cells and thicken the keratin layer, forming a callus. If an area is denuded of epidermis, the area is recolonized through an increase in the rate of cell division in the surrounding epidermis. If a piece of



SKIN DISEASES that involve abnormalities of growth control include both cancers and other diseases considered much less consequential. In normal skin the outermost layer, the epidermis, is made up of a single sheet of basal cells overlain by cells differentiating to form squames, or flat scales, of the protein keratin. Only the cells in the basal layer are capable of division (*shading*). In psoriasis the number of dividing basal cells is increased, so that cells are produced faster than they can be shed from the surface; as a result the epidermis thickens. In the common wart the process of differentiation is slowed, so that cells accumulate at each stage in their life cycle, again causing thickening. Both conditions are noncancerous, and in both the overall organization of the skin is normal; in particular the boundary between the dermis and the epidermis remains well defined. In the skin cancers the organization of the tissues is disrupted and cells escape their normal territorial limits. Basal-cell carcinoma consists of basal cells that continue to divide instead of differentiating and invade the dermis. Squamous-cell carcinoma is made up of cells that differentiate almost normally, but at locations outside epidermis. All four diseases seem to be caused by failures of communication between cells.

skin is implanted into subcutaneous tissue, the epidermis degenerates.

The means by which all these controls are effected are not known. They necessarily involve communication between the cells; that might be accomplished through a concentration gradient of a freely diffusible substance secreted by some cells and detected by others, or it might require direct contact between the cells. Whatever the mechanism is, communication could be interrupted by a defect in the signaling cell or in the recipient cell. As there seem to be many signaling systems in the skin, each operating more or less independently, there should be many distinct disorders of growth control. That is in fact exactly what we find.

Two well-known skin diseases that represent noncancerous abnormalities of growth control are psoriasis and the common wart. In psoriasis the number of multiplying basal cells increases, so that the basal-cell layer becomes about 10 cells thick. Psoriasis seems to be caused by a failure of communication between the dermis and the basal cells. The common wart, which is caused by a virus infection, is a local thickening of all layers of the epidermis. In warts the differentiation of epidermal cells appears to be drastically slowed, so that more cells are present at each level of differentiation. The overall arrangement of the cells remains precisely ordered, however, and the boundary with the dermis is unchanged.

There are two common cancers of the epidermis. The basal-cell carcinoma is made up of cells derived from the basal layer that seem to have escaped the control of the system that normally preserves the boundary between the dermis and the epidermis. The cancerous cells invade the dermis and the underlying tissues, forming an irregular, erosive ulcer; this form of skin cancer is sometimes called rodent ulcer. In spite of its great powers of local invasion a basal-cell carcinoma virtually never metastasizes, suggesting that the cancerous cells still require signals from the dermis in order to multiply.

The second skin cancer, squamous-cell carcinoma, also consists of disordered groups of cells, but unlike those of the basal-cell cancer they undergo almost normal differentiation into squames of keratin. Squamous-cell carcinoma is less invasive locally, but it occasionally gives rise to distant tumors. The cells thus seem to retain some of the cohesive properties of normal epidermal tissue, but they are less dependent on signals from the dermis.

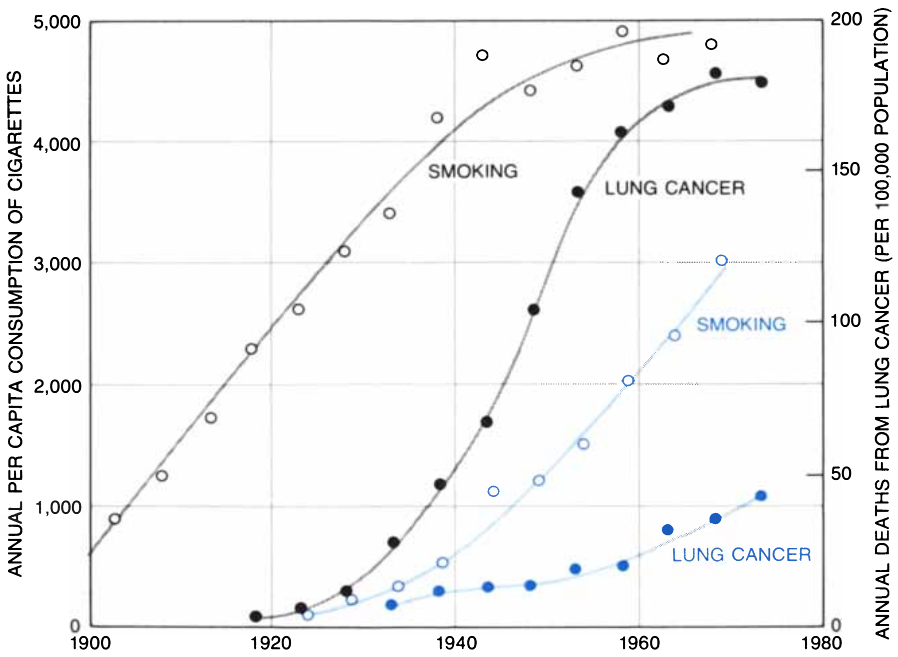
There are many other disorders of growth in the epidermis, reflecting the variety of ways in which the regulatory systems can malfunction. Significantly, it is only when the cells are freed from the constraints of territoriality that they can form a potentially lethal cancer. Furthermore, as the two common skin cancers illustrate, territoriality can be lost in more than one way, and each kind of loss will have a distinctive effect on the behavior of the cells. Cancer therefore cannot be considered a single disease, brought on in every case by the same cellular malfunction. Even though basal-cell and squamous-cell carcinomas both arise in the epidermis, and even though both are caused by sunlight, the cells involved behave differently.

The Immune Response

We cannot expect to understand the behavior of a cancer until we can comprehend the signals passing between cells and the forces that separate groups of dissimilar cells in multicellular organisms. Although little is known about these systems of intercellular communication, it is apparent that the molecules on the cell surface must determine what signals the cell can receive and how the cell interacts with its neighbors. Some abnormalities of cancerous cells should therefore be expressed on their surface.

On the surface of every human cell are molecules called histocompatibility antigens that distinguish it from the cells of other individuals. It is part of the duty of our immune system to destroy cells that bear the wrong histocompatibility antigens. It is natural to wonder whether some means might be found for mobilizing the immune system against any abnormal components that may be present on the surface of the cancer cell. Indeed, since the discovery of cellular immunity at the beginning of the century the suggestion has often been made that one of the main functions of the immune system is to destroy cancers and that the appearance of cancer in old age shows simply that the immune system is failing.

Like the theory that cancers are caused by viruses, that idea is attractive because it offers the hope of a cure, in this case by some form of immunotherapy, without having to wait for further advances in basic biology. Unfortunately the facts are against it. Patients treated with immunodepressants so that they can tolerate an organ transplant and people with inherited defects of the immune system show a greatly increased incidence of certain rare cancers of cells of the immune system, but they do not show an increase in the death rate from any of the common cancers. When very old people are examined for their reactivity to appropriate test antigens, the



CIGARETTE SMOKING AND LUNG CANCER are unmistakably related, but the nature of the relation remained obscure because of the long latent period between the increase in cigarette consumption and the increase in the incidence of lung cancer. The data are for England and Wales. In men (*black*) smoking began to increase at the beginning of the 20th century, but the corresponding trend in deaths from lung cancer did not begin until after 1920. In women (*color*) smoking began later, and lung cancers are only now appearing.

See your travel agent or write to: **Québec Tourism** dept. B52J1, Québec City, Canada G1R 4Y3. **In U.S.** Dept. B52J1, 17 West 50th Street, New York 10020

You've read about them. Now look at them. The available luxury buckets. Did we exaggerate?

Pontiac's new available 5-speed. Fifth acts as an overdrive at highway speeds.

V-6. It's available if you want 231 cubic inches going for you.

and workmanship.

This guarantee is in addition to the New Vehicle warranty but does not apply to repairs required because of misuse, negligence, alteration, accident or lack of reasonable or proper maintenance.

This 5-year/60,000-mile engine guarantee is an added value feature in your 1976 Sunbird. Just try to find an import that can match it!

Not many imports are going to match Sunbird's interior, either. Order the available luxury buckets and it's downright elegant.

Here's the best part. Sunbird's priced right. That's an important part of being a great little car. And the Wide-Track people just won't build anything else.

PONTIAC The Mark of Great Cars.

A remarkable new device is saving two kinds of energy—elbow grease and fuel oil.

Our photograph shows you a new invention in action—a scrubbing machine, named SCAMP™. It's solving an age-old problem for Exxon and other companies who operate large ocean-going ships.

Keeping any ship's hull free of barnacles and other marine growth has always been a headache for sailors. When the growth builds up, it slows a ship down. This wastes fuel. And fuel today is valuable energy.

The traditional way to get rid of marine growth is to haul your sailboat or ship out of the water and start scraping. But that's not so easy when your ship is a tanker that stretches the length of three football fields. Dry-docking this size ship costs you time and money.

Thanks to SCAMP equipment, that's no longer necessary. An affiliate of Exxon developed it to clean the hulls of big ships—while the ships are at anchor, loading or unloading.

This new device is basically remote-controlled. As it travels underwater along the hull of a ship, its three rotating brushes whisk away marine growth. It's so fast it can clean the entire hull of a supertanker in just half a day.

In today's energy-tight world, SCAMP is performing a vital service. It is making the movement of crude oil around the world more efficient and more economical.

A clean hull can reduce the fuel consumption of a large tanker. For example, in a 10,000-mile voyage, a 1,100-foot tanker could save 60,000 gallons of fuel oil.

And that's 60,000 gallons of valuable energy that can be made available for other good uses.

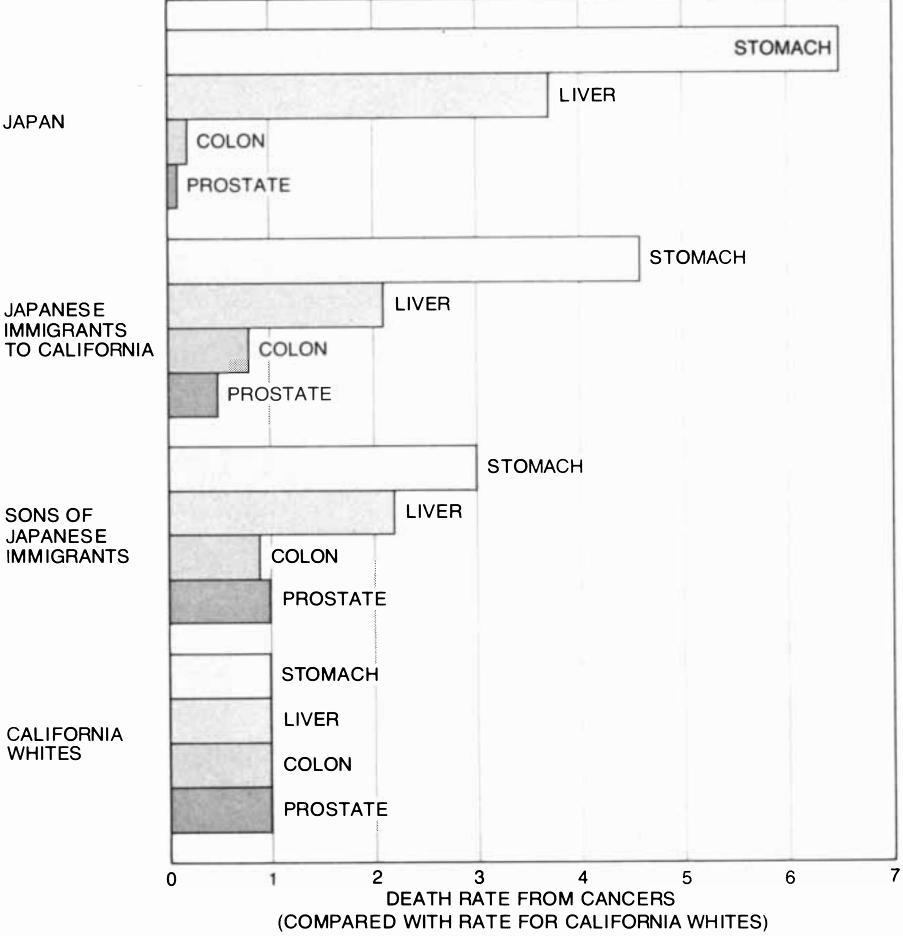
EXXON

annual death rate is found to be higher in those who showed depressed reactivity, but the deaths are not caused by cancers. Similarly, mice with various induced or inherited defects of their immune system may show increased susceptibility to certain tumor viruses, but they are not more susceptible to chemical carcinogens, nor do they show any increase in spontaneous cancers. On the other hand, there is no doubt that the cells of many experimental cancers of animals and certain cancers in man do display abnormal surface components that evoke a limited response from the immune system. The alternative suggestion has therefore been made that a mild immune reaction by the host stimulates the multiplication of cancer cells; in other words, there may be some selection, during the growth of a cancer, for variant cells with novel surface components that provoke an immune response. That, of course, is exactly the opposite of the original idea that the immune system inhibits the development of cancers. Nevertheless, it is not entirely ruled out that some method of immunotherapy could eventually be directed against whatever surface abnormalities of the cancer cell are responsible for its disorganized, invasive behavior.

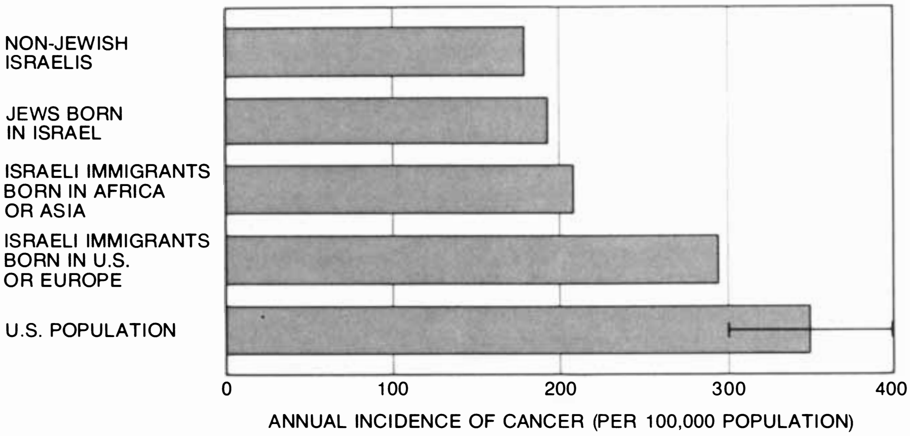
Early Detection

A last aspect of the cancer problem that must be discussed concerns once again the time course of the disease. The most important question from the point of view of the cancer patient is whether his cancer has been detected and excised before it has begun to metastasize. Great efforts are being made to develop screening programs that will detect the common cancers, in particular cancer of the breast and the cervix, very early in their development. Unfortunately these programs are likely to have only limited success.

In order for a cancer to form secondary growths in a distant organ its cells not only must enter the bloodstream or the lymphatic system by direct invasion from the primary site but also must survive and multiply in an alien environment. There is reason to believe the latter capability is often the factor limiting the spread of a tumor. Cancers that are undergoing metastasis do not spread randomly. Cancers of the lung tend to colonize the brain and the adrenal glands; cancer of the prostate commonly spreads to bone; one variety of cancer of the stomach spreads almost exclusively to the ovaries. In other words,



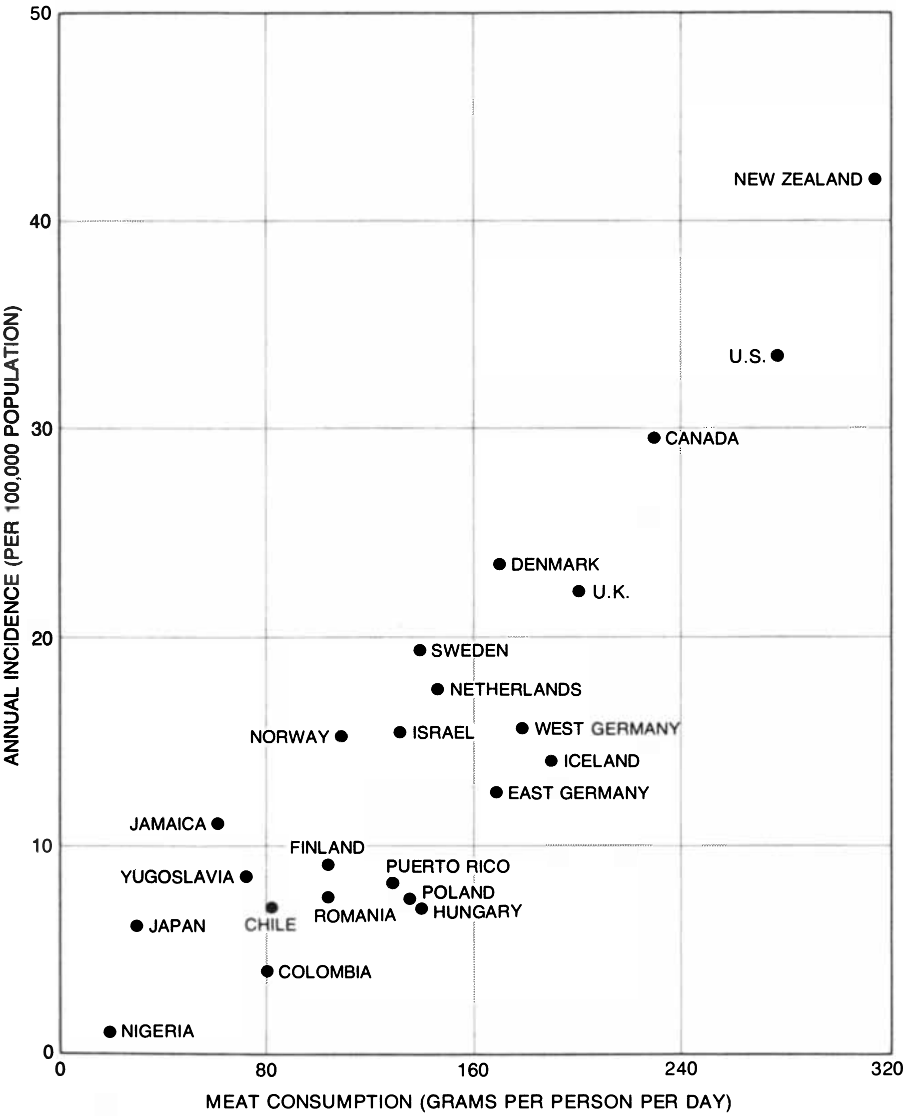
CHANGE IN INCIDENCE of various cancers with migration from Japan to the U.S. provides evidence that the cancers are caused by components of the environment that differ in the two countries. The incidence of each kind of cancer is expressed as the ratio of the death rate in the population being considered to that in a hypothetical population of California whites with the same age distribution; the death rates for whites are thus defined as 1. The death rates among immigrants and immigrants' sons tend consistently toward California norms, but the change requires more than a generation, suggesting that causative agents are factors influenced by culture rather than hazards to which all are exposed equally.



ENVIRONMENTAL FACTORS also explain differences between the overall incidence of cancer in Israel and that in the U.S. The incidence in Israel is relatively low, and it is comparable for the Jewish and non-Jewish (mainly Arab) populations; the incidence in the U.S., although it varies significantly from one region to another, is almost twice as great. The high incidence among Israeli immigrants from the U.S. and Europe probably results from the long latency period of cancers: people exposed to carcinogens in youth may later develop cancers even if they have moved to an environment that has lower levels of the carcinogens.

the ability to grow in foreign sites is not a universal property of cancers. Some, like the basal-cell carcinoma of skin and most cancers arising in the brain, almost never metastasize; others can produce multiple metastases at a stage where the primary cancer is still too small to be detected. It is therefore a misconception to think that the natural history of every cancer consists in growth to a certain size followed by inevitable metastasis. When we consider screening programs for a class of cancers, we should expect to find that some will not spread if they are left untreated for a long time and others will be lethal even when they are detected very early. Hence it will be only for a limited, intermediate group that screening will bring any benefit. To take the most clear-cut example, a carefully controlled screening program in New York has apparently reduced mortality from cancer of the breast by about 30 percent in the group being examined. The death rate seems to have been reduced only for women over the age of 50, however, suggesting that all the breast cancers of younger women that are capable of metastasis spread too early to be intercepted. As a result of screening 31,000 women annually for five years, the number of lives saved is thought to have been about 23, which is probably no more than the number of women in the screened group who died of lung cancer during the same period and who might have been saved by ensuring that none of the women continued smoking cigarettes. Two other programs that have had some success involve screening for cancer of the cervix and, in Japan, for cancer of the stomach. In contrast, an attempt to reduce the high mortality from lung cancer by frequent chest X rays has been a total failure.

Apart from their limited effectiveness, screening programs have an additional shortcoming: for every cancer they detect they reveal perhaps 10 other abnormalities, many of which seem to be precancerous. As long as it is impossible to tell which of these will progress to cancer they must all be treated, which presents an economic problem. It would be beyond the resources of the U.S., let alone any other country, to treat all the precancerous lesions that could be detected with even the limited screening procedures available today if these procedures were applied to the entire population. The survey that found skin cancers in 4 percent of the adult population of rural Tennessee found precancerous skin lesions in a sixth of the population; if all those lesions were to be treated by surgical excision, the cost to the state would be several million dollars, even though cancer of the skin is perhaps the least expensive form of cancer to detect and treat.

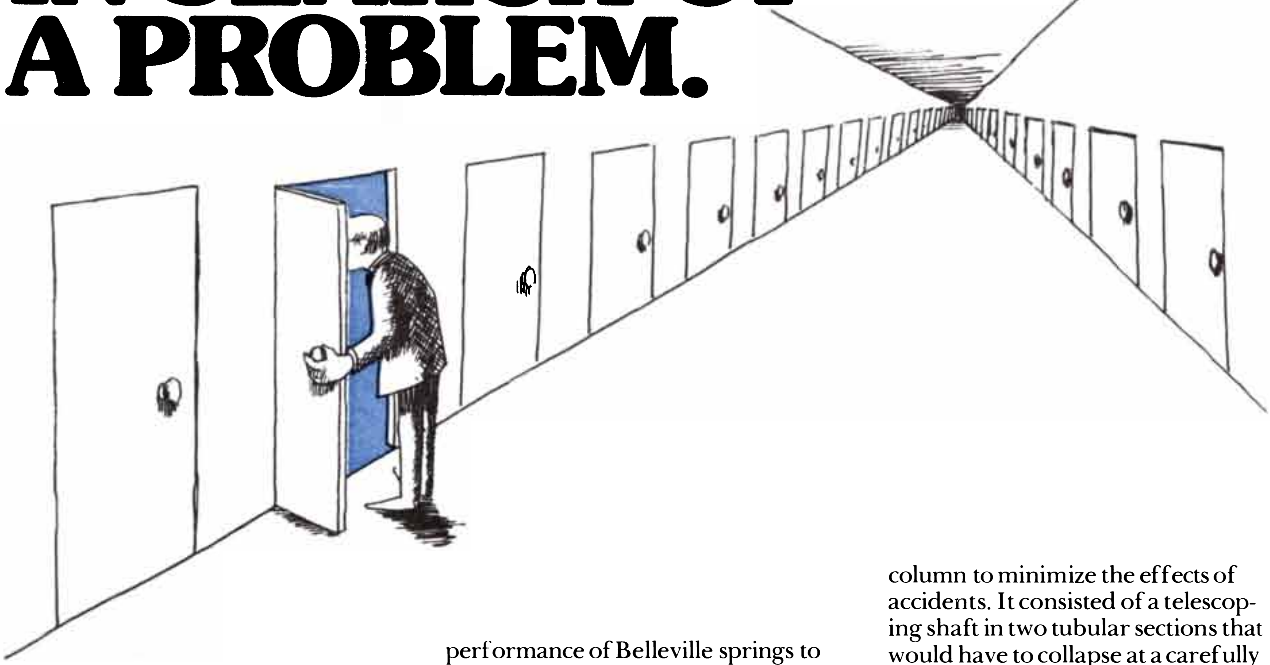


GEOGRAPHY OF A CANCER suggests a probable cause of the disease. The incidence of cancer of the large intestine among women in 23 countries is closely related to per capita meat consumption in those countries. The data are adjusted to eliminate differences in age distribution in the populations. An alternative explanation attributes cancer of the large intestine to a low consumption of cereals. The two hypotheses are hard to distinguish from each other because high meat consumption and low cereal consumption tend to go together.

Conclusions

The cancer problem is of immediate, urgent concern. We all want to know when we can expect to hear of the discovery of a cure for cancer and to be told what steps we can take in the meantime to reduce our chance of dying of cancer. I have tried to give a simple review of what is known about the nature of cancer in man and in particular what can be deduced about the prospects for its prevention or cure. Because the cancer problem arouses so much interest and attracts so much financial support, people engaged in cancer research are under pressure to announce some breakthrough in treatment or, failing that, at least some major advance in our understanding of the disease. As a result it has become unfashionable to suggest that the hoped-for cure may have to wait until we have learned much more about the interactions of cells in simple systems. If this view is correct, however, we should in the meantime be trying to do more in the realm of preventive medicine. Since screening programs seem to be of limited use and too expensive, we are left with the prevention of cancer by seeking out and eradicating its causes. That should not be taken as a counsel of despair. After all, it was largely preventive medicine that eradicated the infectious diseases.

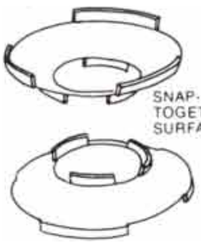
SOLUTIONS IN SEARCH OF A PROBLEM.



In a rational world, first comes the problem then the solution. Right? Right. . .but not always. Sometimes the solution comes first and the problem has to be searched out. A 100-year-old idea, for example, can be adapted to break through technological limitations imposed by older engineering materials. Which brings us to our story.

Merci, M. Belleville

Back in 1866 one Julien Belleville patented a coned metal disc spring that we now call, to no one's surprise, a Belleville spring. It has a convex saucer shape and a center hole. Under load it flattens and returns to its original shape when the pressure is off, even as you and I.

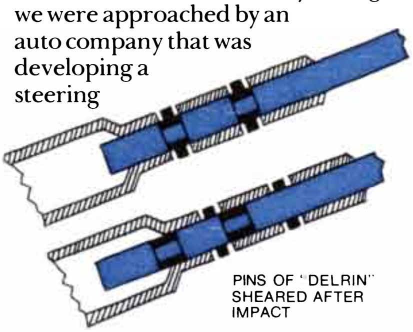


So, we've been working on Belleville springs made of an engineering plastic—Delrin® acetal resin. We've designed springs of "Delrin", molded them as individual discs and as large sheets with *snap-together surfaces.* And we've even extended the performance of Belleville springs to other geometries, thanks to another pretty good invention, the computer.

We've got a lot of answers—for possible use in door closer springs, springs in switches, relays and push buttons, furniture innersprings and, from one of our engineers who thinks young, pogo sticks. Maybe you have a new or even old idea that can expand in function and usefulness when it's matched with our engineering plastics technology.

Born to fail

That isn't the name of a soap opera. Rather it describes an unusual technique: planned failure. We've always known that "Delrin", because of its reproducible properties, could be designed to fail at a predetermined stress. Then several years ago we were approached by an auto company that was developing a steering column to minimize the effects of accidents. It consisted of a telescoping shaft in two tubular sections that would have to collapse at a carefully predetermined stress.



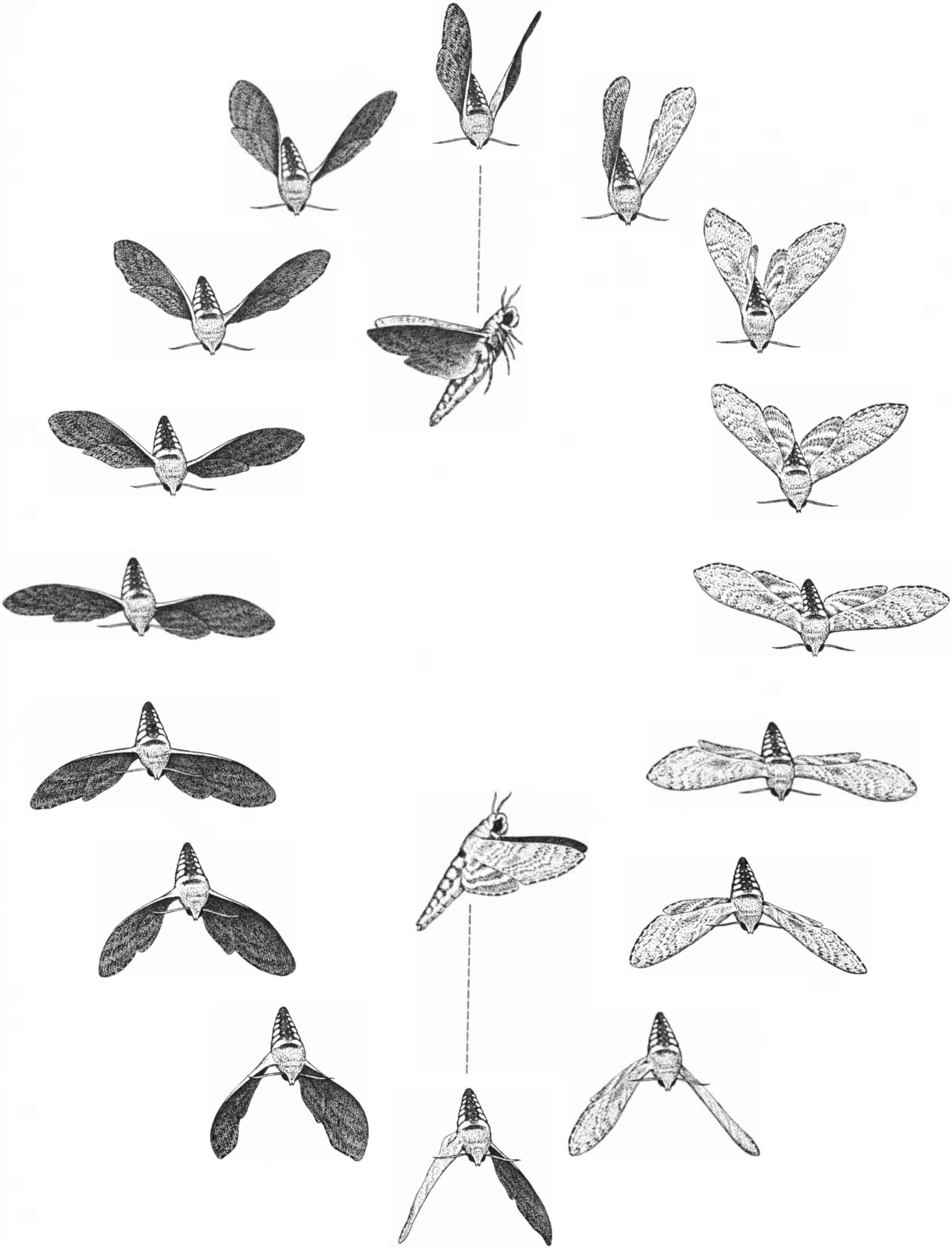
Together with the company we worked out the technology of injection molding "Delrin" into the tubular sections as part of the mold. The design depended on "Delrin" shear pins and bearing collars.

And so we have another apt illustration of a solution looking for a problem and in this case finding it.

Invitation to a Dialog

Surprising breaks with tradition can happen when you look at your problem and our engineering plastics technology with a new, unrestricted curiosity. If our expertise in the properties and uses of engineering plastics may be of help, you can continue this dialog by either writing Dick Johannes, Du Pont Co., Plastics Department, Rm. D-13064, Wilmington, DE 19898 on your company letterhead, or calling Dick at (302) 774-5826.

DU PONT



Unusual Mechanisms for the Generation of Lift in Flying Animals

Some birds and insects that hover cannot generate enough lift by means of standard aerodynamics. A close study of hovering shows that they employ two mechanisms that are best described as clap-fling and flip

by Torkel Weis-Fogh

Flying is a way of life for the great majority of the earth's animals: of the roughly one million living and extinct animal species known, some 750,000 are winged insects and a large number are birds and bats. Most of these creatures fly in what might be called the standard way, utilizing principles of aerodynamics that are well understood. Some of them, however, particularly certain small birds and insects that are able to hover in midair, perform in a way that cannot be explained by established principles of the action of airfoils. In studying such performances in my laboratory at the University of Cambridge I have found that they depend on mechanisms that do not seem to have been previously recognized. The best-understood example is the clap-fling mechanism for the generation of lift. Another example is the flip mechanism, although at present we know little about it.

The lift and propulsion of aircraft depend on the action of airfoils that move through the air steadily. Any nonsteady flows over an airfoil need to be kept to a minimum, since they reduce the efficiency of flight and generate unwanted and possibly dangerous vibrations. In contrast, the muscle-powered flight of birds, bats and insects depends on the flapping of the wings, which introduces a degree of nonsteady airflow.

Nonsteady aerodynamics is thus inherent in natural flapping flight. Its importance varies with, among other things, the size of the animal and its airspeed. The larger the animal is, the faster it flies and the smaller the frequency and amplitude of its wing stroke are, the less significant are the nonsteady effects. When birds and insects are in fast forward flight, the flows therefore approach steady conditions. The wing performance of such fliers can be understood essentially as a sequence of steady-state situations, that is, they can be understood on the basis of ordinary aerodynamics.

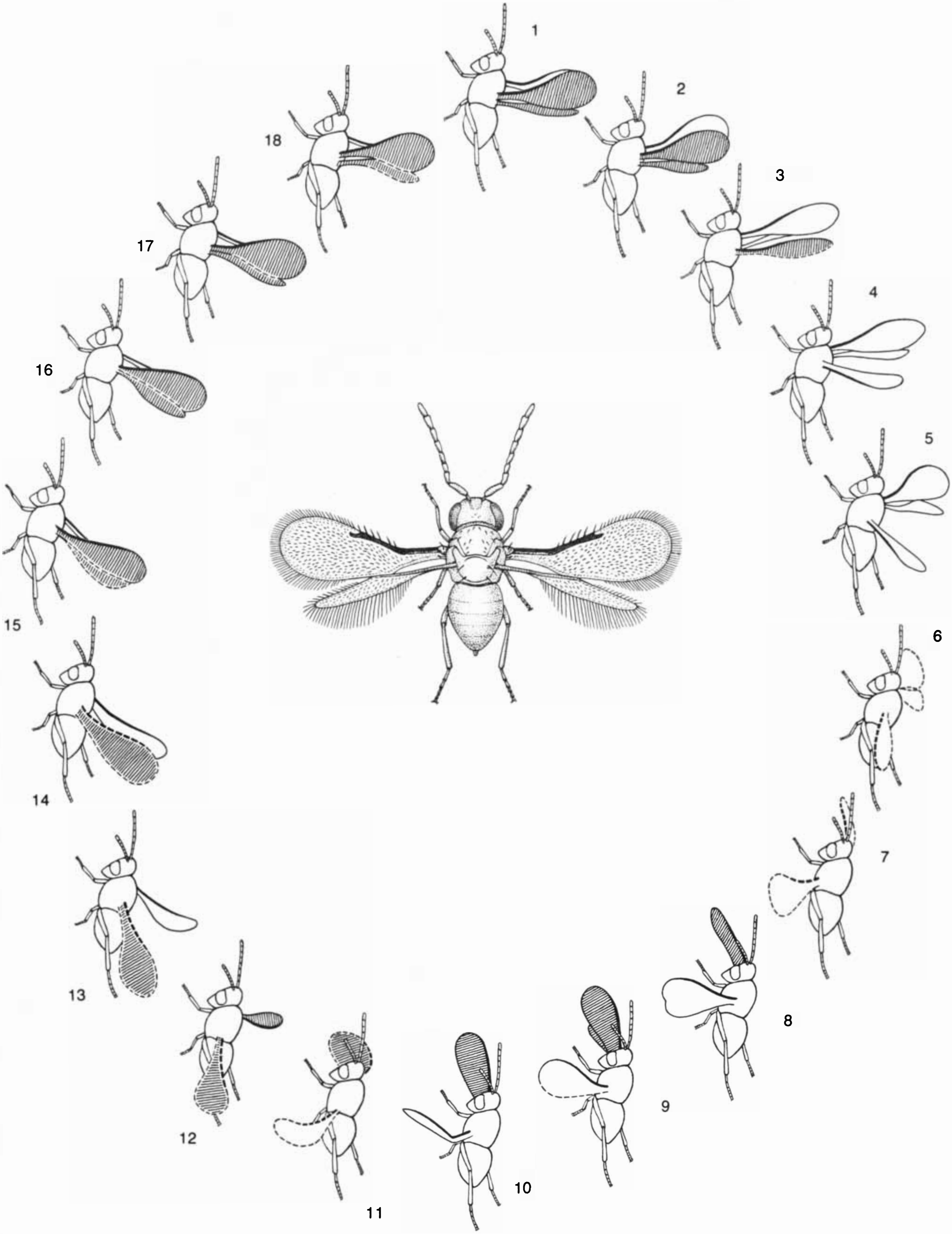
In this kind of flight, which is characteristic of animals ranging from swans to locusts and large flies, the wings act primarily as ordinary airfoils. Each section of wing hits the air at a relatively small angle of attack and generates a force that has a large component of lift perpendicular to the airflow and a much smaller component of drag in the direction of the flow. The principle was explained by the German aeronautical pioneer Otto Lilienthal as early as 1889 and can be understood from his famous analysis of the flying stork.

During the downstroke of the stork's wing the wing's leading edge twists increasingly downward moving outward from the base to the tip. This twisting action has the dual effect of adjusting the angle of attack with respect to the air and of tilting the resulting force forward so that the animal is provided with both propulsion and lift. During the upstroke the leading edge of the wing is twisted upward with respect to the horizontal, so that the angle of attack remains small. This action provides some lift but no propulsion; in fact, its contribution to propulsion is slightly negative.

The same principles apply to the locust and to other animals in fast forward flight. Such flight thus depends on a combination of wing flapping and wing twisting, both of which are much more pronounced in the locust than they are in the stork. It is immediately apparent that at each section of wing the resulting force oscillates widely in magnitude as well as in the three directions of space, which is another way of saying that flapping flight is nonsteady. The fluctuations increase in size and significance when the airspeed is reduced; they reach a maximum when the animal hovers.

Hovering flight is defined as flying without moving forward or backward with respect to a fixed spot on the ground (assuming that the air is still). Hovering flight is not confined to the familiar examples of hummingbirds, dragonflies and the two-winged insects called Syrphinae: the true hover flies. In fact, it is a very common mode of flight. Because the mechanical power for remaining airborne is provided entirely by the wing muscles and is limited to about 200 watts per kilogram of muscle, one can use simple momentum considerations to

HOVERING INSECT, the tobacco-hornworm moth *Manduca sexta*, is portrayed from above in one wing cycle, beginning at top where the moth rapidly twists its wings, which are moving almost horizontally, so that the front edge remains the leading edge during both the downstroke (*right*) and the upstroke (*left*). As in hummingbirds, this pattern is an example of normal hovering. The darkened part of the wings is the underside. Inside the circle one sees the moth from the right side at the beginning of the downstroke and the upstroke.



COMPLETE WING STROKE of a tiny hovering insect, the wasp *Encarsia formosa*, is traced on the basis of high-speed motion pictures. The cycle begins (*1*) and ends (*16–18*) with the wings in the "clapped" position. By Frame 2 the fling has begun, and by Frame 4 the lift produced by the horizontally sweeping wings already equals the weight of the insect. The underside of the wings is darkened. In several frames parts of the wings are represented by broken lines because the film did not reveal them with certainty.

calculate that no animal with a mass larger than about 100 grams (3½ ounces) is able to hover continuously, although pigeons and many other birds are able to take off vertically and to hover for brief periods of time.

Most birds from the size of sparrows downward hover when they approach their nest and when they are feeding and courting. Even certain small bats hover. Among flying insects the ability to hover is the rule rather than the exception. One can argue that this ability has opened up a multitude of small ecological niches in the aerial biosphere and therefore has itself given rise to many species. How, then, do these animals manage to hover?

The answer is not simple, because there is more than one kind of hovering and because the aerodynamic conditions under which the oscillating wings work change with the size of the animal. The relation is best expressed by the ratio between the inertial forces involved in accelerating a given volume of air and the internal viscous forces in that volume. (The ratio is named the Reynolds number after the British engineer and physicist Osborne Reynolds, who described the relation.) In a large hummingbird, for example the giant species *Patagona gigas* (mass 20 grams), the mean value of the Reynolds number for the flow past oscillating wings is 15,000; in a large wasp (.8 gram) it is 4,000; in the small fruit fly *Drosophila* (two milligrams) it is 200, and in the tiny parasitic wasp *Encarsia formosa* (25 micrograms) it is less than 20. (There are even smaller insects in which the Reynolds number approaches unity, but we know nothing about their flight at present.)

The range of Reynolds numbers is significant because ordinary airfoil action is possible at the high and intermediate values but deteriorates sharply at the lower end of the scale. One is therefore faced with the dual problem of how animals hover by means of oscillating wings and how the small insects remain airborne at all. In my laboratory we decided to study these questions by means of mathematical models combined with high-speed cinematography of freely hovering insects, large and small.

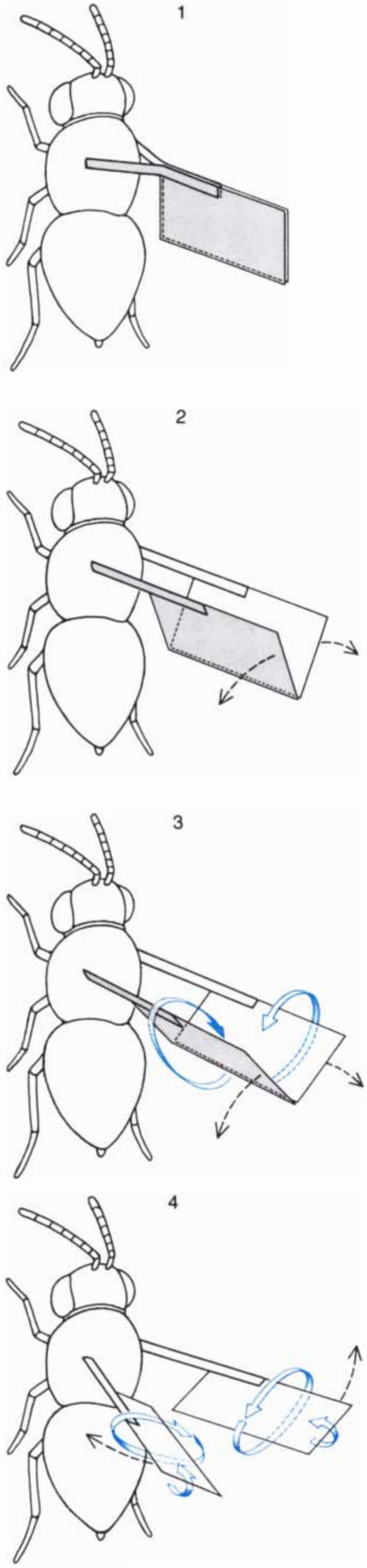
Before World War II, M. Stolpe and K. Zimmer in Germany employed slow-motion films to study hummingbirds feeding from an artificial flower. After the war Crawford H. Greenewalt in the U.S. trained hummingbirds to fly freely in a wind tunnel at airspeeds ranging from 40 kilometers per hour to zero forward speed, which was the equivalent of hovering. These studies showed that during fast forward flight the up-and-down movements of the hummingbird's wings correspond to those of the stork and the locust. The axis of the body is almost horizontal.

As the airspeed is reduced the axis of the hummingbird's body is gradually tilted toward the vertical. During true hovering the bird's wings beat in an almost horizontal figure eight. The angular movements of the wing's axis become regularly sinusoidal, so that the velocities of the wing sections with respect to the air are easy to calculate.

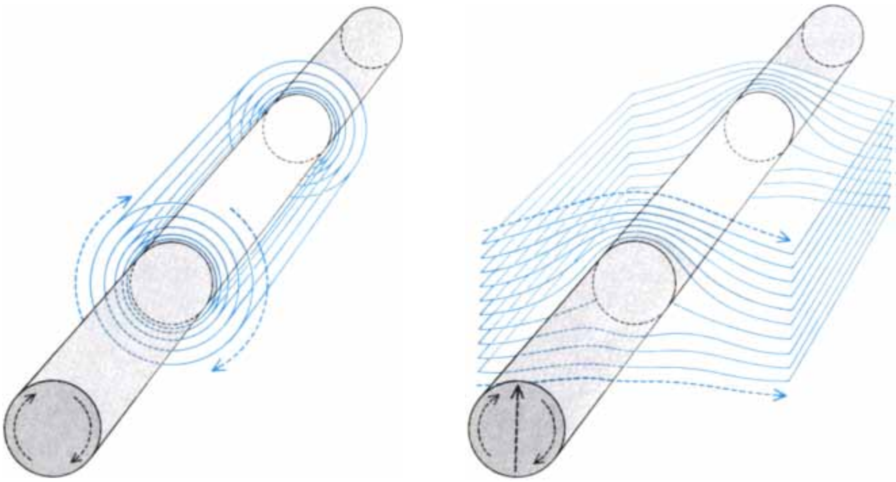
It was also observed that the two wings do not come close to each other and that each wing is twisted rapidly and extensively (by 130 degrees within two milliseconds) at the end of each half stroke. Therefore the angles of attack are large and positive with respect to the horizontal sweep during both phases of the stroke, so that the wings appear to act as ordinary airfoils or rotor blades and to generate vertical lifting forces during the major part of the cycle. At the end of each half stroke, however, the wing is virtually at rest and is influenced only by the downward current that the wing action induces in the still air. That current cannot give rise to any useful force against the pull of gravity.

In contrast to the rotor of a helicopter, the wing must therefore lose its lift near its extreme positions and build it up again after it has been twisted and then starts to move in the opposite direction. Both processes take a bit of time. They entail the formation and shedding of vortexes (a subject to which I shall return), and therefore they involve nonsteady aerodynamics.

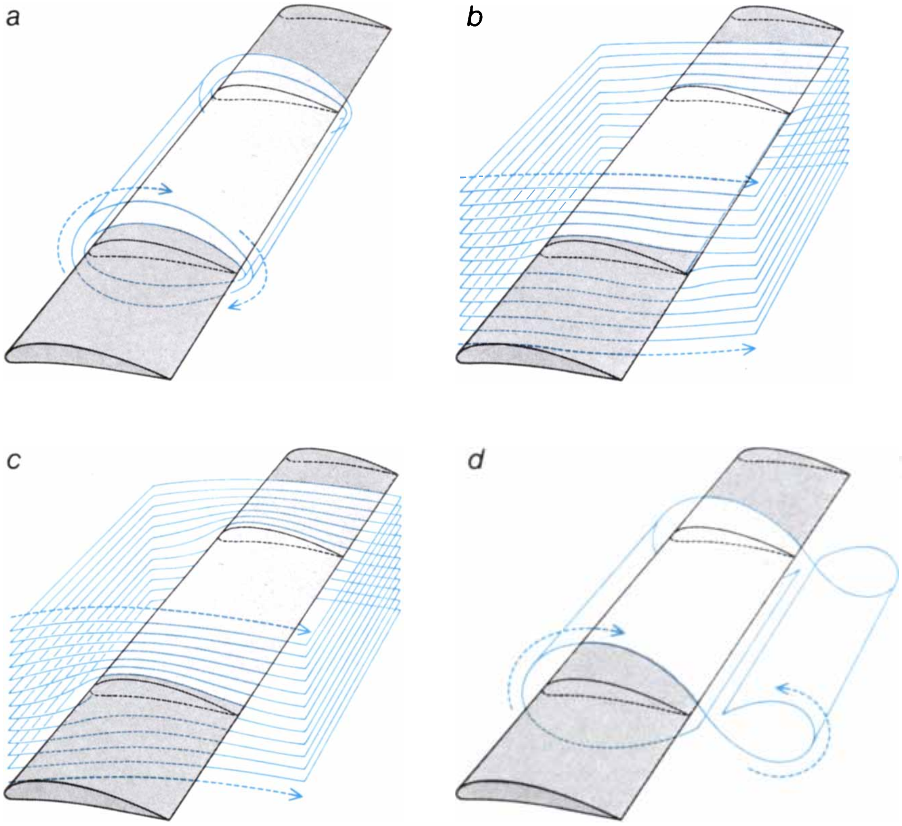
In spite of this limitation of oscillating wings (as opposed to rotating wings), my observations and computations of the average coefficient of lift in hovering



CLAP AND FLING are portrayed schematically on the basis of analysis of the hovering flight of the wasp *Encarsia formosa.* At *1* the two pairs of wings are held firmly together as a vertical plate and in a clapped position; the air is virtually at rest. At *2* the wings are flung open at high speed so that the widening triangular gap between them must be filled by air, which flows in (*3*) as indicated by the bands. Next (*4*) the right and left wing surfaces separate and begin their horizontal sweep, each side carrying the vortex of air formed during the fling. The small bands in *4* indicate tip vortexes.



FORMATION OF VORTEXES is explained with a cylinder that (*left*) spins in a clockwise direction in a fluid that is otherwise at rest with respect to the cylinder. The flow of the fluid, which in the case of a flying or hovering animal (an insect, a bird or a bat) would be air, is depicted in color. It is a circular and purely rotational flow caused by viscous forces in the fluid; the direction of the movement is shown by arrows. A streamline pattern of airflow is created by the spinning cylinder (*right*) after the air is set in a uniform horizontal motion from left to right, a condition that is the equivalent of flying. The result is a compression of the streamlines above the cylinder, a widening below it and a downward component behind it; changes give rise to vertical lifting force termed the Magnus effect.



FLOW PAST AIRFOIL indicates the forces involved with a wing of an animal in hovering flight. First (*a*) one sees only the purely rotational flow, which is shown as the bound vortex that a camera at rest with respect to the undisturbed fluid would record. An airfoil placed (*b*) in an ideal inviscid fluid moving from left to right experiences a purely potential flow that produces neither lift nor drag. When the rotational flow and the potential flow are combined (*c*), a real flow pattern results; the streamlines are compressed above the wing and widened below it, causing a vertical lift and also a downward component of flow behind the wing. If the wing starts to move from right to left (*d*), the first result is a potential flow, as in *b*, but viscous forces quickly begin to create a starting vortex at the trailing edge, so that a bound vortex is formed around the wing as the counterpart of the starting vortex.

hummingbirds, bats and insects [*see illustration on opposite page*] clearly indicate three results of importance for our problem. First, the majority of hoverers hover in essentially the same way as hummingbirds. Second, this "normal" hovering is characterized by coefficients of lift that do not exceed the values one might expect from the steady-state aerodynamics of real wings at the relevant Reynolds numbers. Finally, there are conspicuous exceptions in which the coefficients of lift seem much too large.

It turns out that the animals for which large coefficients of lift are recorded move their wings in a manner somewhat different from the normal one. Examples include the wasp *Encarsia formosa*, butterflies, dragonflies and true hover flies. Many more examples will undoubtedly be found.

In order to understand hovering flight, both normal and exceptional, we must now analyze the nature and generation of aerodynamic lift. It is easiest to start with a long horizontal cylinder that rotates clockwise on its axis at a constant angular velocity, which is symbolized by ω, the small Greek omega [*see top illustration at left*]. If the cylinder could be immersed in an idealized fluid with zero viscosity, its rotation would not affect the fluid. If, however, the cylinder is placed in a real fluid that has both mass and viscosity and is at rest, a cylindrical vortex will gradually build up. It is called a bound vortex.

This vortex is characterized by a circulation (Γ) that has the same direction and value at any distance from the surface of the cylinder. The molecules of air at the surface have the velocity r, where r is the radius of the cylinder. The circulation is then velocity times circumference ($\Gamma = 2\omega r^2$). So far no other forces are involved except for a small drag that resists the rotation and maintains the circulation.

If the entire mass of fluid is set in uniform translational motion (say from left to right), the combination of rotation and translation means that the resulting velocity is increased above the cylinder and decreased below it. The result, according to what physicists know as Bernoulli's theorem, is a decreased pressure above and an increased pressure below, so that the cylinder "feels" a force perpendicular to the horizontal movement. This force is the lift (L). To put it another way, the spinning cylinder feels a transverse force, which is known as the Magnus effect. It does not matter whether we are dealing with a spinning cylinder or with a real wing; if circulation is generated in one way or another, the relation between the lift, the circulation and the translation of the undisturbed fluid is the same per unit of axial length.

Against this background one sees that an airfoil or a wing in steady motion through the air is a device by means of which circulation is created and maintained in the form of a vortex bound to the wing. This bound vortex is then superposed on the flow pattern that the wing profile would produce in an ideal fluid. The pattern is termed the potential flow. Moving past the wing, it gives rise to neither lift nor drag, and its streamlines converge toward the horizontal behind the wing. When the bound vortex is combined with the ideal flow, the result is the smooth streamlined flow that can be calculated and can also be observed by means of smoke trails when a real wing is placed in a real fluid. Behind the trailing edge the fluid develops a downward momentum corresponding to the lift generated, as is the case with the spinning cylinder.

But how is the bound vortex created and maintained? The answer is crucial for an understanding of airfoils in general and for the problems of natural hovering flight in particular. It was provided by the German physicist Ludwig Prandtl in 1912 and can be described as follows.

If a wing suddenly starts from rest (an impulsive start), the initial flow pattern resembles the potential flow in an ideal inviscid fluid. It is of course useless insofar as it does not involve any downward momentum and hence any lift. In a real fluid, however, there is an area close to the trailing edge where the shearing viscous forces become large. Indeed, they are much larger than was once thought possible in air. The result is that a vortex—the starting vortex—is formed behind the wing. It has a directional sense opposite to the sense of the bound vortex. (It is a general rule in fluid dynamics that a vortex cannot be created without a countervortex of the opposite sense but of equal strength and circulation; in this instance the countervortex is the bound vortex.)

After the wing has moved by two or three chord lengths, or widths, through the air the smooth flow characteristic of the steady state is established. However, since it takes time to build up the starting vortex and since the starting and bound vortexes interact destructively in inverse proportion to their distance apart, the resulting lift is small at first and does not approach the steady value until about three chord lengths away from the starting point. The impulsive start of the wing of a hovering animal at each end of the stroke therefore hampers rather than facilitates the normal airfoil action. Consequently the abnormally high coefficients of lift observed for certain insects become even more interesting.

One should bear in mind that the useful bound vortex owes its existence and continuation to viscous shearing forces and that flight by means of normal airfoils would be impossible in an ideal fluid. On the other hand, if the viscous forces become large in relation to the inertial forces, as they do at low Reynolds numbers, the creation of circulation becomes difficult and the vortexes tend to die out quickly. This is the reason high coefficients of lift cannot be generated by small insects, with the result that the lift becomes relatively small and the drag large.

A value of about 2 is acceptable as a coefficient of lift for a bird wing operating at a Reynolds number of 7,000; the membranous wing of a locust, which works at 2,000, may reach 1.3. In contrast, as Steven Vogel found while working at Harvard University, in *Drosophila*, where the Reynolds number is 200, the coefficient of lift cannot exceed .9. We therefore found it particularly interesting to study the hovering and slow climbing flight of the tiny wasp *Encarsia formosa*, which was readily available because the operators of greenhouses use it to control a destructive aphid.

E. formosa has a wing length of only .6 millimeter. In general the creature resembles most four-winged insects except that its wings, like those of most other very small insects, have a marginal brim of hairs. The only way we could analyze its wing movements with reasonable adequacy was to make motion pictures of freely flying insects at from 7,000 to 8,000 exposures per second, a rate that was only just sufficient because the wings beat at 400 hertz (cycles per second).

When we disregarded the brim of hair and assumed that the wing movements were sinusoidal, we found the average coefficient of lift was as high as 5—an entire order of magnitude higher

ANIMAL	AIRBORNE MASS (GRAMS)	REYNOLDS NUMBER	COEFFICIENT OF LIFT
SMALL BAT (*PLECOTUS AURITUS*)	9	14,000	1.3
HUMMINGBIRDS:			
LARGE (*PATAGONA GIGAS*)	20	15,000	1.8
MEDIUM (*AMAZILIA FIMBRIATA*)	5	7,500	2
BEETLES (COLEOPTERA):			
ELEPHANT-DUNG BEETLE (*HELIOCOPRIS SP.*)	13	23,000	.5
COCKCHAFER (*MELOLONTHA VULGARIS*)	.6	4,700	.6
MOTHS (LEPIDOPTERA):			
PRIVET HAWK (*SPHINX LIGUSTRI*)	1.6	6,300	1.2
TOBACCO HORNWORM (*MANDUCA SEXTA*)	2.1	6,700	1.2
BUTTERFLIES (LEPIDOPTERA):			
GREEN-VEINED WHITE (*PIERIS NAPI*)	.04	1,400	2.2
BEES AND WASPS (HYMENOPTERA):			
BUMBLEBEE (*BOMBUS TERRESTRIS*)	.9	4,500	1.2
HONEYBEE (*APIS MELLIFERA*)	.1	1,900	.8
HORNET WASP (*VESPA CRABRO*)	.6	4,200	.8
CHALCID WASP (*ENCARSIA FORMOSA*)	.000025	15	5
TWO-WINGED INSECTS (DIPTERA):			
CRANE FLY (*TIPULA SP.*)	.03	770	.8
MOSQUITO (*AËDES AEGYPTI*)	.001	170	.6
FRUIT FLY (*DROSOPHILA VIRILIS*)	.002	210	1
HOVER FLIES (*SYRPHUS SPP.*)	.025	400	2 TO 6
DRAGONFLIES (ODONATA):			
AESHNA GRANDIS	.7	2,000	2 TO 3

FLIGHT CHARACTERISTICS of hovering animals are compared. The colored entries indicate unusual aerodynamic performance. The Reynolds number expresses the ratio between inertial forces involved in accelerating a volume of air and the internal viscous forces in the air; ordinary airfoil action is possible at the higher values but deteriorates sharply at the lower ones. A high coefficient of lift is unusual at low Reynolds numbers and proves to be the result of the distinctive wing movements of the clap-fling and flip patterns of the hoverers.

than would be expected at the low Reynolds number of from 15 to 20. Recently C. P. Ellington in my laboratory took the hairs into account as if they were a solid surface, and he also made his calculations on the basis of the true nonsinusoidal movements. Even so, the coefficient of lift was 1.7 and hence considerably in excess of what is possible in a situation of steady flow.

A clue to the resolution of this apparent paradox came from a frame-by-frame analysis of the films. They showed that *Encarsia* hovers in the normal way, with its body axis vertical and its wings sweeping almost horizontally. The tiny insect therefore uses the lift principle rather than a drag mechanism.

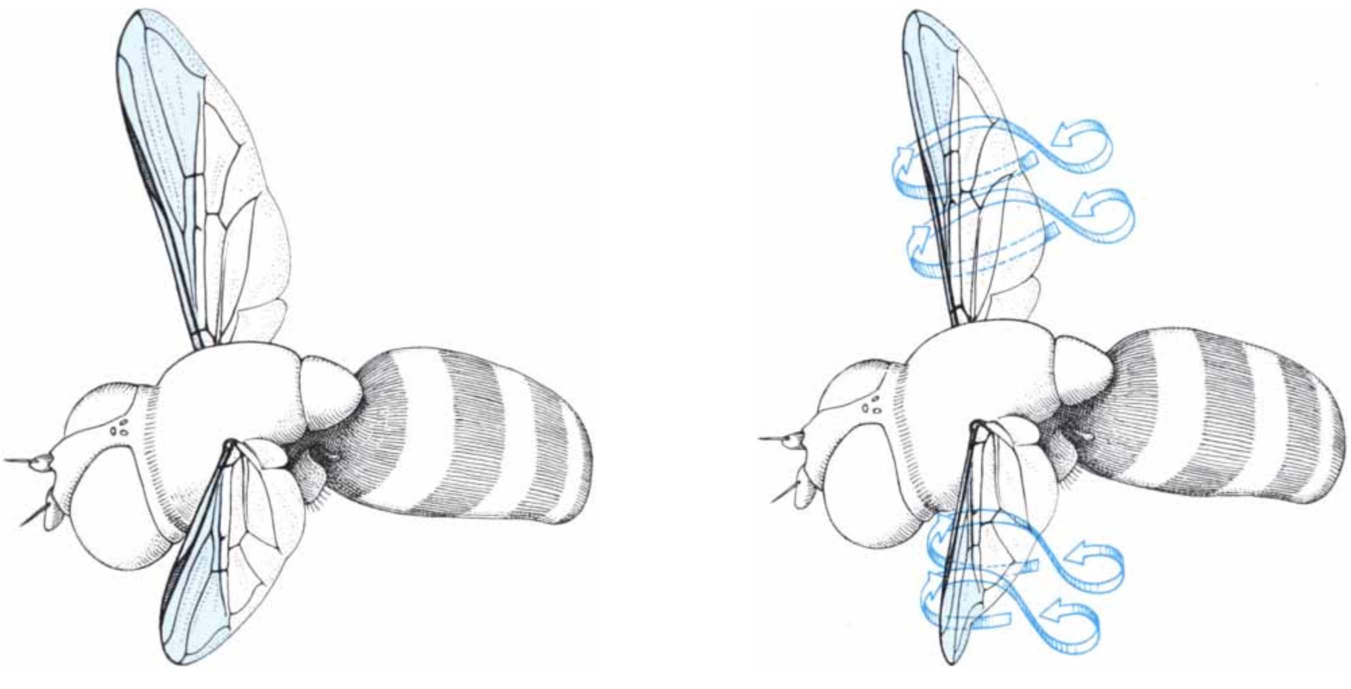
Toward the end of each upstroke, however, the two pairs of wings (two interlocked wings on each side) meet and form a single vertical plate over the back of the insect. It is as if the insect clapped its wings behind its back [*see illustration on page 82*]. They rest there for half a millisecond and then are suddenly flung open, like the rapid flinging open of a book, while the hind margins of the two pairs remain in touch. This action is what we call the fling. It happens during every wing cycle and in whatever way the insect flies.

After the fling the wings from the two sides separate and, in a hovering wasp, swing almost horizontally through the air. At the end of the downstroke the wings do not meet in a similar clap under the body. The sweeping action stops well before the point of clapping. The wings are then suddenly twisted through an angle of 120 degrees. This is the action we call the flip; it resembles the flipping of a pancake. The analysis showed that sufficient lift to overcome the weight of the insect's body was generated long before the horizontal angular velocity of the wing axis had reached its maximum, both after the fling and after the flip. The finding is in contrast to what one would expect according to ordinary airfoil action.

The novel clap-fling mechanism is best appreciated if one views the hovering wasp obliquely from behind and represents the two wings on each side by a single rectangular plate suspended from the muscle-filled thorax by means of a short elastic rod [*see illustration on page 83*]. During the clap the air is virtually at rest with respect to the wings until the right and left wings are suddenly flung open, thereby creating a growing wedge-shaped space between their upper surfaces. The space is undoubtedly filled immediately by air rushing in from the surroundings. This potential flow takes the form of two symmetrical vortexes of equal strength but opposite directional sense.

When the left and right wings split apart along the hind margins of the hind wings and begin their horizontal sweep, each side is already provided with a circulation. It is not created by the horizontal movement or by the viscous forces but by the potential flow generated by the fling. Therefore lift is generated as soon as the wings start their horizontal sweep. (The specialist may notice that the Helmholtz and Kelvin theorems have not been violated.) Sir James Lighthill of the University of Cambridge succeeded in analyzing the exact form of the potential flow theoretically, and by mathematical analysis we found that the lift should equal the body weight of the insect early in the downstroke, as is actually observed.

So far we have not taken into account viscous effects in the boundary layer: the layer of stagnant air immediately adjacent to the surface of the wing. They and certain aspects of the three-dimensional flow patterns were also analyzed by Lighthill. He concluded that they tend to enhance or at least not to invalidate the main effect caused by the fling. The general conclusion is therefore surprising: Contrary to what aerodynamicists have assumed so far, it is possible to create circulation and lift by means of wings without the intervention of viscous forces, that is, in an ideal inviscid fluid. As for the hovering insect,



FLIP MECHANISM is proposed to explain the hovering of the hover flies and the dragonflies. The model here is a hover fly. Whereas most hovering insects hover with their body more or less vertical and their wings moving in a horizontal figure eight, the hover flies and dragonflies hover with their body more nearly horizontal and their wings sweeping obliquely up and down through a small angle. At left the wings have completed the upstroke and are almost flat plates. The front parts are structurally reinforced and the rear surface is soft and flexible. At the beginning of the downstroke (*right*) the front part is suddenly twisted downward rapidly; the rear area is unaffected. This "flip" action could give rise to a system of vortexes, like those portrayed by the colored bands.

instead of starting the downstroke with zero circulation and building it up gradually during the succeeding sweep, animals that employ the clap-fling mechanism build up the circulation before the downstroke and not during it. The half stroke thus begins with coefficients of lift that are much higher than is possible in steady-state aerodynamics. In the course of the wing stroke the value must decrease toward the steady-state level, although so far we know little about the change with time. Instead of approaching the steady-state level from below, as in ordinary airfoils, these animals approach it from above.

Dimensional considerations indicate that the clap-fling mechanism does not depend critically on the size of the animal and may be useful in much larger animals, particularly fliers with a small wing loading or a low aspect ratio, that is, a low ratio of wing width to wing length. In my laboratory we have now seen the mechanism in the butterfly *Pieris,* in the small aphid that is parasitized by *Encarsia* and in *Drosophila.* Our calculations indicate that in each case it improves performance by a factor of from three to five. The findings may also apply to the flycatcher bird (*Ficedula hypoleuca*) when it hovers in front of its nest, since Ulla Norberg of the University of Göteborg has found that its coefficient of lift is close to 4.

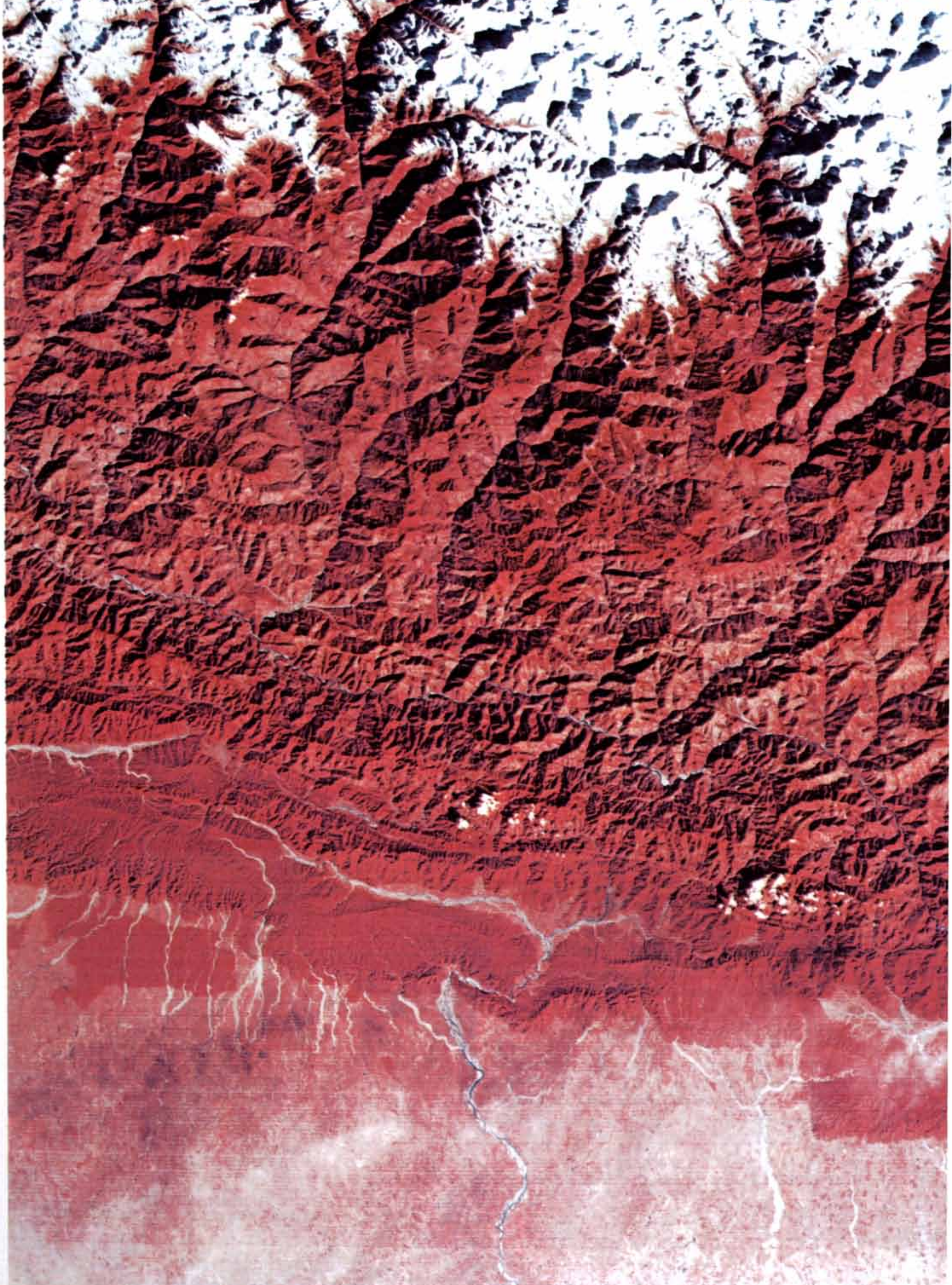
Under emergency conditions even the rock dove *Columba livia* and its domestic descendants, the pigeons, appear to make use of the clap-fling mechanism during a vertical takeoff. As early as 19 B.C. Virgil noted (in the fifth book of *The Aeneid*) that the rock dove starts a steep climb with one or two loud claps of the wings. In 1890 E. J. Marey, the famous French student of animal locomotion, demonstrated that the sounds are caused when the animal claps its two wings above its back. Multiple-flash photographic exposures by R. H. J. Brown of the University of Cambridge show that the claps initiate a true fling movement. My own rough calculations indicate that in this situation the wing of the pigeon starts with a circulation of a magnitude that otherwise could be achieved only much later in the downstroke. Virgil's poetic description of the rock dove that rushes up from its nest on a cliff with a few loud claps before it glides silently through the peaceful air may well be based on a real transition from nonsteady aerodynamics to steady-state aerodynamics.

The dragonflies (Odonata Anisoptera) and the hover flies (the Syrphidae) merit further consideration. When one of these insects hovers, its body axis remains horizontal as in forward flight and its wings beat obliquely up and down but through a rather small stroke angle. The coefficient of lift cannot possibly be smaller than 2 or 3 and it is likely to be much higher. In fact, R. A. Norberg of the University of Göteborg has recently found that it is 4.2 in the dragonfly *Aeshna juncea.* How is this result to be explained?

I have examined the wings of those insects and found that the anterior (leading) part is heavily reinforced by stiff ribs or veins, whereas the posterior (trailing) area is soft and pliable. In insects the wings are thin membranes and contain no muscles. The essential twisting movements are produced exclusively at the root of the wing, next to the body. If one calculates how fast a torsional wave can travel from the base to the tip in the anterior part of the wing, one finds that it has reached the tip well before the torsion or the twist can affect the trailing part. This conclusion is confirmed by the remarkable flash photographs of these insects made by Stephen Dalton and also by our recent films of hovering fruit flies and true hover flies.

We then have a possibility similar to but different from the clap-fling situation. If one views the wing of a hover fly obliquely from above just before the downstroke begins, one sees an almost flat surface [*see illustration on opposite page*]. Suddenly the leading area is twisted downward, while for a short time the trailing area remains virtually unaffected and motionless with respect to the air. This flip action may cause air to move to form a pair of vortexes. One of them is mainly around the anterior part of the wing; the countervortex is farther back. The trailing edge may then represent a stagnation line. When the wing starts its downward movement, it is already provided with a circulation of the correct sense for generating lift, and the countervortex is left behind.

It is too early to evaluate the validity of this notion. We have a long period of research ahead of us before we can ascertain the significant factors and provide quantitative estimates of air circulation, lift and drag and of how they change with time. The main point is that these insects and the other fliers I have discussed depend to a significant extent on nonsteady aerodynamic effects, hitherto unknown, which for them are beneficial rather than being a nuisance, as they are in man-made aircraft.



The Subduction of the Lithosphere

The rocky shell of the earth grows outward from mid-ocean ridges. Ultimately it plunges into the mantle below, giving rise to oceanic trenches, earthquakes, volcanoes, island arcs and mountain ranges

by M. Nafi Toksöz

The lithosphere, or outer shell, of the earth is made up of about a dozen rigid plates that move with respect to one another. New lithosphere is created at mid-ocean ridges by the upwelling and cooling of magma from the earth's interior. Since new lithosphere is continuously being created and the earth is not expanding to any appreciable extent, the question arises: What happens to the "old" lithosphere?

The answer came in the late 1960's as the last major link in the theory of sea-floor spreading and plate tectonics that has revolutionized our understanding of tectonic processes, or structural deformations, in the earth and has provided a unifying theme for many diverse observations of the earth sciences. The old lithosphere is subducted, or pushed down, into the earth's mantle. As the formerly rigid plate descends it slowly heats up, and over a period of millions of years it is absorbed into the general circulation of the earth's mantle.

The subduction of the lithosphere is perhaps the most significant phenomenon in global tectonics. Subduction not only explains what happens to old lithosphere but also accounts for many of the geologic processes that shape the earth's surface. Most of the world's volcanoes and earthquakes, including nearly all the earthquakes with deep and intermediate foci, are associated with descending lithospheric plates. The prominent island arcs—chains of islands such as the Aleutians, the Kuriles, the Marianas and the islands of Japan—are surface expressions of the subduction process. The deepest trenches of the world's oceans, including the Java and Tonga trenches and all others associated with island arcs, mark the seaward boundary of subduction zones. Major mountain belts, such as the Andes and the Himalayas, have resulted from the convergence and subduction of lithospheric plates.

In order to appreciate the gigantic scale on which subduction takes place, consider that both the Atlantic and the Pacific oceans were created over the past 200 million years as a consequence of sea-floor spreading. Thus the lithosphere that underlies the world's major oceans is less than 200 million years old. As the oceans opened, an equivalent area of lithosphere was simultaneously subducted. A simple calculation shows that the process involved the consumption of at least 20 billion cubic kilometers of crustal and lithospheric material. At the present rate of subduction an area equal to the entire surface of the earth would be consumed by the mantle in about 160 million years.

To understand the subduction process it is necessary to look at the thermal regime of the earth. The temperatures within the earth at first increase rapidly with depth, reaching about 1,200 degrees Celsius at a depth of 100 kilometers. Then they increase more gradually, approaching 2,000 degrees C. at about 500 kilometers. The minerals in peridotite, the major constituent of the upper mantle, start to melt at about 1,200 C., or typically at a depth of 100 kilometers. Under the oceans the upper mantle is fairly soft and may contain some molten material at depths as shallow as 80 kilometers. The soft region of the mantle, over which the rigid lithospheric plate normally moves, is the asthenosphere. It appears that in certain areas convection currents in the asthenosphere may drive the plates, and that in other regions the plate motions may drive the convection currents.

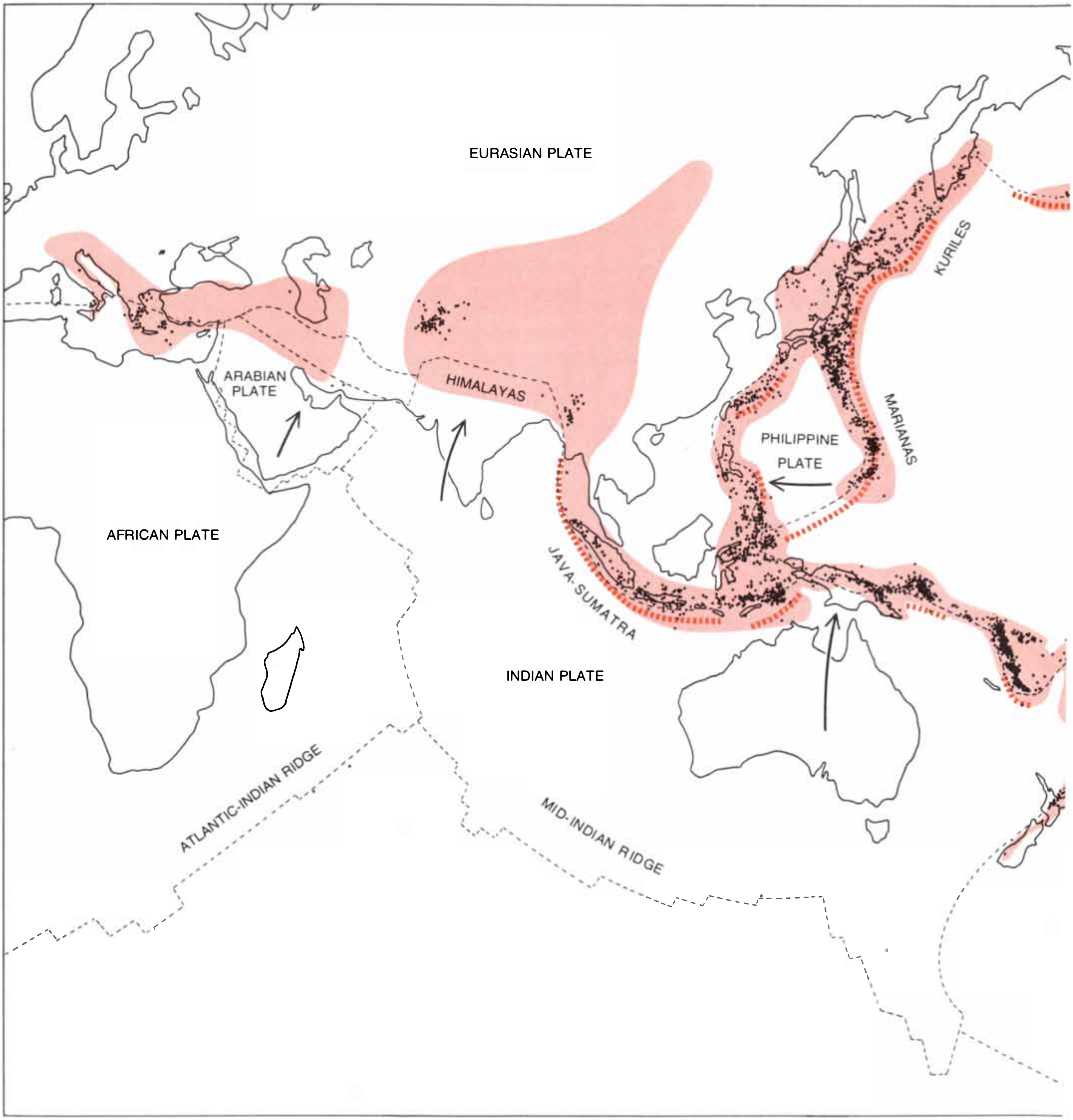
The mid-ocean ridges mark the region where upwelling material forms new lithosphere. The ridges are elevated more than three kilometers above the average level of the ocean floor because the newly extruded rock is hot and hence more buoyant than the colder rock in the older lithosphere. As the lithosphere spreads away from the ridge it gradually cools and thickens. The spreading rate is generally between one centimeter and 10 centimeters per year. The higher velocities are associated with the Pacific plate and the lower velocities with the plates bordering the Mid-Atlantic Ridge. At a velocity of eight centimeters per year the lithosphere will reach a thickness of about 80 kilometers at a distance of 1,000 kilometers from the ridge. Under most of the Pacific abyssal plains a thickness of this value has been confirmed by measurements of the velocities of seismic waves.

Where two plates move toward each

HIMALAYAS OF NEPAL, shown in this false-color picture made from the Earth Resources Technology Satellite (ERTS), are a zone in which continental lithosphere is being subducted. In most subduction zones oceanic lithosphere plunges under continental lithosphere. Here the lithosphere of the Indian subcontinent (*bottom*) is being subducted under the snow-covered Himalayas (*top*), raising the mountain range in the process. The area covered by the picture is 125 kilometers (78 miles) across. Mount Everest is one of the peaks on the ridge at the very edge of the picture in the upper right-hand corner. The main boundary fault between the two lithospheric plates runs from left to right in the valley that is marked by two clusters of cloud that are visible at lower center and lower right.

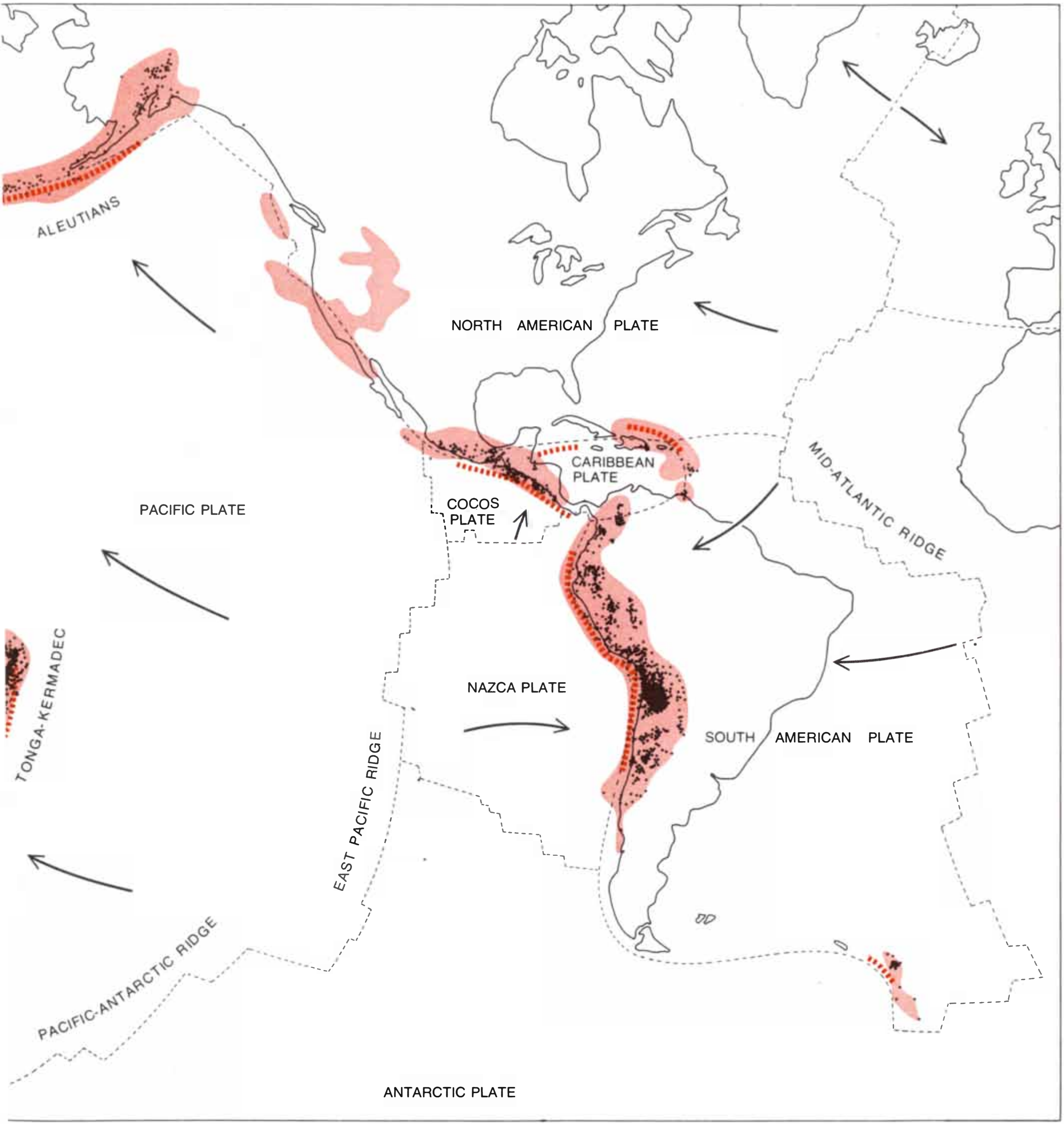
other and converge, the oceanic plate usually bends and is pushed under the thicker and more stable continental plate. The line of initial subduction is marked by an oceanic trench. At first the dip, or angle of descent, is low and then it gradually becomes steeper. Profiles across trenches of the reflections of seismic waves clearly show the downward curve of the top of descending oceanic plates.

Several factors contribute to the heating of the lithosphere as it descends into the mantle. First, heat simply flows into the cooler lithosphere from the surrounding warmer mantle. Since the conductivity of the rock increases with temperature, the conductive heating becomes more efficient with increasing depth. Second, as the lithospheric slab descends it is subjected to increasing pres-

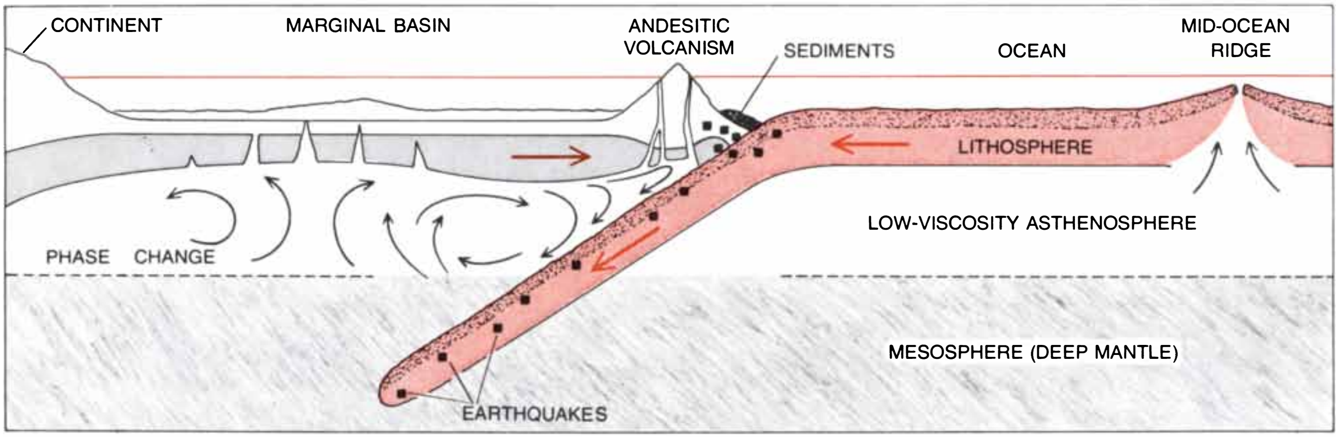


TECTONIC MAP OF THE EARTH depicts the principal lithospheric plates and their general direction of motion (*arrows*). New material is continually added to the plates at mid-ocean ridges by the upwelling and cooling of magma from the earth's mantle. It moves outward and is eventually returned to the mantle by subduction. There it is slowly consumed. The subduction process creates deep oceanic trenches (*broken lines in color*) and island arcs such as those bordering the western and northern Pacific. On the islands of the arcs are many active volcanoes. Young mountain belts in Europe and Asia identify zones where continental lithosure, which introduces heat of compression. Third, the slab is heated by the radioactive decay of uranium, thorium and potassium, which are present throughout the earth's crust and add heat at a constant rate to the descending material. Fourth, heat is provided by the energy released when the minerals in the lithosphere change to denser phases, or more compact crystal structures, as they are subjected to higher pressures during descent. Finally, heat is generated by friction, shear stresses and the dissipation of viscous motions at the boundaries between the moving lithospheric plate and the surrounding mantle. Among all these sources the first and fourth contribute the most toward the heating of the descending lithosphere.

The temperatures inside a descending



spheric plates converge; around the Pacific young mountain ranges result from the subduction of oceanic plates. The areas in color identify the general location of the great majority of earthquakes that occurred at all depths between 1961 and 1967; they are based on maps made by H. J. Dorman and M. Barazangi of the Lamont-Doherty Geological Observatory. Most earthquakes have a magnitude below 6.5 and occur at a shallow depth (between five and 15 kilometers). The locations of deep earthquakes, those occurring below 100 kilometers, are given by the black dots. All the deep earthquakes take place in cold descending slabs of the oceanic type.



FORMATION AND SUBDUCTION OF LITHOSPHERE are shown in this cross section of the crust and mantle. New lithosphere is created at a mid-ocean ridge. A trench forms where the lithospheric slab descends into the mantle. Earthquakes (*small squares*) occur predominantly in the upper portion of the descending slab. Arrows in soft asthenosphere indicate direction of possible convective motions. Secondary convection currents in asthenosphere may form small spreading centers under marginal basins.

lithospheric plate have been calculated theoretically over the past five years by geophysicists in Britain, Japan and the U.S. Although different approaches were taken in the calculations, the results are in good agreement. For example, our group at the Massachusetts Institute of Technology has computed the progressive heating of plates penetrating into the mantle at various velocities over periods ranging from several hundred thousand years to more than 10 million years. A typical calculation based on our model of the phenomenon shows what happens to a plate descending at the rate of eight centimeters per year (the velocity characteristic of the Pacific subduction zones) at three points in time: 3.6, 7.1 and 12.4 million years after the beginning of subduction [*see illustration on opposite page*].

NAME	PLATES INVOLVED	TYPE	LENGTH OF ZONE (KILOMETERS)	SUBDUCTION RATE (CENTIMETERS PER YEAR)	MAXIMUM EARTHQUAKE DEPTH (KILOMETERS)	TYPE OF SUBDUCTING LITHOSPHERE
KURILES-KAMCHATKA-HONSHU	PACIFIC UNDER EURASIAN	A	2,800	7.5	610	OCEANIC
TONGA-KERMADEC-NEW ZEALAND	PACIFIC UNDER INDIAN	A	3,000	8.2	660	OCEANIC
MIDDLE AMERICAN	COCOS UNDER NORTH AMERICAN	B	1,900	9.5	270	OCEANIC
MEXICAN	PACIFIC UNDER NORTH AMERICAN	B	2,200	6.2	300	OCEANIC
ALEUTIANS	PACIFIC UNDER NORTH AMERICAN	B	3,800	3.5	260	OCEANIC
SUNDRA-JAVA-SUMATRA-BURMA	INDIAN UNDER EURASIAN	B	5,700	6.7	730	OCEANIC
SOUTH SANDWICH	SOUTH AMERICAN SUBDUCTS UNDER SCOTIA	C	650	1.9	200	OCEANIC
CARIBBEAN	SOUTH AMERICAN UNDER CARIBBEAN	C	1,350	0.5	200	OCEANIC
AEGEAN	AFRICAN UNDER EURASIAN	C	1,550	2.7	300	OCEANIC
SOLOMON–NEW HEBRIDES	INDIAN UNDER PACIFIC	D	2,750	8.7	640	OCEANIC
IZU-BONIN-MARIANAS	PACIFIC UNDER PHILIPPINE	D	4,450	1.2	680	OCEANIC
IRAN	ARABIAN UNDER EURASIAN	E	2,250	4.7	250	CONTINENTAL
HIMALAYAN	INDIAN UNDER EURASIAN	E	2,400	5.5	300	CONTINENTAL
RYUKYU-PHILIPPINES	PHILIPPINE UNDER EURASIAN	E	4,750	6.7	280	OCEANIC
PERU-CHILE	NAZCA UNDER SOUTH AMERICAN	E	6,700	9.3	700	OCEANIC

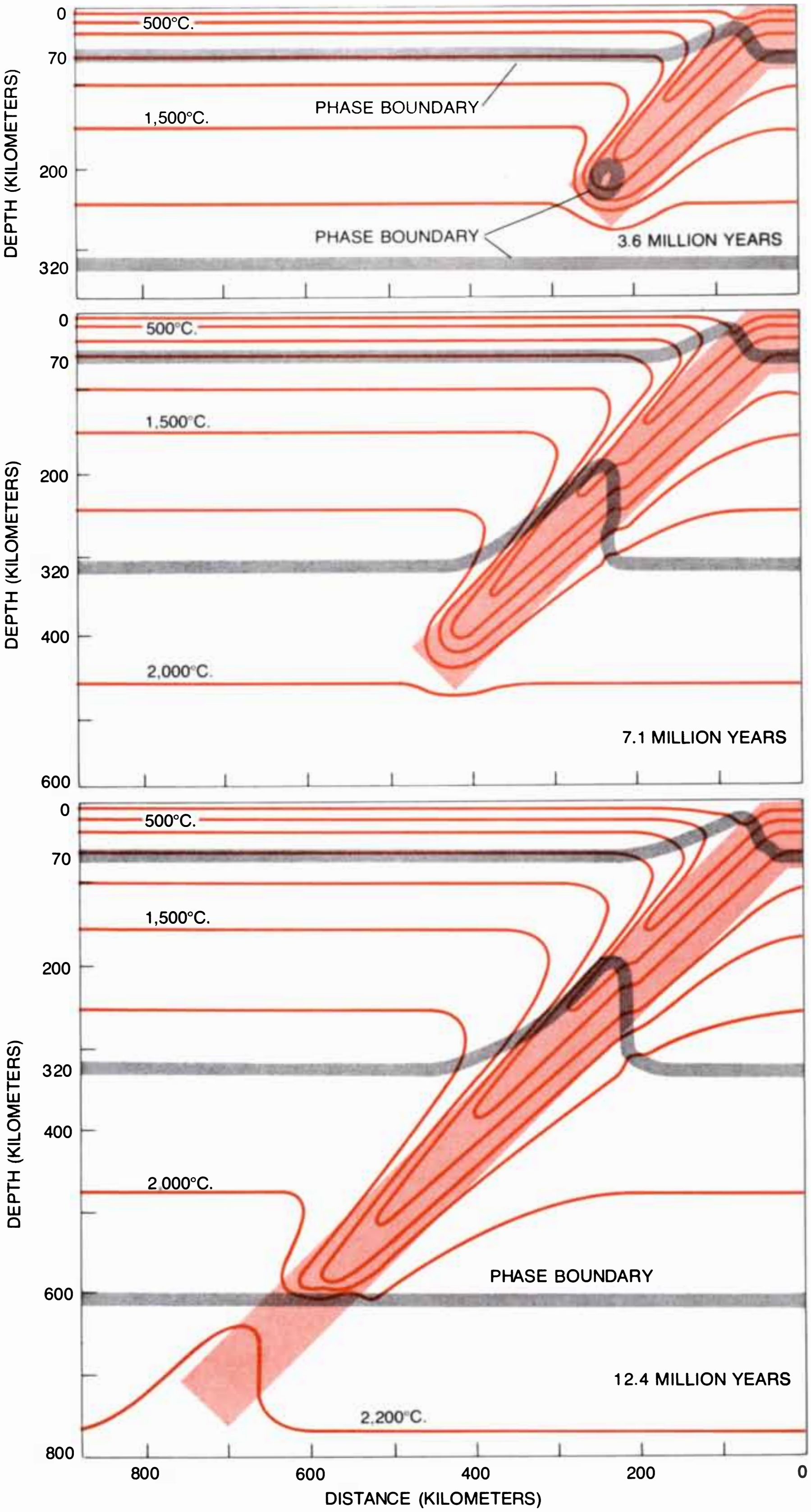
MAJOR SUBDUCTION ZONES and some of their principal characteristics are listed. One of the smallest plates, the Nazca plate, is associated with the longest single subduction zone, embracing almost the entire west coast of South America. It also has the second-highest subduction rate: 9.3 centimeters per year perpendicular to the arc of the earth's surface. In general the more rapidly a plate descends, the greater is the maximum depth of earthquakes associated with it. (A major exception is the subduction zone under the Philippines.) The five principal types of subduction zone (*A—E*) are depicted schematically in the illustration on page 94.

In this model the interior of the descending plate remains distinctly cooler than the surrounding mantle until the plate reaches a depth of about 600 kilometers. As the plate penetrates deeper its interior begins to heat up more rapidly because of the more efficient transfer of heat by radiation. When the plate goes beyond a depth of about 700 kilometers, it can no longer be thermally distinguished as a structural unit. It has become a part of the mantle. Significantly, 700 kilometers is a depth below which no earthquake has ever been recorded. Apparently deep earthquakes cannot occur except in descending plates; therefore the occurrence of such earthquakes implies the presence of sunken plate material.

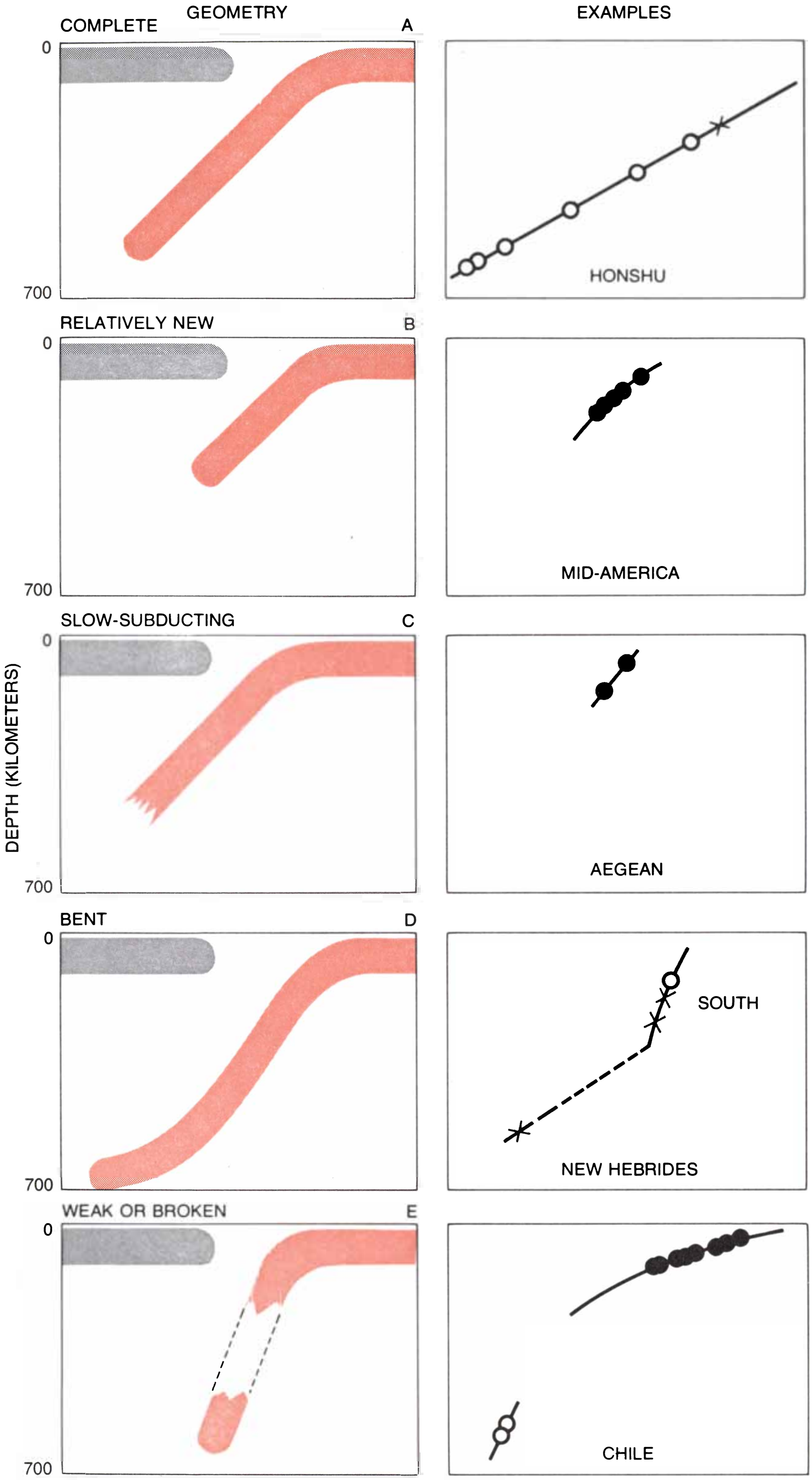
The descending lithosphere does not always, however, penetrate to 700 kilometers before it is assimilated. A slow-moving plate will attain thermal equilibrium before reaching that depth. For example, at a velocity of one centimeter per year the subducting plate will be assimilated at a depth of about 400 kilometers. If subduction ceases altogether, the subducted segment of the lithosphere will lose its identity and become part of the surrounding mantle in roughly 60 million years. At half that age a stationary plate will already have become too warm to generate earthquakes. These calculations make it clear why we can identify only those subducted plates that are associated with the latest episode of sea-floor spreading. Although there are surface geological expressions of older subduction zones, the plates subducted under these regions cannot be identified in the earth's mantle. The old slabs are lost not only because of the assimilation process but also because of the motion of the surface with respect to the mantle.

So far I have been describing ideal subduction zones without major complications. Such zones are found, for example, under the Japanese island of Honshu, under the Kuriles (extending to the north of Japan) and under the Tonga-Kermadec area (to the north of New Zealand). In many other areas the lithosphere descends in a more complicated manner.

In new subduction areas the descending slab may have penetrated a good deal less than 700 kilometers, as is the case under the Aleutians, the west coast of Central America and Sumatra. In other areas where the subduction rate is low the slab may be assimilated well before it reaches that depth; the subduction



EVOLUTION OF DESCENDING SLAB is described by computer models developed in the author's laboratory at the Massachusetts Institute of Technology. These diagrams depict the fate of a slab subducting at an angle of 45 degrees and at a rate of eight centimeters per year. Phase changes, induced by increasing pressure, normally occur at depths of 70, 320 and 600 kilometers. In the descending slab the first two phase changes occur at shallower depths because of the slab's lower temperature. The phase conversions to denser mineral forms help to heat the slab and to speed its assimilation. When the slab reaches the temperature of the surrounding mantle at a depth of 700 kilometers, it loses its original identity.



FIVE MAJOR TYPES of subducting oceanic slabs can be identified. Examples of each type are shown to the right of the schematic diagrams. In the examples the solid lines represent the location of all earthquakes projected onto a cross section. The symbols on the lines identify particularly large earthquakes from which the direction of stress was determined. Open circles indicate compression along the length of the slab; filled circles indicate tension along the length of the slab, and crosses show stresses that do not lie in the plane of the cross section. Many subduction zones exhibit a "seismic gap" between 300 and 500 kilometers where no earthquakes occur. It is not known whether this is because the slab is broken (*Type E*) or because stresses are absent at that depth. Examples given are based on a survey conducted by Bryan L. Isacks of Cornell University and Peter Molnar of M.I.T.

of the Mediterranean plate under the Aegean Sea is an example. In still other areas the subduction starts at a shallow angle, gets steeper at intermediate depths and bends again nearly to the horizontal at about 500 kilometers. Such a sigmoid configuration is observed dramatically under the New Hebrides in the South Pacific. The double bend may be attributable to low resistance in the upper asthenosphere and much greater resistance at a depth of 600 kilometers, resulting from an increase either in the density or in the strength of the mantle, possibly both. Another anomalous situation is found under Peru and Chile, where there is a marked absence of earthquakes at intermediate depths, indicating a stress-free zone or possibly a broken slab.

Most frequently the oceanic lithosphere is subducted under an island arc, as is generally the case in the western Pacific. Here, however, there are many other combinations and complications. For example, a small oceanic plate, such as the Philippine plate, may get trapped between two trenches. Or an oceanic plate may be subducted under a continent, as in the case of the Nazca plate, which plunges under the Andes. The Andes can be regarded as being equivalent to an overgrown island arc. Elsewhere transform faults such as the San Andreas fault may interrupt subduction boundaries. In other cases multiple subduction zones may develop within relatively small areas. Finally, subducting plates may bring two continents together, with major tectonic consequences. Continental collisions place major restrictions on plate motions because the buoyancy of the continental crust, which is less dense than the mantle, resists subduction. Collisions of this type create major mountain belts, such as the Alps and the Himalayas.

Continental subduction is qualitatively different from oceanic subduction because it is a transient process rather than a steady-state one. When continental crust moves into a subduction zone, its buoyancy prevents it from being carried down farther than perhaps 40 kilometers below its normal depth. As plate convergence continues the crust becomes detached from the plate and is itself underthrust by more continental crust. That creates a double layer of low-density crust, which rises buoyantly to support the high topography of a major mountain range. It is possible that the long oceanic slab below the surface ultimately becomes detached and sinks; in any case it is no longer a source of earthquakes. After this stage further deformation and compression may take place behind the line of collision, producing a high plateau with surface volcanoes, like the plateau of Tibet. Eventually the plate convergence itself will stop as resisting forces build up. It now seems that continental collisions are probably a major factor in the periodic reorientations of the relative motions of the plates.

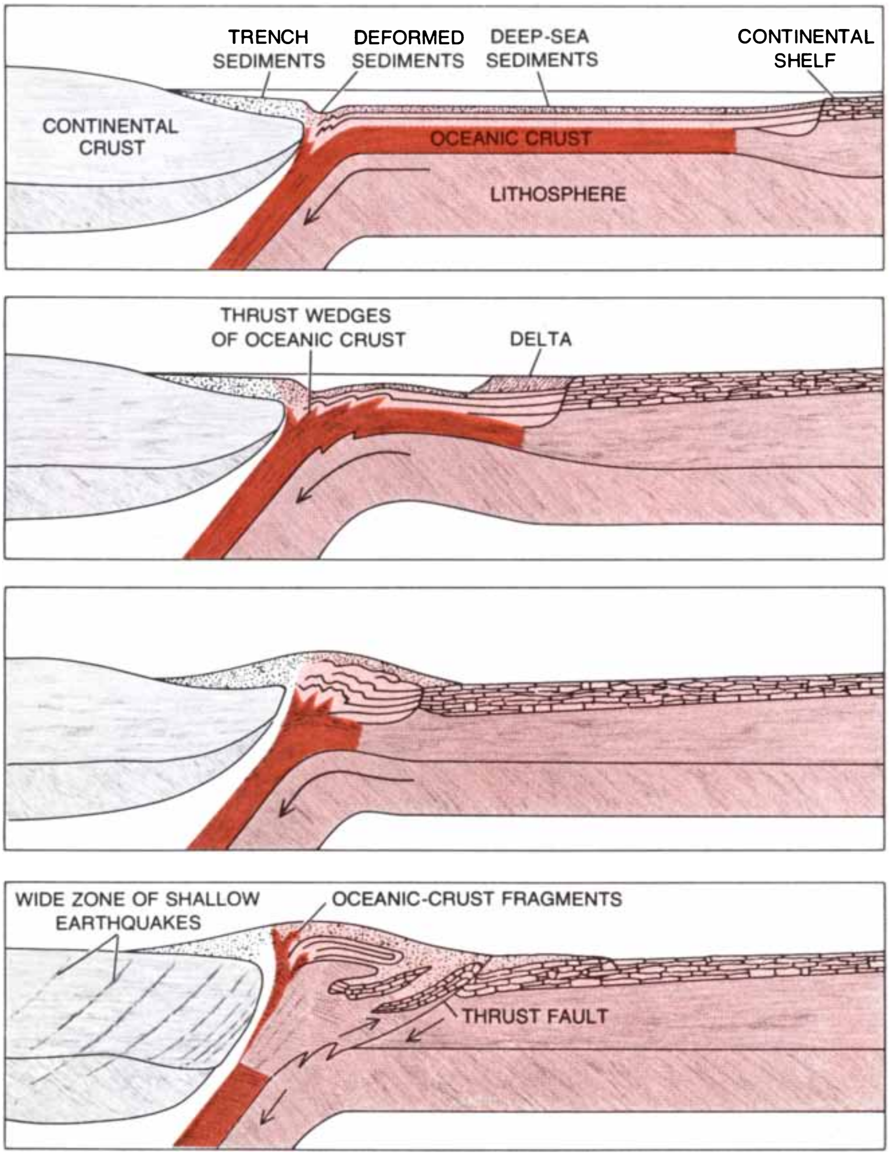
It is clear that an understanding of the geological, geochemical and geophysical consequences of lithospheric subduction helps to explain many major features of the earth's surface. At the same time the observable features enable us to test the validity of theoretical subduction models. A wide variety of features can be investigated. For the sake of brevity I shall mention only the geological characteristics of the trench sediments and the subducting crust, the andesitic magmas associated with island-arc volcanoes, and heat-flow and gravity anomalies. The measurable quantities related to these features are primarily sensitive to the properties of subduction down to a depth of about 100 kilometers. The most definitive observations on the deeper parts of subducting plates are seismic observations. The velocity and attenuation of seismic waves, and most significantly the indication the waves give of the locations of deep- and intermediate-focus earthquakes, outline the extent of the relatively cool and rigid zone of the descending lithosphere.

With the passage of time the deep oceanic trenches created by descending plates accumulate large deposits of sediment, primarily from the adjacent continent. As the sediments get caught between the subducting oceanic crust and either the island arc or the continental crust they are subjected to strong deformation, shearing, heating and metamorphism. Profiles of seismic reflections have identified these deformed units. Some of the sediments may even be dragged to great depths, where they may eventually melt and contribute to volcanism. In this case they would return rapidly to the surface, and the total mass of low-density crustal rocks would be preserved.

A prominent feature of subduction zones is volcanism that gives rise to andesite, a fine-grained gray rock. Where the magma for these volcanoes originates is not definitely known. Most geochemical and petrological evidence favors a depth of about 100 kilometers for the magma source. The magma may come from the partial melting of the subducted oceanic crust, as A. E. Ringwood of the Australian National University suggested in 1969. The shearing that takes place at the top of the descending plate may provide the heat required for partial melting. Convective motions in the wedge of asthenosphere above the descending plate may also contribute to magma sources by raising asthenospheric material to a depth where it could melt slightly under lower pressure.

The flow of heat through the earth's surface tells us something about the thermal characteristics of shallow layers. (It is influenced only indirectly by deeper phenomena.) Trenches have low heat flow (less than one microcalorie per square centimeter per second); island arcs generally have a high and variable heat flow because of their volcanism. High heat flow is also associated with the marginal basins behind the island arcs, for example the Sea of Japan, the Sea of Okhotsk, the Lau Basin west of Tonga and the Parece Vela Basin behind the Marianas arc.

These basins are underlain by relatively hot material brought up either by convection currents behind the island arc or by upwelling from deeper regions. The convection is induced in the wedge



COLLISION OF CONTINENTS occurs when an oceanic slab that is subducting at the edge of one continent (*left*) is itself part of a lithospheric plate bearing a second continent (*right*). Such a collision took place when the Indian lithospheric plate, traveling generally northward for 200 million years, subducted under Eurasian plate. This kind of subduction eventually ends, but not before crust of subducting plate has been detached and deformed and has pushed up a mountain range (in the case of the Indian plate the Himalayas).

of asthenosphere above the descending lithospheric plate by the downward motion of the plate. Since it takes time for such currents to be set in motion, high heat flow would not be expected in basins behind the youngest subduction zones. Indeed, the observed heat-flow values in the Bering Sea behind the Aleutians are normal.

Gravity anomalies associated with subduction zones are large and broad. A descending lithospheric plate is cooler and denser than the surrounding mantle; therefore it gives rise to a positive gravity anomaly. The hot region under a marginal basin would show a density lower than normal and hence would create a negative gravity anomaly. Changes in the character of the crust from the ocean to an island arc or a continent add more anomalies. A combination of all these anomalies is needed to account for the gravity observations that have been made across subduction zones [*see illustration on opposite page*]. The gravity evidence provides strong support for the subduction models, but it is not conclusive because of the uncertainties as to the depth of the masses that give rise to the anomalies.

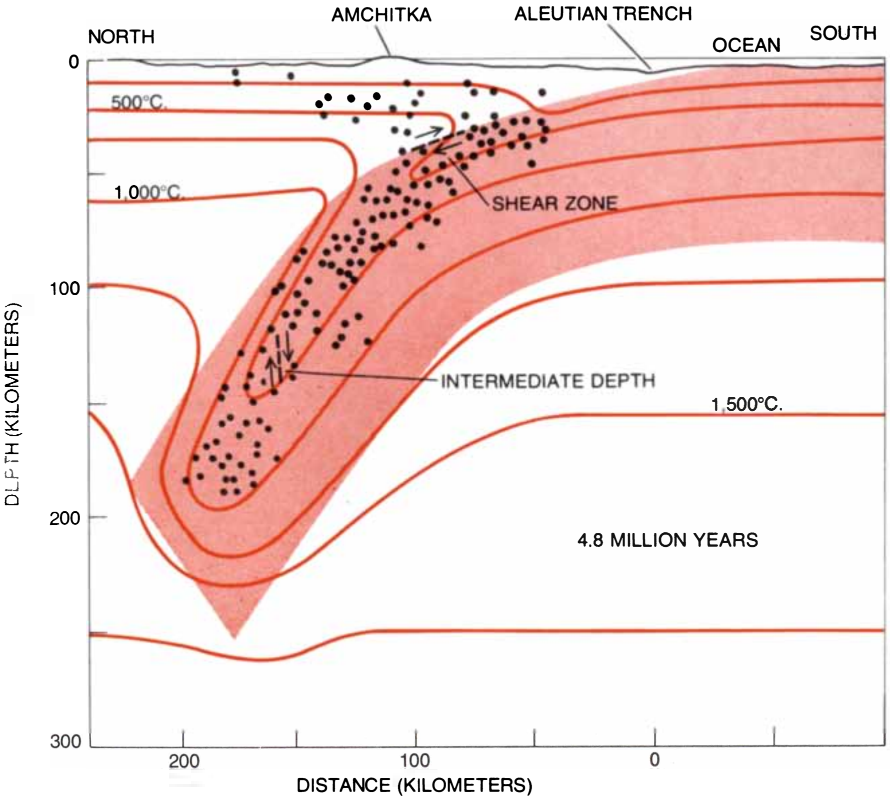
The most compelling evidence for the subduction of the lithosphere comes from seismology. Most of the world's earthquakes and nearly all the deep- and intermediate-focus earthquakes are associated with subduction zones. The hypocenters of the earthquakes and their source mechanisms can be explained by the stresses in the subducting plate. The models that explain the seismic-wave observations outline the location of the subducted cool lithospheric plates. In some areas (Japan, the Aleutians, the Tonga Trench, South America) the data are abundant and convincing.

The general picture that emerges is as follows. At shallow depths, where the edges of the two rigid lithospheric plates are pressing against each other, there is intense earthquake activity. Many of the world's greatest earthquakes (for example the Chile earthquake of 1960, the Alaska earthquake of 1964 and the Kamchatka earthquake of 1952), as well as many smaller ones, occur along the shear plane between the subducting oceanic lithosphere and the continental or island-arc lithosphere. Some normal-faulting (tensional) earthquakes on the ocean side of a trench are caused by arching of the lithosphere. Other earthquakes result from the tearing of the lithosphere and other adjustments in this zone of intense deformation.

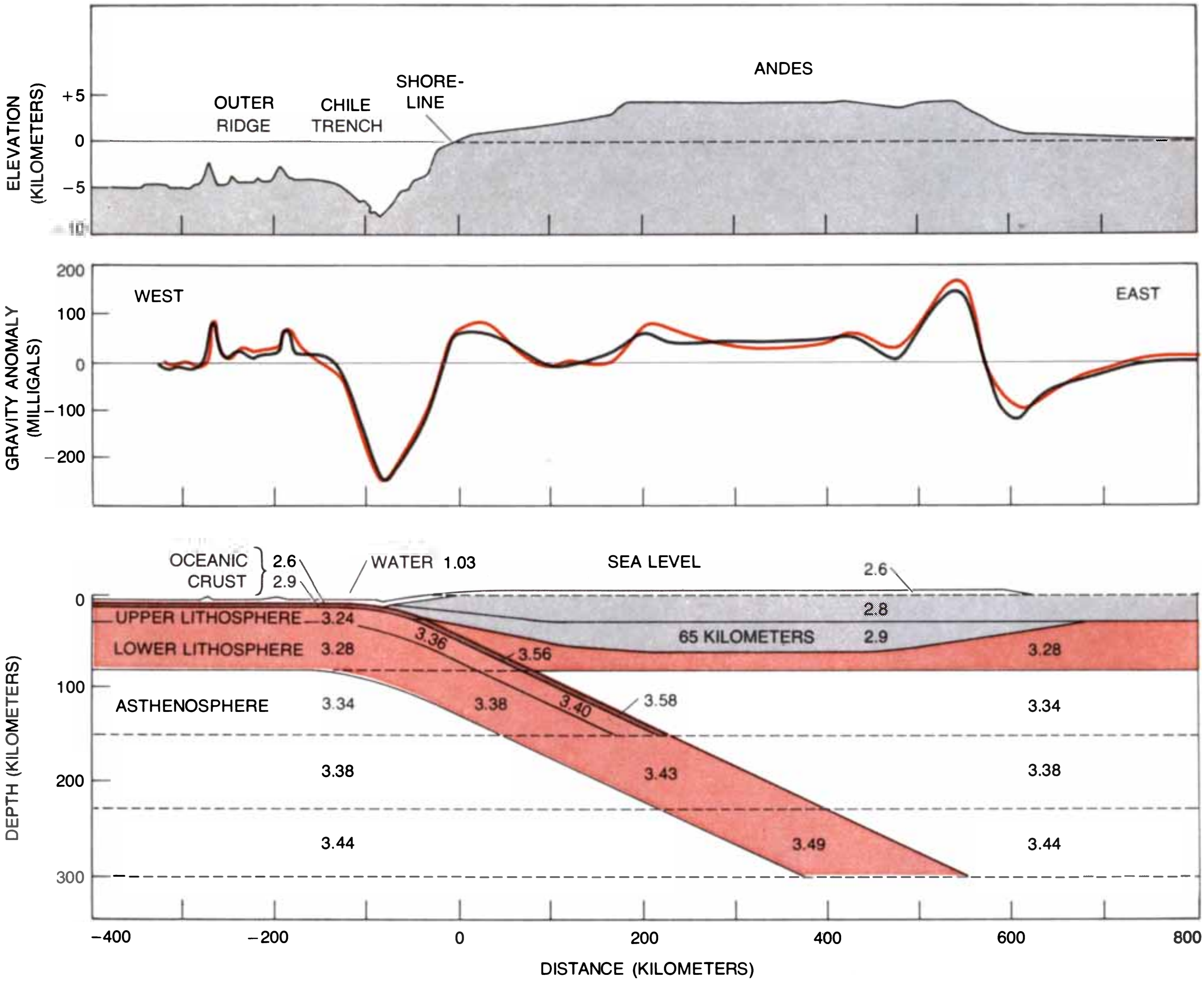
The deep- and intermediate-focus earthquakes generally occur along the Benioff zone, a plane that dips toward a continent. At first this plane was thought to be the shear zone between the upper surface of the descending lithospheric plate and the adjoining mantle. Detailed studies conducted by Bryan L. Isacks of Cornell University and Peter Molnar of M.I.T. and others over the past 10 years have shown, however, that the forces needed to account for the observed earthquakes could not be provided by the shearing process. These studies, combined with more precise determinations of the location of earthquake foci under several island arcs, indicate that the deep- and intermediate-focus earthquakes occur in the coolest region of the interior of the descending plate. The stresses generated by the gravitational forces acting on the dense interior of the slab and the resistance of the surrounding mantle to the slab's penetration are also highest in the coolest region. Moreover, the cool and rigid interior of the slab acts as a channel to transmit stresses. The computed directions of the stresses are consistent with the directions that have been deduced from earthquakes.

These concepts can be tested in areas where detailed studies of earthquakes have been made. Two such regions are the Aleutians and Japan. At Amchitka Island in the central Aleutians the nuclear explosions named Longshot, Milrow and Cannikin provided energy sources with precisely known locations and times. From the travel times of the seismic waves going through the subducting lithosphere the location of the coolest region was determined precisely. The dense network of seismic stations installed in the area also provided precise locations of earthquakes. The shallow earthquakes are concentrated along the thrust plane and the deeper ones along the coolest region [*see illustration at left*].

The islands of Japan constitute probably the most intensively studied seismic belt in the world. The velocities of seis-



ALEUTIAN EARTHQUAKES mark the general location of the subducting Pacific plate in that region. The precise location of the cold descending slab in relation to the earthquakes was determined with the help of seismic waves from nuclear tests on Amchitka Island, which showed that the waves travel more rapidly through the cold slab than through the surrounding mantle. A computed simulation of seismic records revealed that intermediate-depth earthquakes (*dots*) occur in the cold center of the slab, as shown here, and not, as had been thought, at the shear zone on its upper face. At shallower depths the earthquakes occur in shear zone and in overriding plate. Arrows show the slip planes and the sense of motion.



EFFECT OF A SUBDUCTING PLATE ON GRAVITY is clearly represented in the gravity anomaly that has been measured over the west coast of Chile and the Andes. The diagram at the top is a topographical cross section of the region. The observed gravity anomaly, given in milligals, is shown by the black curve in the middle diagram. The colored curve is the anomaly calculated on the basis of the lithospheric model shown at the bottom. (One gal, named for Galileo, is one 980th the normal gravity at the earth's surface; thus an anomaly of —260 milligals over the trench corresponds to a gravity deficit of about .026 percent.) The model includes the trench, which gives rise to the gravity low, and cold dense slab, which has opposite effect. Densities given in model are in grams per cubic centimeter. Model was worked out by J. A. Grow and Carl O. Bowin of the Woods Hole Oceanographic Institution.

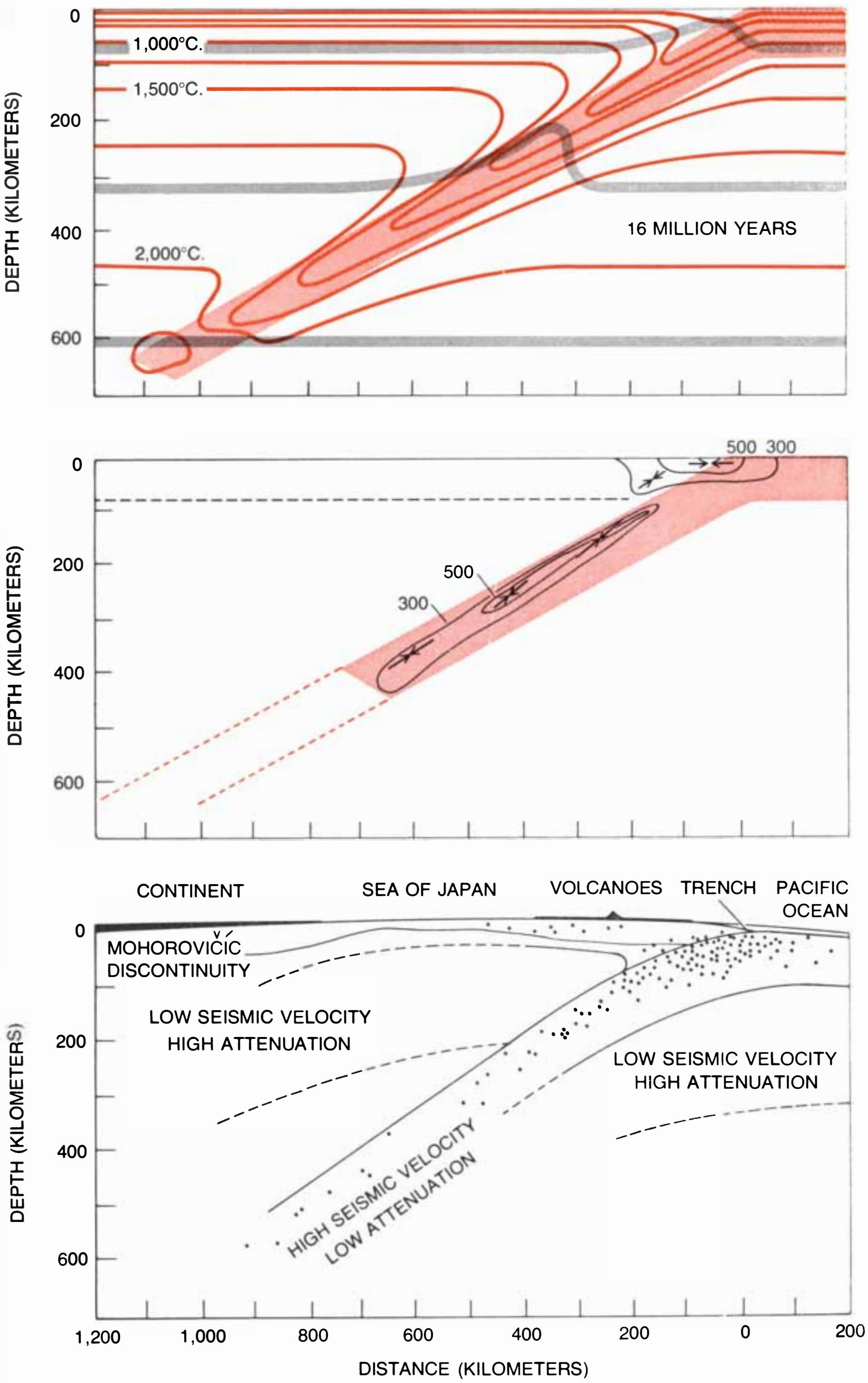
mic waves, the characteristics of the waves' attenuation, the precise locations of earthquake hypocenters and the focal mechanisms all fit the subduction model in this region. The descending plate shows high velocities and low attenuation, which is a measure of the nonelastic damping of high-frequency seismic waves. There are numerous shallow earthquakes along and near the boundary where the plates meet near the surface. Deep- and intermediate-focus earthquakes are in the coolest region of the slab where the stresses are highest [*see illustration on next page*]. In other subduction zones the locations of earthquakes are not as precisely known. Nevertheless, wherever adequate data exist, for example for the areas of the Tonga Trench and of Peru and Chile, the deep- and intermediate-focus earthquakes are found to occur in the interior of the subducting plate along the coolest region.

The absence of earthquakes below a depth of 700 kilometers can now be explained. The descending lithosphere heats up below that depth and can no longer behave as a rigid elastic medium susceptible to faulting or brittle fracture. Moreover, below that depth the stresses are small, and they are relieved by slow plastic deformation rather than by the sudden failure associated with an earthquake.

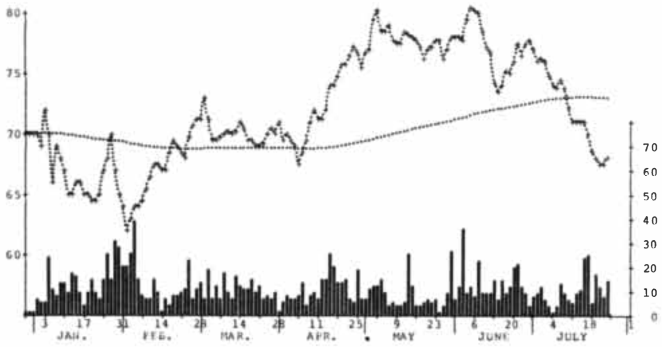
The gravitational energy associated with large masses of subducting cool, dense material is large even in terms of the total energy associated with plate motions. The gravitational forces are largely balanced by the resistance of the mantle to the penetration of the descending lithosphere. The net force acting on the plates in the subduction zone is still enough to play a major role in global plate motions. Other forces that contribute are the horizontal flow of convection currents under the plates and the outward push of the material coming to the surface at the mid-ocean ridges.

Not all the problems of plate motions and subduction have been solved. It is puzzling, for example, that the Pa-

cific plate can move laterally for 6,000 kilometers before it subducts. It is not known why some subduction zones are where they are. It is not clear why plate motions change at certain times. These are minor problems, however, compared with the understanding of continental drift, earthquakes, volcanism and mountain building that has been gained. The theory of plate tectonics is a concept that unifies the main features of the earth's surface and their history better than any other concept in the geological sciences.



EARTHQUAKES IN JAPAN AREA are caused by several westward-dipping slabs of Pacific lithosphere. The author has calculated a temperature model for a typical Japan slab (*top*). This in turn has been used to calculate the stresses generated within the upper portion of the slab (*middle*). The stresses result from the interaction between the slab's tendency to sink because of its high density and opposing forces: friction near surface and viscous drag in asthenosphere. Nonhydrostatic stresses are computed in bars (one bar is 14.7 pounds per square inch). Arrows show direction of compression. Calculated stresses account well both for distribution of earthquakes (*bottom*) and their mode of initiation.



An important peripheral to the 9800 series of calculators is the 9871 printer/plotter. Its bidirectional platen and programmable horizontal and vertical tabulation permit the calculator-controlled preparation of graphs, charts, and text, such as are combined in this statistical portrait. Price of the printer is $3400*.

A powerful new desk-top programmable that gives you four dimensions of computation.

The HP 9815A provides keystroke calculation for scratch-pad-type problem solving; dedicated problem handling with automatic set-up; programmability for sophisticated problems; and interfacing capability for extended I/O, plus instrumentation control, data acquisition, and data processing. Add its range of peripherals, and its price, and you have a flexible, economical, on-site "computer" for virtually any lab.

Running a program on the new 9815 desk-top programmable calculator is this easy: position the mode switch to AUTO-START, plug in a prerecorded, high-speed data cartridge, and turn on the power. The calculator automatically loads and executes the first program on the cartridge. You can choose from HP's library of prerecorded program pacs in such fields as statistics, electrical engineering, medicine, and surveying. Successive users can quickly personalize the calculator for their particular need.

Program and data storage is compact: a single data cartridge holds up to 48 programs, each up to 2,008 steps long, providing 96,384 bytes of memory. Bidirectional search and interchangeable data and program storage capabilities give the calculator the power and speed to handle complex programs.

The 9815 contains the computer functions necessary to give you complete programming language and power. Besides facile programmability, it has an extensive preprogrammed keyboard for quick, single-stroke calculations: 28 scientific function keys and 4 arithmetic keys. It has HP's own uncomplicated logic system, to give you answers you can trust. It has a built-in printer that prints instructions, and gives you a permanent, labeled copy of your calculation, plus a numeric display.

The fourth dimension of this new desk-top programmable is its interfacing capability. An optional $200* two-channel I/O module allows a choice of seven different HP peripherals to work with the 9815, including the new HP 9871A printer/plotter.

The 9815 also accommodates the Hewlett-Packard Interface Bus, which allows it to control, gather, and process data from as many as 15 HP-IB-compatible instruments. The 9815 also works well with BCD instruments and devices with 8-bit parallel interfaces.

Quite a repertoire for a price of $2900*.

The Synthesis of Diamond at Low Pressure

Natural diamonds and most man-made ones form at high pressure. It is also possible to synthesize diamond by growing it from existing diamonds in a low-pressure gas rich in carbon

by B. V. Derjaguin and D. B. Fedoseev

When the English chemist Smithson Tennant showed in 1797 that diamond is a form of carbon (diamond and graphite are the only crystalline forms of the element), he set in motion a new kind of alchemy in which the objective was to make diamonds out of graphite or some other carbonaceous material. The requirements were to achieve the conditions of high pressure and temperature under which natural diamond is formed deep in the earth. After much effort by many people the objective was attained in 1955 by a group of workers at the General Electric Company.

The pressure required is on the order of 55,000 atmospheres (808,500 pounds per square inch) and the temperature is about 2,000 degrees Celsius. Both the pressure and the temperature must be sustained for a considerable period of time. After 1955 a number of investigators, including the members of our group at the Institute of Physical Chemistry in Moscow, undertook to explore the possibility of synthesizing diamonds at low pressure. We have now devised means of enlarging a seed crystal of diamond by exposing it to a carbonaceous gas at a pressure of less than one torr (.001 atmosphere).

Diamond is so unyielding that it is the hardest of all materials. Graphite yields so easily in bulk form that one of its principal uses is as a lubricant. These differences arise from the different crystal structure of the two materials [*see illustrations on pages 104 and 105*].

The crystal structure of diamond is cubic. An atom of carbon at the center of the cube is surrounded symmetrically by four other carbons, all at the same distance and arranged at the corners of a regular tetrahedron. The great hardness of diamond results from the strength of the carbon-carbon bonds and the symmetrical arrangement of the atoms.

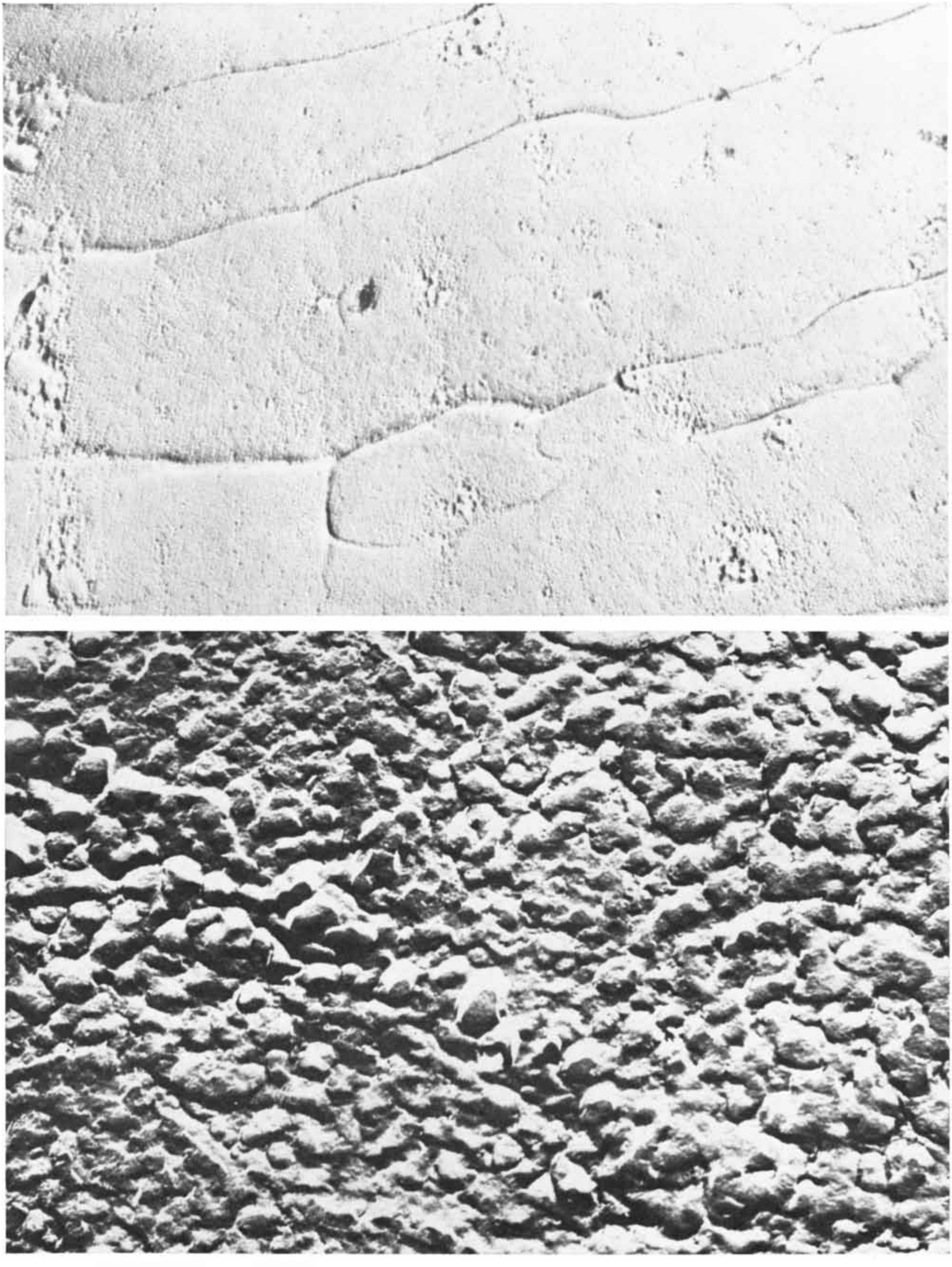
Graphite is organized in layers, each of which consists of hexagons with carbon atoms at the corners. Within each layer the pattern is not greatly different from that in diamond, but the layers are fairly widely spaced and the bonds between the atoms in one layer and the atoms in another are weak. Graphite serves as a lubricant because the layers can slip over one another under the influence of weak mechanical forces. On the other hand, the strong bonds between the atoms within a layer impart to graphite a high tensile strength in the direction of the layer. This is the reason that "whiskers," or threadlike crystals, of graphite have a high tensile strength.

The synthesis of diamond rests on those differences in the arrangement of carbon atoms, but all of that was unknown to the early investigators. The story of the synthesis really begins with the efforts of various people in the 17th and 18th centuries to determine what diamond is. In 1675 Isaac Newton, comparing the refraction of light in diamond and in certain combustible fluids, concluded that diamond must be combustible. Robert Boyle substantiated the conclusion by showing that diamond is altered in flame. In 1694 the Florence Academy conducted public experiments in which diamond was burned by light focused by a large convex lens. Antoine Laurent Lavoisier, investigating the products of diamond combustion, discovered in 1772 that they make clear lime water turbid: a test that is indicative of the formation of carbon dioxide.

Tennant achieved his demonstration that diamond is carbon by burning a diamond in a closed gold vessel filled with oxygen. The combustion produced carbon dioxide with a content of carbon exactly corresponding to the weight of the diamond burned. Two years later Jean-François Clouet and H. Guillton of France also reached the conclusion that diamond is pure carbon. They substantiated the conclusion by heating pure iron with diamond and producing an excellent steel; carbon is an essential component of steel, and their pure iron could only have become steel by obtaining carbon from the burning diamond. In the same year (1799) Guillton discovered that graphite is also carbon.

The first experiments aiming at the creation of artificial diamonds were carried out in the 19th century. In 1823 V. Karazin, the founder of Kharkov University, obtained crystals of a refractory substance that had a high proportion of carbon. Dmitri Ivanovich Mendeleev, the creator of the periodic table of the elements, judged the crystals to be similar to diamond. Six years later Charles Cagniard de la Tour carried out in France a number of experiments that essentially repeated the work of Karazin.

In 1880 the Scottish experimenter James B. Hannay reported that he had made diamonds by heating a mixture of hydrocarbons, "bone oil" and lithium at red heat in sealed wrought-iron tubes. (All but three of 80 tubes exploded.) A number of experimenters who later tried to make diamonds in this way were unsuccessful, and the work was more or less forgotten until 1943, when a small exhibit labeled "Hannay's Diamonds" was found in an obscure corner of the British Museum. F. A. Bannister and Kathleen Lonsdale examined the objects by X-ray diffraction and concluded that they were diamonds. It has never been determined, however, whether or not the diamonds are artificial.



GROWTH OF DIAMOND at low pressure is shown in these electron micrographs. At the top a facet of a seed crystal of diamond is shown before the growth process. At the bottom is the facet after it had been exposed in a carbonaceous gas for several hours at a pressure below one atmosphere and a temperature of 1,050 degrees Celsius. Growth rate was from .01 to .1 micrometer per hour.

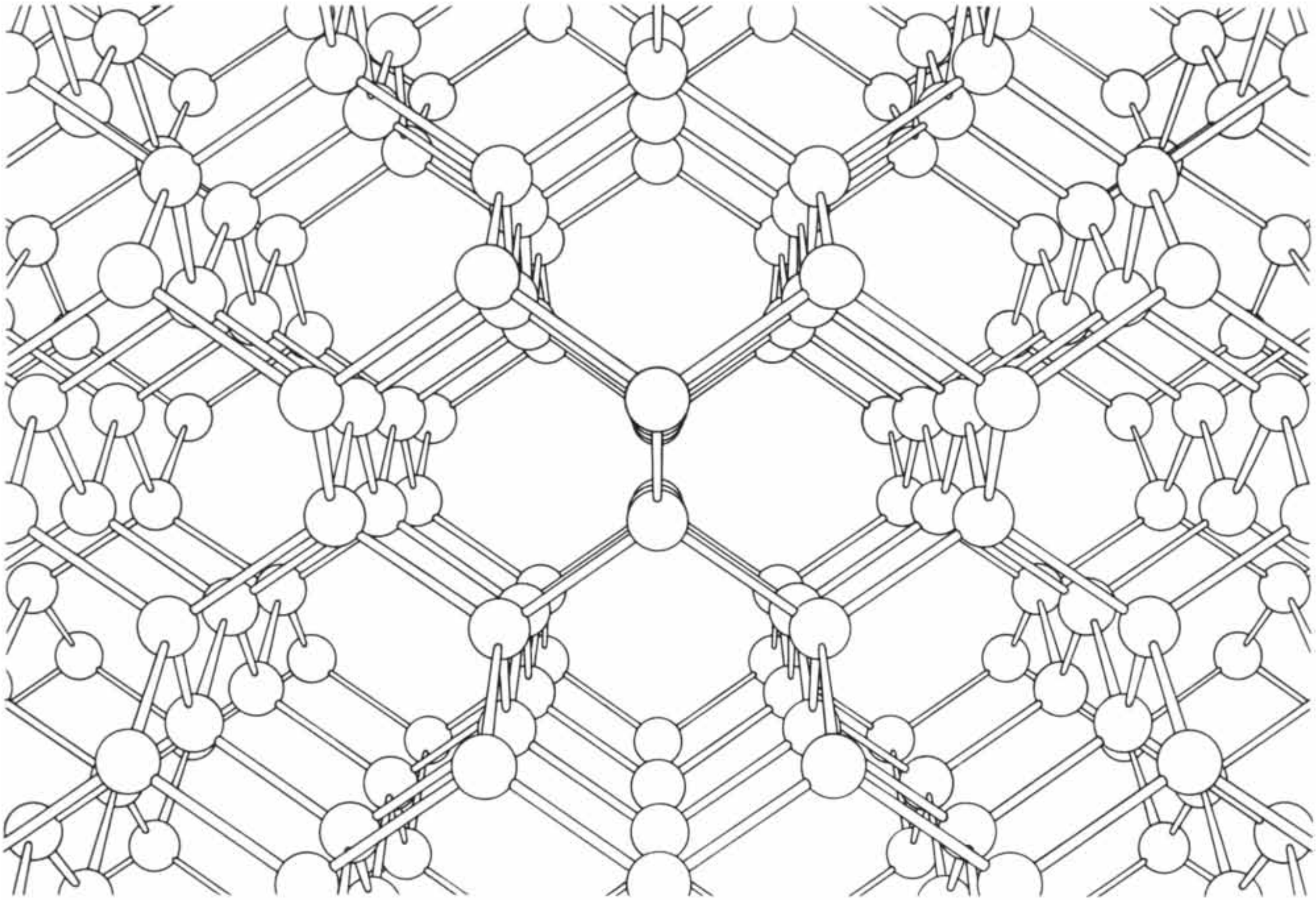
Another 19th-century experimenter who maintained he had made diamonds was the French chemist Henri Moissan, the inventor of the electric furnace. In his furnace he heated carbon-saturated iron in a crucible to 3,000 degrees C. and plunged the crucible into cold water. A high pressure was set up within the mass by the formation of a solid crust around it. When Moissan dissolved a cooled ingot in acid, he discovered grains that did not react with the acid and were capable of scratching ruby. The density of the grains was within the diamond range (diamond having a density of 3.51 grams per cubic centimeter, compared with 2.25 for graphite), and on burning they formed carbon dioxide. In 1893 Moissan reported that his grains were diamonds. Subsequent experimenters also thought they had made diamonds by similar methods, but the consensus now is that the pressure generated in such experiments could not have been nearly high enough to make diamonds and that the materials produced were probably varieties of the hard mineral spinel.

It was also in 1893 that K. Khrushchev, professor of mineralogy at the Petersburg Medical Academy, produced (independently of Moissan's work) dark transparent crystals by crystallizing carbon in molten silver. The crystals could scratch corundum, the hardest material after diamond, and they formed carbon dioxide on combustion. Nonetheless, modern work indicates that the crystals could not have been diamonds.

All the work we have described so far involved efforts to synthesize diamond from solutions of carbon. Probably the first attempt with a carbonaceous gas was made in 1911 by W. Bowlton, who tried to build up single crystals of diamond from illuminating gas. The attempt failed.

Complementing the efforts to make diamonds were numerous attempts to define theoretically the conditions for converting graphite into diamond. The problem was most successfully tackled by O. I. Leipunskii of the Institute of Chemical Physics of the Academy of Sciences of the U.S.S.R., who showed that all the efforts to make diamonds had been pursued under conditions in which graphite is thermodynamically a more stable form of carbon than diamond is. To convert graphite into diamond requires, first, subjecting the graphite to a pressure of some 55,000 atmospheres and, second, preheating it to about 2,000 degrees C. Under these conditions at least some of the graphite will be converted into diamond.

Achieving these conditions for the manufacture of diamonds called for solving difficult engineering problems connected with attaining high pressure and temperature simultaneously and in the same space. The effort necessitated using special structural materials and making numerous measurements. The General Electric group that first succeeded in achieving the appropriate conditions included Francis P. Bundy,



LATTICE OF DIAMOND, which is one of the two crystalline forms of carbon (the other is graphite), consists of infinitely repeating tetrahedral arrays of carbon atoms (*circles*). The lattice structure is portrayed here as though the viewer were inside it. Each carbon atom lies at the center of a regular tetrahedron whose vertexes are defined by the four nearest neighboring atoms. The atoms are connected by covalent bonds; it is the strength and symmetry of the bonds that give rise to the hardness of diamond.

H. Tracy Hall, Herbert M. Strong and Robert Wentorf. Synthetic diamonds were made at about the same time in Sweden and thereafter in Britain and other countries.

In the U.S.S.R. diamonds were synthesized by a group of workers under the direction of L. F. Vereshchagin of the Institute of Physics of High Pressures of the Academy of Sciences. Important contributions to the development of the diamond-synthesis technique and of industrial processes for producing diamonds have been made by a group under the direction of V. Bakul at the Institute for Ultrahard Materials in Kiev. At that institute the new synthetic ultrahard materials named Kubonite and Slavutich have recently been developed; in some respects they are not inferior to diamond.

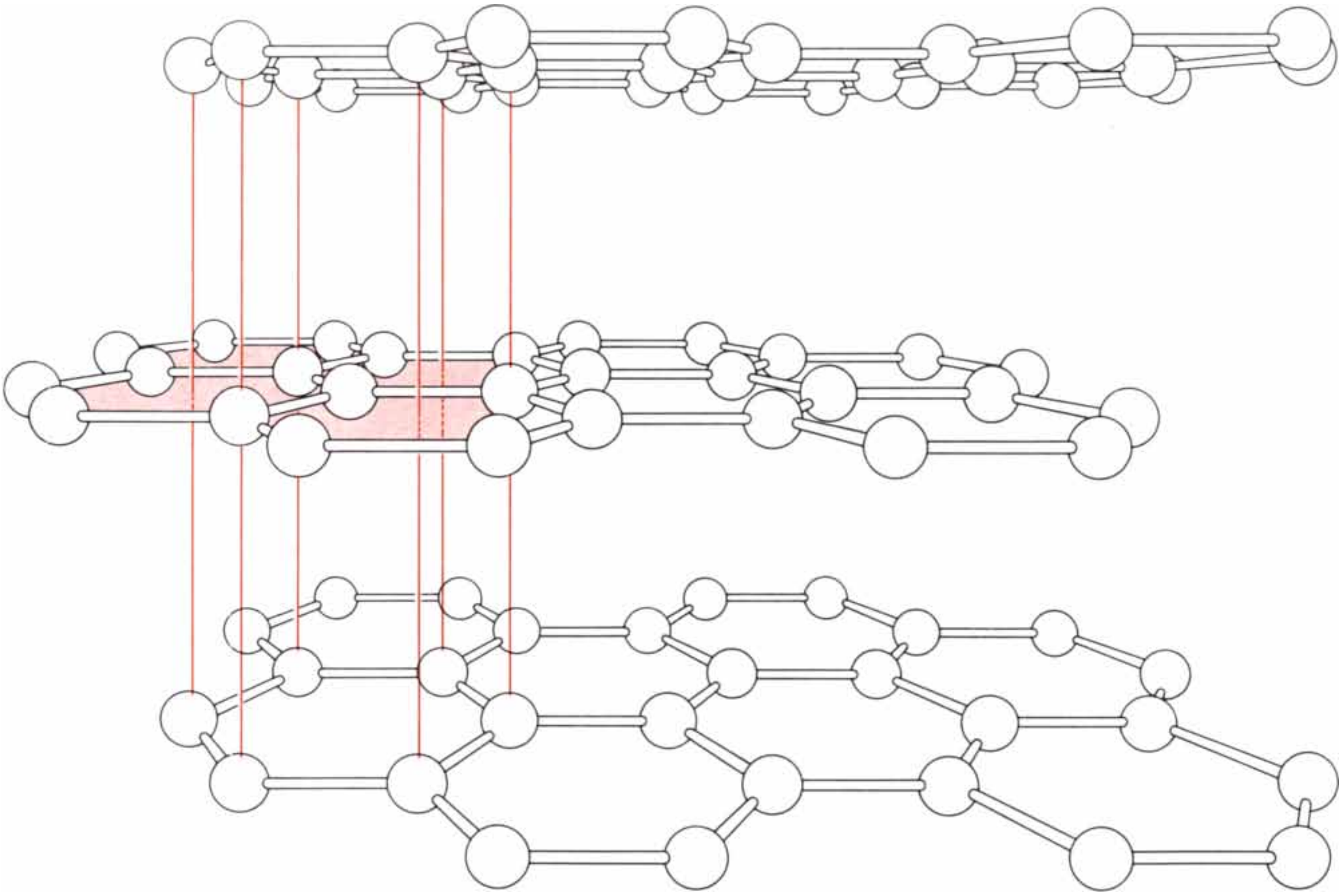
The fact that graphite and diamond have the identical composition naturally suggested attempting to convert graphite into diamond by applying high pressure to the graphite and thereby pushing the carbon atoms into the denser crystal lattice that is characteristic of diamond. If the carbon atoms are to shift from the graphite lattice to the diamond lattice, however, they must be given added mobility. That mobility is achieved by raising the temperature.

The various thermodynamic conditions under which graphite can be converted into diamond can be plotted on a chart as an equilibrium curve, which separates the two states of carbon [*see illustration on page 107*]. At each fixed temperature graphite will inevitably be converted into diamond at a pressure exceeding the equilibrium one, whereas at any pressure below equilibrium the diamond made out of graphite is metastable, so that it reverts to graphite.

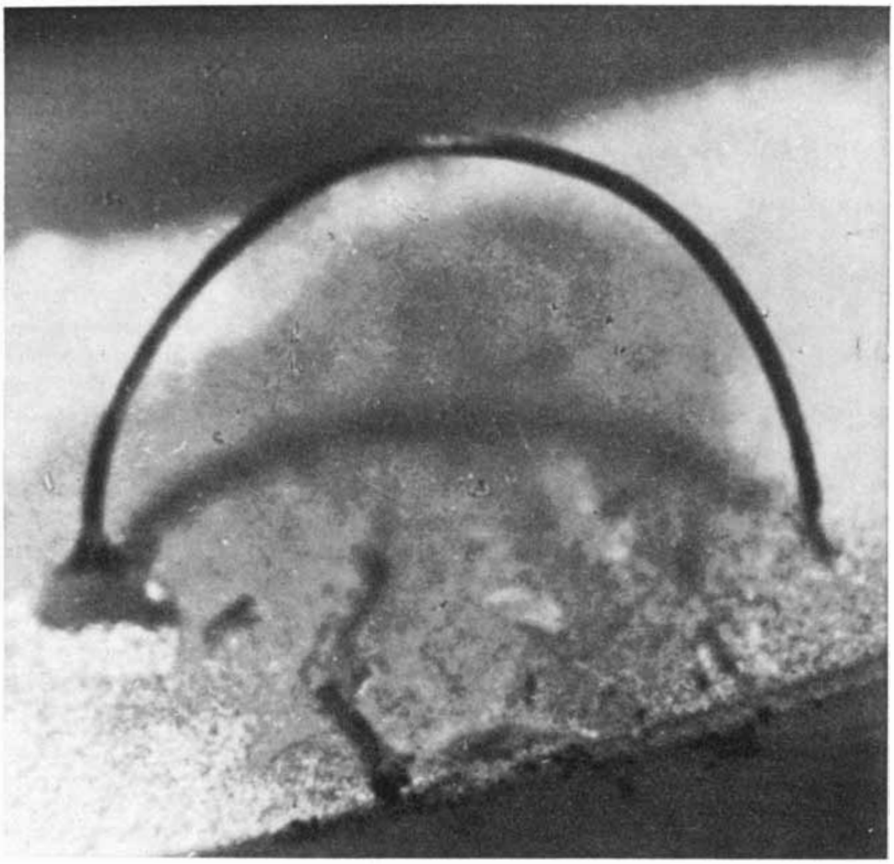
The situation changes, however, if one considers the possibility of the growth of a diamond crystal that is already available. Diamond can grow at the pressures and temperatures in the region of its metastability, provided that an environment of carbon atoms is created around the diamond by some appropriate means. The growth can be attributed to the fact that the concentration of carbon atoms collecting at the surface of the diamond is higher than the concentration of atoms with which diamond can be in equilibrium at a given temperature. In other words, the carbon atoms must be above the saturation concentration. Examples of ways of achieving the necessary concentration include evaporating carbon from graphite (by heating it to a high temperature) and dissolving carbon in a molten metal.

A supersaturated concentration of carbon atoms around diamond can also be achieved with a carbonaceous gas. If the molecules of the gas can decompose, giving off carbon atoms, at the temperature of the diamond seedbed, then at a sufficient density of gas the concentration of carbon atoms might reach the saturation level. Under those conditions a diamond crystal could grow.

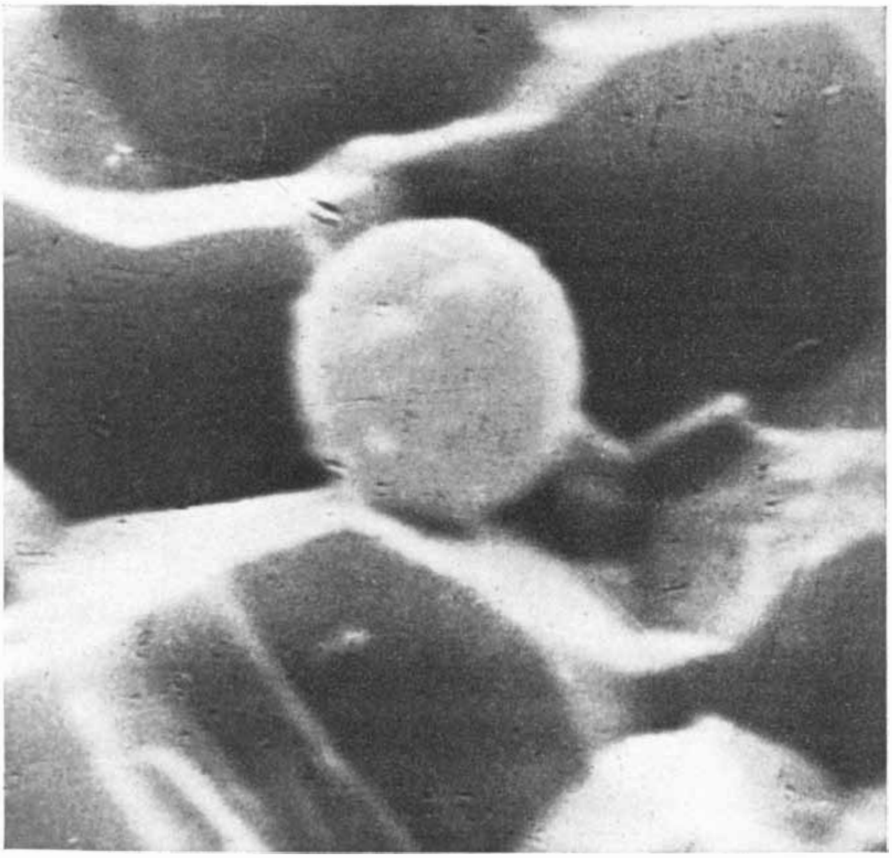
Compared with a process using graph-



LATTICE OF GRAPHITE has its carbon atoms bonded together in hexagons arrayed in planes. Within a plane graphite is harder than diamond, but the bonds between planes are weak. This property accounts for the usefulness of graphite as a lubricant: the weakly bonded planes slide readily in relation to one another. The objective in making synthetic diamond is to convert the crystal structure of graphite into the crystal structure of diamond or to make diamond grow from diamond without developing as graphite.



DIAMOND "WHISKER," or threadlike single crystal, grown on a facet of diamond in 90 minutes by a pulse-radiation process developed by the authors is 1.5 millimeters long.



NONWHISKER CRYSTAL was also grown by the authors' low-pressure process. The crystal is the round object in the center of the micrograph. Its diameter is 120 micrometers.

ite as the source of carbon atoms, the gas method is more advantageous in that it does not require maintaining a zone raised to a temperature much higher than the temperature needed for diamond to grow. The gas approach is the basis of a method for growing diamond crystals that was discovered in 1956 by one of us (Derjaguin) and B. V. Spitsyn. A similar method is described in a U.S. patent obtained by William G. Eversole of the Linde Division of the Union Carbide Corporation in 1958.

In our method diamond facets were made to grow in an atmosphere of carbon tetraiodide at a temperature of 1,000 degrees C. In Eversole's method the growth of diamond powder was mainly effected with hydrocarbons such as methane. In both methods the molecules of gas, impinging on the heated diamond surface, decomposed and gave off carbon atoms.

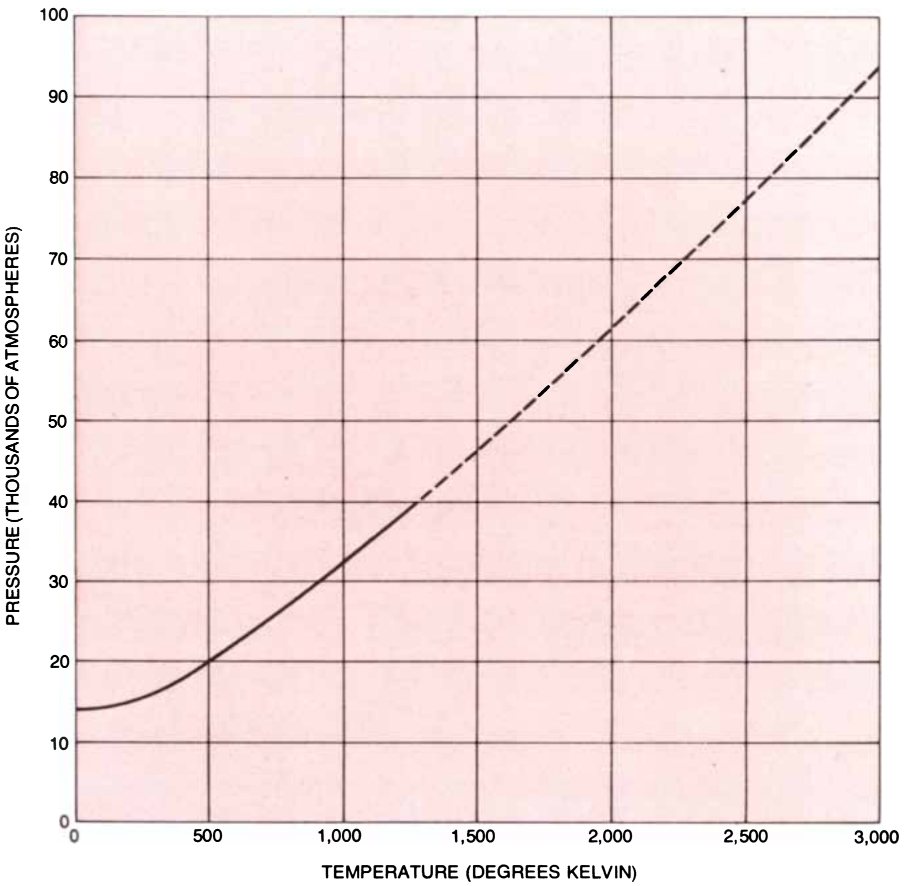
Since the concentration of carbon atoms that can be in equilibrium with graphite is lower than the concentration that can be in equilibrium with diamond (by a factor of about two at 1,000 degrees C.), it is always possible that in the gas method graphite will grow simultaneously with diamond. Graphite nuclei are likely to form in the gas, and they tend to accept the carbon vapor faster than the diamond does. The diamond-growing process must therefore be interrupted from time to time so that the graphite can be removed by chemical means.

There is a fundamental difference between the growth of graphite on a diamond surface and the growth of the diamond. The diamond growth proceeds as the extension of a crystal lattice; one can visualize it as a continuation of the bricklaying operation by which the seed crystal was formed. For graphite crystals to form on diamond, however, the carbon atoms must array themselves in a pattern quite different from that of the underlying diamond. The process is inherently less probable than the growth of the diamond crystal, and its improbability tends to limit the growth of the graphite on the diamond crystal. (This kind of limitation is characteristic of the birth of any new phase, such as the formation of a bubble from a liquid or of a crystal from a gas or a liquid—or of a crystal from another kind of crystal.)

Work by several investigators has evolved a body of theory to explain why the carbon atoms accumulating on the surface of diamond tend to continue the bricklaying operation rather than to form graphite crystals. Nonetheless, if the growth of diamond crystals were continued for several hours, graphite crystals would accumulate on the diamond surface and might eventually cover it, thereby stopping the growth of the diamond. It is as though the bricks were supplied too quickly, so that they could not be laid in the underlying pattern.

There are various ways to avoid the overgrowth of graphite. In the method outlined in the Eversole patent the growth of the diamond is interrupted periodically so that the graphite can be removed by transferring the material into an autoclave containing a hydrogen atmosphere at a temperature of 1,000 degrees C. and a pressure of between 50 and 200 atmospheres. The graphite carbon reacts with the hydrogen much more readily than the diamond carbon does. As a result the graphite is removed, leaving the diamond crystals clean and ready for further growth when they are put back into the carbon-saturated environment.

Our group has found a more efficient process that does not call for high pressure: we oxidize the graphite with the oxygen in air at atmospheric pressure. The entire process of growing diamond out of a carbonaceous gas can thus consist of growing the diamond, together with graphite, followed by removing the graphite by oxidation. Using air to remove the graphite from the diamond surface enables us to carry out both operations in a single reactor.



EQUILIBRIUM CURVE portrays the conditions of pressure and temperature under which graphite will become diamond. The area below the curve represents graphite and the area above the curve diamond. At any given temperature graphite will be converted to diamond at a pressure above the equilibrium pressure. Broken part of curve is extrapolated.

In investigating the growth of diamond in a carbonaceous gas it is advantageous to have the seed diamonds in powder form. The reason is that powder has a fairly large specific surface, that is, a fairly large surface per unit of weight of the material. Many of our experiments on the kinetics of diamond synthesis therefore were done with powder having a particle size of up to one micrometer. Such a diamond powder has a specific surface of about 10 square meters per gram. We directly weigh the material during both the growth phase and the cleaning phase, so that we can tell at all times what is happening to the mass of the sample.

A typical apparatus for one of our experiments has as its principal component a vertical quartz reactor that is heated by a special furnace to a temperature of 1,100 degrees C. Placed inside the reactor is a quartz cup containing a weighed amount of diamond powder. The cup is suspended by a thin quartz thread on a quartz spiral; a change in the extension of the spiral reveals a change in the weight of the powder.

The reactor is evacuated and methane (CH_4) is pumped through it. The temperature in the reactor is simultaneously raised to a predetermined value. After a period during which the diamond particles grow the temperature is lowered and air is admitted to the reactor to remove the graphite that has accumulated. Then the cycle of growth and cleaning is repeated.

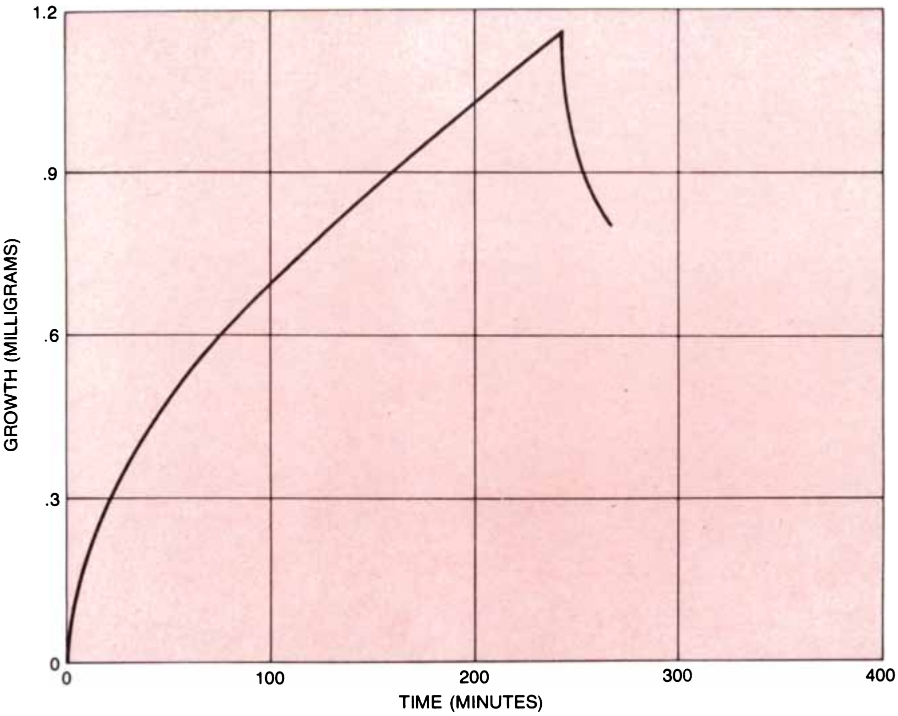
The results of a typical cycle can be portrayed in a chart [*see top illustration on next page*]. In this experiment the buildup conditions included a temperature of 1,020 degrees C. and a methane pressure of .07 torr. Over four hours, which was the duration of the cycle, the gain in weight of the diamond was about 2.7 percent. Five subsequent cycles increased the weight of the starting diamond powder by about 9.5 percent.

The measurements of the gain in weight show that the rate at which the average dimensions of the powder increase is quite low, varying within a range of from several angstroms to several scores of angstroms. If, in order to increase the rate, the temperature or the concentration of gas were increased, the formation of graphite would be accelerated too. As a result the useful buildup time (the amount of time before we have to start oxidizing graphite) would be reduced, and the process would become less efficient from the engineering viewpoint.

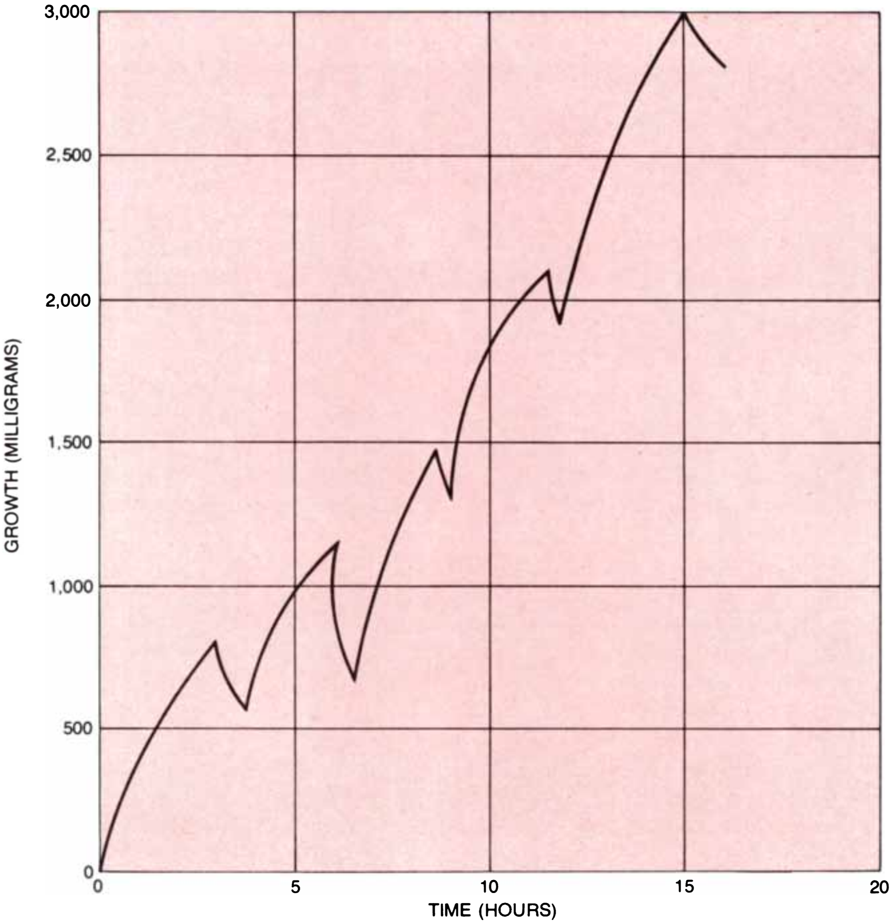
One means of accelerating the process is to apply vibrations to the powder. In this way all the particles of powder are suspended in the gas, so that each particle is more effectively surrounded by methane. This technique leads to a substantial improvement in the growth of the diamond.

We tested our powders in various ways to determine that what had grown was in fact diamond. Analysis by electron diffraction and X-ray diffraction showed that the new material had the structure of diamond. Physical and chemical analyses involving the determination of elemental chemical composition, density and the like also showed that the material was diamond.

Our method of building up fine diamond powder by resorting to the oxidation of graphite yields a considerable gain in diamond weight in a reasonable period of time in spite of the fact that



SINGLE BUILDUP CYCLE in the authors' apparatus is charted. The conditions in the apparatus were a temperature of 1,020 degrees C. and a methane pressure of .07 torr. During the four hours of the cycle the gain in weight of diamond was about 2.7 percent. The point at which the curve drops represents the time when air was admitted to the reactor for the purpose of removing graphite, which invariably accumulates during synthesis of diamond.



DIAMOND GROWTH during five cycles amounted to about 9.5 percent over the weight of the diamond powder with which the buildup was started. Comparable slope of rising portions of the curve shows that the growth rate is restored after each removal of graphite.

the linear rate of growth is quite low. If one undertook to grow larger crystals (say one centimeter in diameter) at the same linear rate, it would take not less than 10 years to double the mass. Fortunately we have been able to develop a number of processes whereby large crystals can be grown at much higher linear rates without the evolution of appreciable amounts of graphite. One of the processes does not differ in principle from the conditions in which we grow diamond powder. Nonetheless, the growth rate for large crystals can be increased to a range of from .01 to .1 micrometer per hour without giving rise to excessive graphite.

It is natural to ask why diamond powder cannot be grown at such high linear rates. One reason is supplied by the theory of crystal growth. To start building each new layer on a facet of diamond requires the formation of a two-dimensional critical nucleus. The probability of the formation of such a nucleus per unit of time is proportional to the surface area. Therefore the mean interval elapsing between the formation of two seeds is inversely proportional to the surface area. For very small seed diamonds, such as the particles of a powder, many faces are so small that considerable time elapses between one nucleation and the next. On larger faces the probability of the nucleation of at least one growth step to keep the deposition of new layers going is much greater.

The relation does not apply indefinitely. When the surface area of a crystalline facet exceeds certain limits, it will either not be coated until a second critical seed appears on it or coating will be prevented because the facet is not perfectly smooth but rather is composed of terraces divided by steps. A terrace will not be coated unless a critical seed forms on it, since the layer of "bricks" forming on one terrace cannot easily climb up or down the steps and go onto a neighboring terrace. Hence the maximum linear growth rate of a facet is equal to the spacing between two adjacent layers divided by the mean time elapsing between the formation of two flat critical seeds on the same terrace. The size of terraces not only is different for different crystals and on different facets but also can vary with time. For example, it can diminish as the crystal grows. In such a case the growth of the crystal slows down.

The blocking of diamond growth by graphite proceeds quite differently. The graphite forms three-dimensional nuclei that are capable of growing both laterally and in thickness. In contrast to the two-dimensional diamond nuclei the three-dimensional graphite nuclei are incapable of coating a facet with a layer one atom thick. Instead the graphite layers form one on top of the other and simultaneously grow at the edges. The time it takes for the graphite crystal to stop or seriously retard the growth of diamond does not depend on the size of the diamond crystal or diminish with its growth. That is probably why in experiments with fairly large diamond crystals (about three or four millimeters across) we succeeded in obtaining a higher linear growth rate than we did with powder. Moreover, with the larger crystals we did not have to supersaturate the gas with carbon atoms to the extent that was necessary with powder, and this lower saturation helped to reduce the evolution of graphite.

More recently we have used a pulsed regime of crystallization to increase the linear growth rate still further without simultaneously speeding up the evolution of graphite. A portion of a diamond facet is heated by focusing on it the radiation from a high-pressure xenon gas-discharge tube. A chopper interrupts the rays regularly, so that the heat is delivered to the facet in pulses. During a pulse the carbon around the diamond becomes highly supersaturated. The diamond grows rapidly, but at the same time many graphite nuclei form. If the pulses are short enough and the dark periods between them are long enough, however, the graphite nuclei do not grow to a critical size. During the dark periods, when the temperature of the diamond and the concentration of carbon atoms near the diamond are dropping, the graphite nuclei tend to revert back to methane. Under the pulse regime the linear growth rate of diamond has reached values as high as micrometers per hour.

Our pulse method has also on occasion given rise to threadlike whiskers of diamond. The rate at which the diamond whiskers grow in length is remarkable, amounting to as much as half a millimeter per hour. Whiskers are a metastable crystalline form of carbon, so that it is surprising they develop at all. Still more surprising is the fact that when they stop growing in length, they start growing in breadth; they eventually become spherical or assume a shape approaching a regular polyhedron. Diamond whiskers hold the promise of contributing significantly to the synthesis of diamond on an industrial scale.

The Role of Pupil Size in Communication

Changes in attitude can be detected by measuring changes in pupil size. It now appears that enlarged or constricted pupils can also affect the attitude and responses of the person who observes them

by Eckhard H. Hess

When we say that someone's eyes are soft, hard, beady, cold or warm, we are in most instances referring only to a certain aspect of that person's eyes: the size of the pupils. The commonplace idea that the eyes can express emotion has been confirmed by experiment. Ten years ago I described in these pages how the viewing of a pleasing image is accompanied by a measurable dilation of the viewer's pupils, and how in general changes in pupil size are objectively correlated with emotions and mental activity [see "Attitude and Pupil Size," by Eckhard H. Hess; SCIENTIFIC AMERICAN, April, 1965]. More recently I have been interested in another aspect of changes in pupil size: the role of such changes in nonverbal communication. I have found that pupil size serves as a signal between individuals, usually at an unconscious level. It is obvious that the eyes play many roles in nonverbal communication, as when someone averts his eyes in talking with someone else. Here, however, I am referring only to the role of the pupil.

The changes in emotions and mental activity revealed by changes in pupil size are clearly associated with changes in attitude. Accordingly the measurement of changes in pupil size, which I have named pupillometrics, has become a useful tool in the study of attitudinal change. In my laboratory at the University of Chicago changes in pupil size are measured while the subject views slides projected on a screen. In order to adapt the subject's eyes to the brightness of the stimulus slide, he is first shown a control slide that has the same average brightness as the stimulus slide that is to follow. The subject views the control slide for 10 seconds and then the stimulus slide for 10 seconds. The difference between the average size of the pupil when the subject is viewing the control slide and its average size when he is viewing the stimulus slide is recorded as the pupil response. One of our methods for measuring pupil size is to photograph the subject's eye during the experiment with a motion-picture camera. Later the film is projected on a screen and the pupil size is measured with a millimeter scale. We have also been using an electronic pupillometer that scans the eye and automatically measures the diameter of the pupil while the experiment is in progress.

The usefulness of pupillometrics in the study of attitudinal change has recently been further validated by the results of a study in my laboratory by Paul W. Beaver. He presented color slides of several kinds of food to 20 people who had missed their previous lunch or dinner and who had not eaten at all for the past five to eight hours. Another group of 20 subjects who had eaten lunch or dinner two hours before were also shown the slides. When the pupil responses of the viewers were measured, it was found that the increase in pupil diameter among the hungry subjects was greater than that among the sated subjects. In fact, in some instances the sated subjects showed a constriction of the pupil. The results demonstrate that even a temporary change in attitude can be detected by measuring pupil response.

Our studies of the pupil as an indicator of attitude led us to consider the possibility that one person uses another person's pupil size as a source of information about that person's feelings or attitudes. In one experiment I showed two photographs of an attractive young woman to a group of men. The photographs were identical except that in one the woman's pupils had been retouched to make them larger and in the other they had been retouched to make them smaller. None of the men reported noticing the difference in pupil size, but when they were asked to describe the woman, they said that the woman in the picture with the large pupils was "soft," "more feminine" or "pretty." The same woman in the picture with the small pupils was described as being "hard," "selfish" or "cold." There could be little doubt that the large pupils made the woman more attractive to the men.

Women used to put the drug belladonna, meaning "beautiful lady," into their eyes because they thought it made them more beautiful. The active principle of belladonna is atropine, which causes the pupils to dilate. Indeed, an eyewash preparation containing atropine was popular among women not many years ago, until the U.S. Food and Drug Administration put a stop to its sale.

Where did the notion that larger pupils make a woman look more attractive come from? It would be easy to dismiss it as mere folklore, but it clearly has a basis in reality. For one thing, younger people have larger pupils than older people, so that large pupils are associated with an obvious constituent of physical attractiveness. What is really appealing about large pupils in a woman, however, is that they are an indicator of interest, which can be interpreted as sexual interest. Moreover, when men view a picture of a woman with large pupils, their own pupils dilate. In other words, seeing large pupils gives rise to larger pupils.

The pupil responses of men and women, all stably married and presumably heterosexual, were investigated by Thomas M. Simms, who was then working at the University of Toronto. The subjects were shown two pictures of a man, one with large pupils and the other

PHOTOGRAPHS OF TWO WOMEN were retouched so that each woman had large pupils in one photograph and small pupils in the other. Male subjects were shown eight different pairs of the photographs: all the possible combinations of the two women. As subjects viewed each pair they were asked in which of the two pictures did the woman appear to be more sympathetic, more selfish, happier, angrier, warmer, sadder, more attractive, more unfriendly and so on. When the question concerned a positive attribute, male subjects tended to choose the photograph of the woman with the large pupils. When the question concerned a negative attribute, they tended to choose the photograph of the woman with the small pupils. Neither woman, however, was consistently chosen as being the more attractive or the more unfriendly. The selection in most instances appeared to be made unconsciously on basis of pupil size.

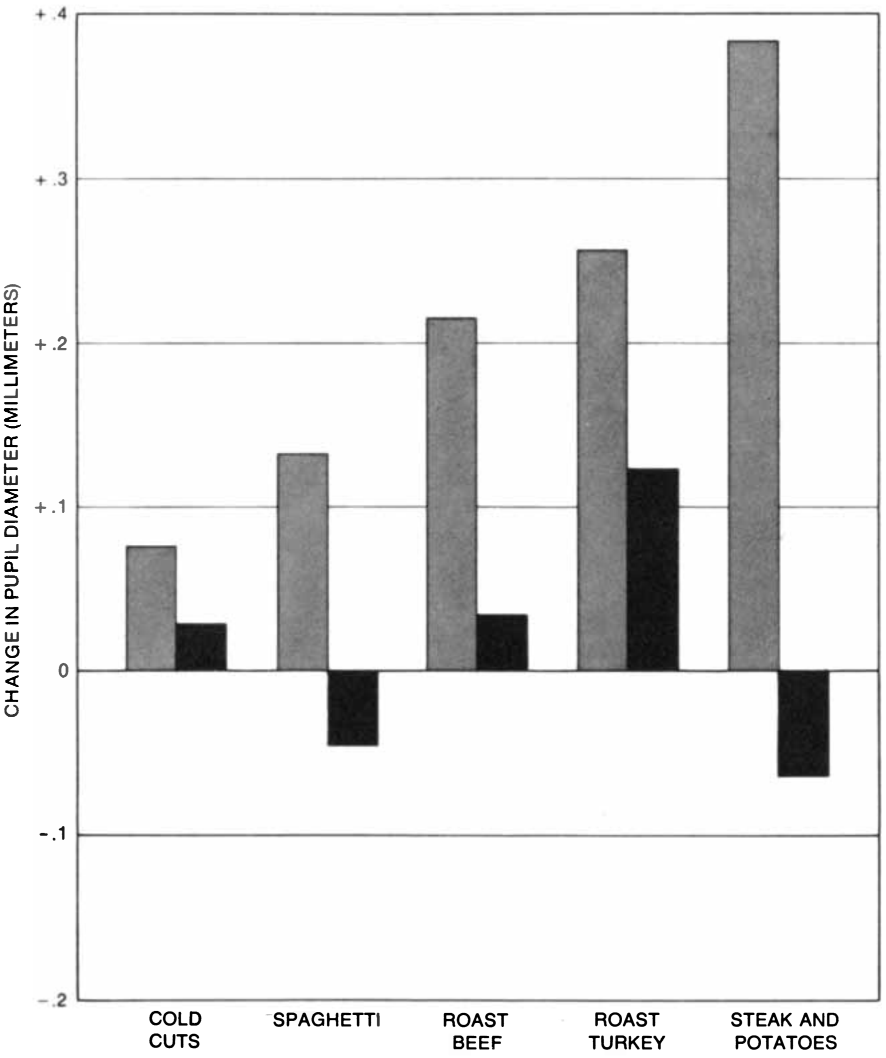
with small pupils. They were also shown two pictures of a woman, one with large pupils and the other with small pupils. The pupils of the male subjects dilated the most when they viewed the picture of the woman with the large pupils. Similarly, the pupils of the women dilated the most when they viewed the picture of the man with large pupils. With both men and women the dilation in response to the picture of a person of the opposite sex with small pupils was much less.

Even more interesting were the pupil responses of the men and women to the pictures of the person of their own sex. The men showed almost no increase in pupil size as they viewed either picture of the man. The women, on the other hand, showed a smaller pupil response to the picture of the woman with the large pupils than they did to the picture of the woman with the small pupils. This finding is supported by the results of another study carried out by Robert A. Hicks, Tom Reaney and Lynn Hill of California State University at San Jose. In interviews with a group of women they found that the women preferred a picture of a woman who had small pupils to a picture of the same woman with large pupils. These findings suggest that women's magazines will not increase their newsstand sales by printing pictures of women with large pupils on the cover.

Additional evidence comes from a study conducted by John W. Stass and Frank N. Willis, Jr., of the University of Missouri at Kansas City. They introduced a subject to two individuals of the opposite sex and asked the subject to select one of them as a partner for an experiment. One of the proposed partners had been given eye drops to dilate his pupils; the other had not. Both men and women tended to choose the person with the large pupils. Stass and Willis also observed that eye contact—a direct exchange of gazes—during the introduction was a factor that increased the likelihood that the individual with the large pupils would be chosen. Most of the subjects were not able, however, to say whether they had used the large pupils or the eye contact as the basis for choosing their partner.

A study of men who identified themselves as homosexuals, conducted by Simms at the University of Toronto, further confirms the effect of pupil size in sexual communication. Simms found that male homosexuals distinctly prefer a picture of a woman with constricted pupils to a picture of the same woman with dilated pupils. Apparently the signal of sexual interest that is transmitted by the dilated pupils of a woman does not appeal to male homosexuals.

Another interesting finding by Simms is that heterosexual "Don Juans," men who identified themselves as being more interested in having sexual relations with many women than in forming a lasting relationship with one woman, have the same pupil response to pictures of women with large and small pupils that male homosexuals do. This finding suggests that such men also have an aversion to women whose pupils indicate sexual interest.



DIFFERENCES IN PUPIL RESPONSE of hungry subjects (*gray bars*) and of sated subjects (*black bars*) to color slides of various foods are shown. The subjects first viewed a control slide, then a slide of a food, and the change in pupil size was measured. The hungry subjects had not eaten at all for from five to eight hours. The sated subjects had eaten a meal two hours before. All the food slides produced dilation of the pupils in hungry subjects, whereas in two instances the slides produced a constriction of the pupils in the sated subjects. The results indicate the pupil response can be a valid measure of attitude change.

In the experiments I have been describing the test pictures were retouched photographs of a complete face. I set out to examine whether or not similar responses might be elicited by a purely schematic pair of eyes. Because of my interest in ethology, the study of the biological basis of behavior, I wondered if large pupils might not act as a "releaser." Ethologists have shown that a small portion of an animal may in itself be sufficient to release a specific pattern of behavior in another animal. For example, a robin will usually attack another robin that intrudes on its territory. David Lack of the University of Oxford found that a robin will also attack a single red feather that is put in an upright position in some strategic place such as a tree limb. The feather, which apparently symbolizes the red breast of a robin, releases the attack behavior of another robin. Could the size of the pupils release a pattern of behavior that is innate or perhaps learned very early in life?

I drew three kinds of schematic eyes that consisted simply of a circle with a

• 40 MPG highway
• 28 MPG city
• Our highest EPA rating

Chevette

That's right.

40 mpg highway, 28 mpg city, based on EPA test results with standard 1.4-litre engine and 4-speed manual transmission. (Remember: These mileage figures are estimates. The actual mileage you get will vary depending on the type of driving you do, your driving habits, your car's condition, and available equipment.)

Chevette is international in design and heritage, with engineering concepts proved around the world.

Its wheelbase is about the same as the VW Rabbit's.

It has more front-seat head room than a Datsun B-210 and more front-seat leg room than a Toyota Corolla.

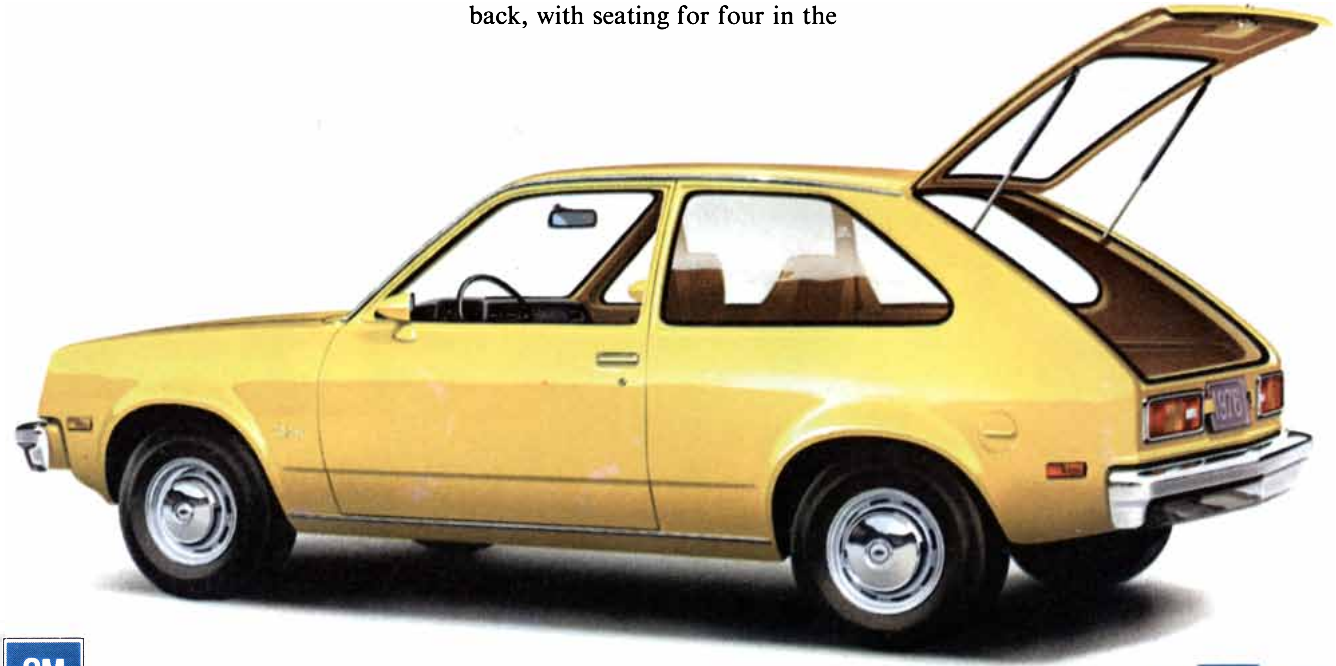
It has one of the shortest turning circles in the world, making it extremely nimble.

Chevette is a two-door hatchback, with seating for four in the Chevette Coupe and its Rally, Sport and Woody versions. (A two-passenger model called the Scooter is also available.)

It is protected by 17 anti-corrosion methods.

In short, it's a new kind of American car. Sold and serviced by all 6,030 Chevrolet dealers from sea to shining sea.

It's about time.



GM
MARK OF EXCELLENCE

Chevrolet

Chevrolet's new kind of American car.

In California, see your Chevrolet dealer for EPA mileage figures and engine/transmission combinations available on California emission-equipped cars.

Why International Paper is helping to develop a 1,000,000-acre forest on land it doesn't own

We want to make sure there'll still be enough wood products around when your children grow up.

Industry sources estimate Americans will use about twice as much paper and wood in the year 2000 as they use today. And the U.S. Forest Service predicts that our nation's commercial timberlands won't be able to keep up with the demand.

One of our solutions is to help private landowners increase their yield. They own about 60 percent of America's forest lands—yet produce only 30 percent of the wood fiber. (Forest products companies own only 13 percent of America's forest lands—and produce 34 percent of the wood fiber.)

We're looking especially to people who own land close to our operations in the South. In 1976 we'll expand our program to the Northeast and West Coast.

How we help landowners

We do it through the Landowner Assistance Program.

We'll show a private landowner how to prepare a site, plant, protect, thin, and harvest—at no charge.

This way, he can get the most from his forest land—in some cases, he can actually *double* his yield.

We'll even find a contractor to do the actual work. Or do the job ourselves at cost.

For this help, IP gets the right to buy a landowner's timber at competitive prices.

We've got more than 300,000 acres in the Landowner Assistance Program now. We're aiming for 1,000,000 before 1980.

A big help. But it's only one thing we're doing to increase the world's wood-fiber supply.

Higher yield from our own lands

We've developed a Supertree—a southern pine that grows taller, straighter, healthier and faster than ordinary pines.

We're experimenting with a new machine that can harvest an entire tree—taproots and all. The roots used to be left in the ground.

We're moving ahead on fertilization techniques. Tree Farm programs. Research.

Will all this be enough to keep the world's fiber supply going strong?

It'll help. But more must be done.

At International Paper, we believe forest products companies, private landowners and the government should work together to develop more constructive policies for managing America's forests.

The wrong policies can make tree farming impossible and force the sale of forest land for other purposes.

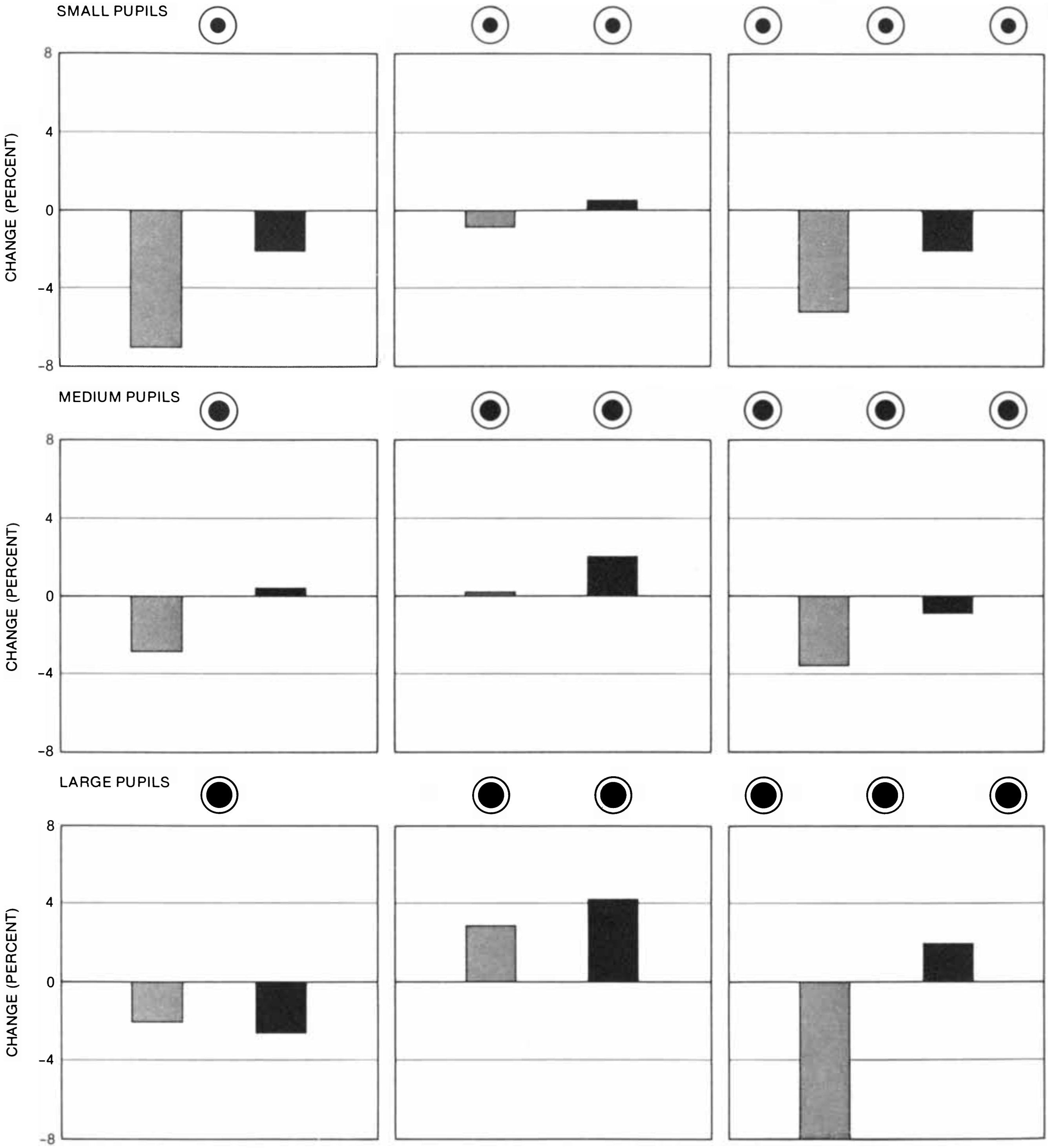
The right policies can assure continuation of America's forests—a renewable natural resource.

If you'd like more information about what has to be done to assure the world's fiber supply, write Dept. 145-A, International Paper Company, 220 East 42nd St., New York, New York 10017.

INTERNATIONAL PAPER COMPANY
220 EAST 42ND STREET NEW YORK NEW YORK 10017

round black area in its center. One set of schematic eyes had small pupils, a second set had medium-sized pupils and the third set had large pupils. I prepared slides that displayed each size of pupil first as a single eye, then as a pair of eyes and finally as three eyes. In the slide showing a pair of eyes the two circles were slightly separated so that they would resemble a pair of eyes in a face. In the slide showing three eyes the spacing of the circles was the same as that of the two eyes of the pair.

I showed the schematic eyes to a group of subjects and observed how they responded in terms of their own pupil size. The responses to the single eye and the triple eye did not vary systematically in relation to the size of the schematic pupils. When I displayed the paired eyes, however, there was a significant



SCHEMATIC EYES with small, medium and large pupils were tested to determine their effect on the pupil size of subjects who viewed them. The pupil responses of men (*gray bars*) and women (*black bars*) to the single eye and the triple eyes did not vary systematically in relation to the size of the pupil in the schematic eyes. The paired schematic eyes, however, produced a significant change in the size of the subjects' pupils. The paired eyes with the largest pupils caused the largest dilation. The finding that schematic eyes with large pupils will cause dilation suggests that observing large pupils releases a pattern of behavior that may be innate.

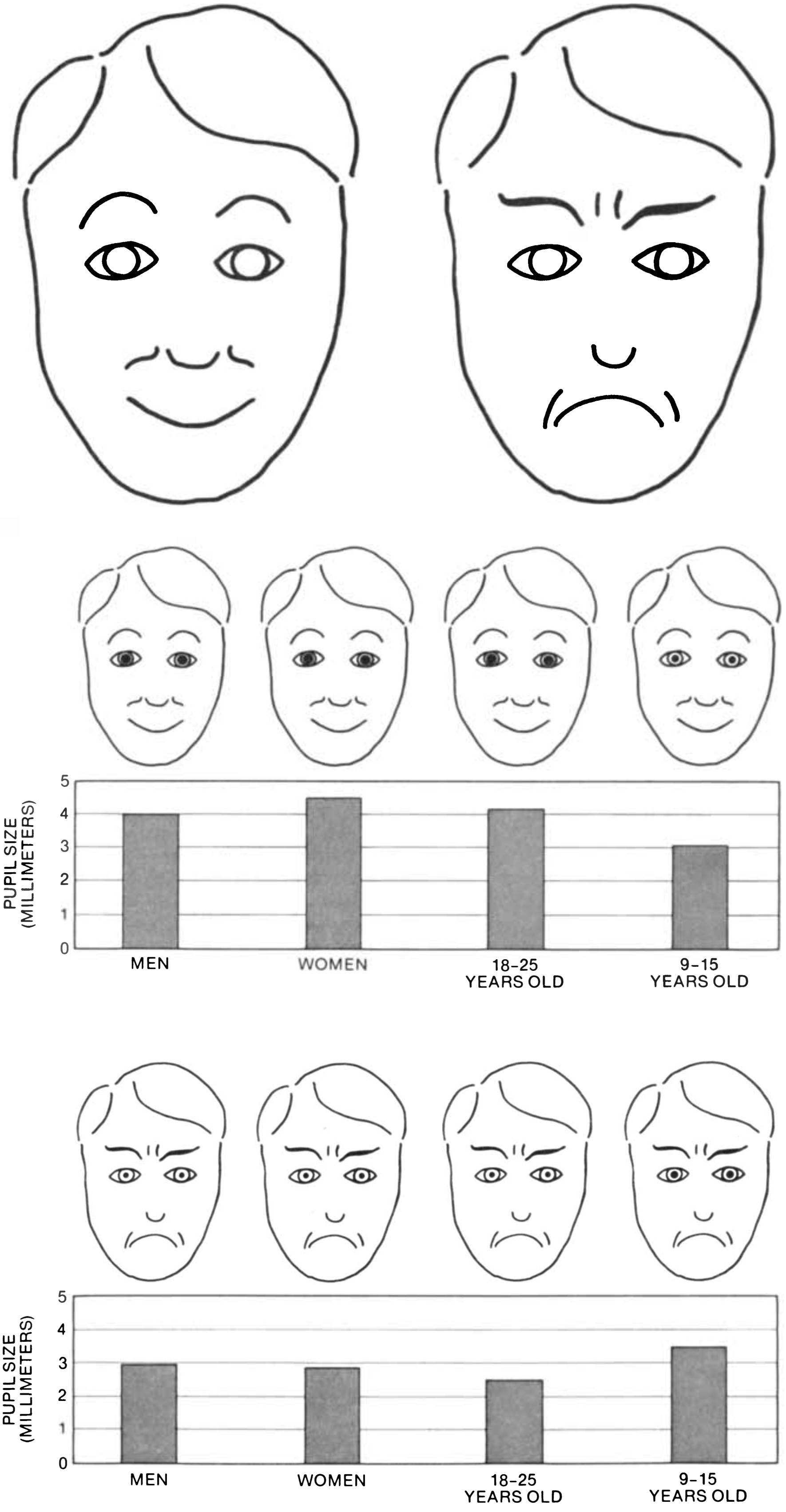
change in the dilation of the subjects' pupils. The paired schematic eyes that had the largest pupils caused the viewers' pupils to dilate much more than paired schematic eyes with smaller pupils did. In female subjects the dilation of the pupils was greater than that in male subjects. The fact that even a pair of schematic eyes will give rise to a dilation of the pupils strongly suggests that the dilation response is innate and not learned.

Similar results were obtained in an experiment with schematic eyes by Richard G. Coss of the University of California at Los Angeles. He, however, got a smaller pupil-dilation response to the paired schematic eyes than I did, possibly because he had placed the schematic eyes so close together that they did not resemble a pair of eyes in a face.

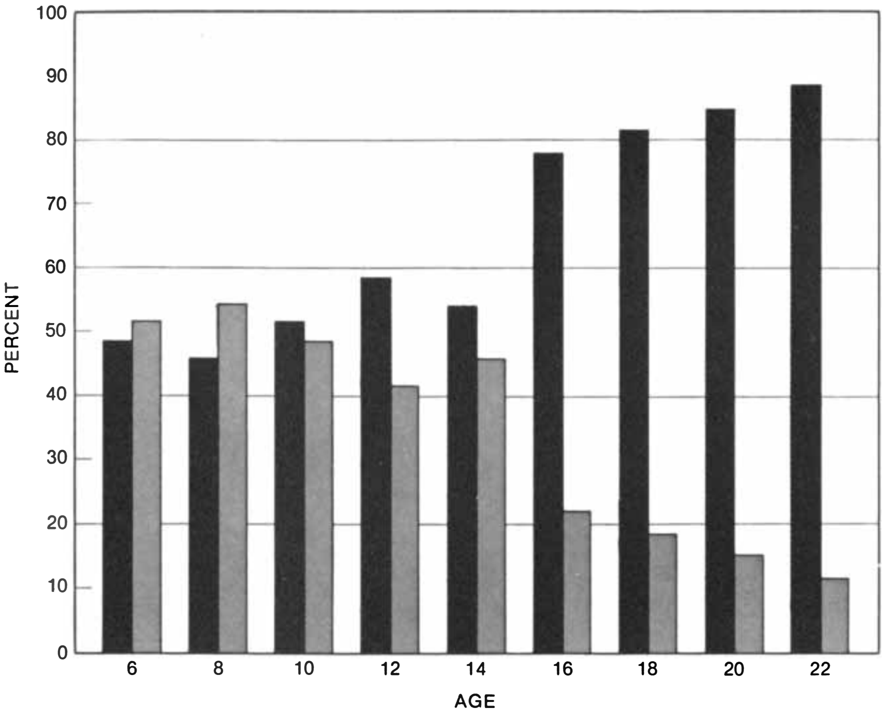
In children the absolute—not relative—size of the pupils is larger than it is in adults. Whatever other reasons there are for this difference, having large pupils is probably advantageous to a child in that it makes him more appealing to the adults who take care of him. Very young children have other features that may release the caring response in adults. An infant's head is large in proportion to his body. His eyes are large and are located below the middle of his face. His limbs are short and fat. The "lovable" cartoon characters created by Walt Disney and other artists tend strongly to have these babyish features and large pupils.

While studying the behavior of infants Janet Bare Ashear of the University of Chicago noticed that she seemed to elicit more smiles from infants than some of her fellow workers did. It turned out that in average room lighting her pupils were larger than those of most other people. To find out if large pupils in an adult do affect the smiling behavior of infants, she arranged to visit 16 infants in their homes. The infants were between three months and three and a half months old, an age when they smile at adults and have not yet developed a fear of strangers. Ashear made two visits to each home, and on each visit she interacted with the infant, talking and smiling, and recorded the number of times the infant smiled. On one of her visits she had her pupils artificially dilated with the drug phenylephrine hydrochloride. On the other visit her pupils had been artificially constricted with another drug, pilocarpine hydrochloride.

The infants smiled more often when Ashear's pupils were dilated than they did when her pupils were constricted. Although the experiment was only a pi-



DRAWINGS OF TWO FACES WITHOUT PUPILS were given to subjects who were asked to draw in the size of pupil that best fits the face. One face was smiling and the other was scowling. Men and women drew larger pupils on the happy face than on the scowling face, and so did a group of college students ranging in age from 18 to 25 years. Younger people between the ages of nine and 15 tended to draw pupils of about the same size on both faces, which indicates that they do not attribute different meanings to pupils of different sizes.



AGE DIFFERENCES in perceiving a face with large pupils as being happier than a face with small pupils were found in study of individuals ranging in age from six to 22. The subjects were shown drawings of two faces and were asked to choose the happier one. Subjects up to the age of 14 were just as likely to choose the face with the small pupils (*gray bars*) as the one with the large pupils (*black bars*). Subjects who were 16 years of age or older, however, strongly tended to choose the face with the larger pupils as being the happier one.

lot study and was obviously open to experimenter bias, the results nonetheless suggest that in infants the positive response to large pupils may not be learned but is part of the infants' perceptual development. Perhaps large pupils act as a releaser in infants as well as in adults.

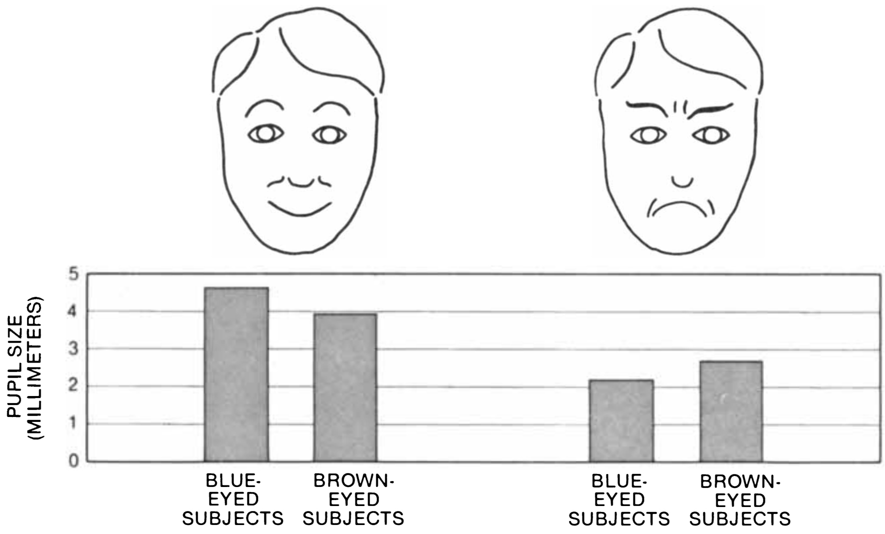
An unexpected result of Ashear's pilot study was the reactions of the mothers. When the experimenter's pupils were constricted, the mothers said that she appeared to be "harsh," "hard," "brassy," "cold," "evasive" and "sneaky." One of the mothers said that Ashear appeared to be trying to hide something from her. When Ashear came with dilated pupils, she was described as being "naïve," "young," "open," "soft" and "gentle."

Children's books sometimes have illustrations that make use of different pupil sizes to depict good and bad characters. One finds that the "wicked witch" has tiny pupils and the "beautiful princess" very large pupils. There is an interesting difference between older books for children and more recent ones. Many of the older books have illustrations in which the pupils of the characters are rendered in various sizes; many of the newer books have illustrations in which the pupils of the characters are the same size.

I have conducted an experiment with drawings of faces that have no pupils. The faces were drawn about three-fourths the size of an average adult face. One face was smiling and the other was scowling. When I gave these faces to 10 men and 10 women and asked them to "draw in the size of pupils that you think best fits the face," I found that 15 of 20 subjects drew larger pupils on the happy face than on the scowling face [*see illustration on preceding page*].

I also had students in several of my classes draw pupils on the same two faces. The results were unequivocal: 47 of 50 students drew larger pupils on the happy face than they did on the scowling one. When I tested a group of younger people, between the ages of nine and 15, I found that they tended to draw the pupils on both faces the same size; in fact, the pupils on the scowling face were on the average slightly larger.

One of my students, James Dickson McLean, investigated the response of individuals ranging in age from six to 22 to drawings of faces that had pupils of different sizes. In one experiment the subjects were asked to choose the "happier" of two female faces. One face had large pupils and the other had small ones. McLean found that up to the age of 14 a person does not necessarily perceive larger pupils as being happier than smaller pupils. His finding agrees with the results I obtained in my study of youngsters between the ages of nine and 15. McLean concluded that the turning point for attributing different meanings to pupils of different sizes comes at about the age of 14. It may be, however, that the answers given by children reflect not their actual perception but their understanding of the question (or lack of understanding). By testing the pupil responses of children to pictures of faces with large pupils and small pupils, we



BLUE-EYED SUBJECTS drew larger pupils on a sketch of a happy face and smaller pupils on a scowling face than brown-eyed subjects. In addition, when viewing a picture that normally causes dilation or constriction, blue-eyed people show a greater change in pupil size.

may find they have the same dilation responses that adults do.

Of particular interest was another finding by McLean: blue-eyed subjects were more likely to judge large pupils as being happy than brown-eyed subjects. When we asked another group of subjects to fill in the pupils on drawings of happy faces and angry ones, we found that the blue-eyed subjects drew larger "happy" pupils and smaller "angry" pupils than the brown-eyed subjects [*see bottom illustration on opposite page*]. We also found that blue-eyed people have a stronger pupil response than brown-eyed people when they view a picture that causes pupil dilation or constriction. To be more precise, with respect to the total range of response from the smallest pupil size to the largest the range is greater for blue-eyed people than it is for brown-eyed people. (This statement applies, of course, only to changes in pupil size resulting from emotions or attitudes.)

I shall conclude with the results of a study we have just completed. We took two identical photographs of a woman and in one made the pupils large and in the other made the pupils small. The same was done with a second pair of photographs of another woman. We showed these photographs in pairs to a group of male subjects. As the subjects viewed each pair they were asked which of the two women was more attractive, more selfish, happier, more unfriendly and so on. The subjects were shown eight different pairs of photographs, that is, all the possible combinations of the two women.

Neither woman was consistently chosen as being the more attractive or the more selfish or whatever. The subjects, however, strongly tended to choose the photograph with the large pupils when the question concerned a positive attribute. When the question concerned a negative attribute, the subjects tended to choose the photograph with the small pupils. The tendency to associate positive attributes with large pupils and negative attributes with small pupils was stronger in blue-eyed subjects than in brown-eyed subjects.

Why do blue-eyed people respond more to large and small pupils than brown-eyed people? It is of course easier to see a pupil surrounded by a blue iris than it is to see one surrounded by a brown iris. Perhaps it is not unwarranted to assume that the response has been favored by evolutionary selection more in blue-eyed people than in brown-eyed people.

MATHEMATICAL GAMES

On map projections (with special reference to some inspired ones)

by Martin Gardner

"Similar scenatas are projectilised from Hullulullu, Bawlawayo, empyreal Raum and mordern Atems."

—James Joyce, *Finnegans Wake*

If only the earth were flat, as Martin Luther, John Calvin and the fathers of the Catholic Church believed, the labors of cartographers would be greatly simplified. Indeed, flat-earth maps of the first six centuries of the Christian Era presented no serious geometrical problems. A few learned churchmen agreed with Pythagoras, Plato, Aristotle and Archimedes that the earth was round, but to most churchmen such a belief was heresy.

Early medieval maps of the world were "Scripture-preserving." They were either rectangular to preserve the "four corners" in Isaiah 11:12 and in Revelation 7:1, or they were circular or oval to preserve the "circle of the earth" in Isaiah 40:22. There was, of course, no need for meridians and parallels. Jerusalem was exactly in the center, as Ezekiel 5:5 suggests. The top of the map pointed east and included the site of Eden. The land was surrounded by the "great waters" that had once flooded the earth and also by the sources of the "four winds" (Daniel 7:2, Revelation 7:1) that blew so erratically toward the Holy City.

After the eighth century the rotundity of the earth gradually became acceptable to the Church, with such eminent Catholics as Thomas Aquinas and Dante Alighieri defending it. It slipped out of favor among the Reformers, but during the Renaissance it quickly won the day. The rapid increase in travel and exploration, particularly the great sea voyages, made it necessary to have better maps, and that naturally revived a troublesome mathematical question: How can a portion of the earth's surface be drawn on a plane so that all distances are accurately represented?

The answer is that it cannot. The side of a cylinder or cone will map perfectly onto a plane, but the surface of a sphere will not. You can flatten a cylinder or cone without distorting its surface, but even a small region of a sphere's surface will not press flat without cracking, folding or stretching. Every flat map of all or part of the earth distorts something. The cartographer's tricky task is to design maps that will show the least distortion or no distortion of those properties the map's user deems desirable. At the same time the distortion of all other properties should be minimal. We shall take a quick look at some classical methods of mapmaking before turning to methods that result in more bizarre maps.

One of the most desirable features of a map is that angles between any two lines on the map be the same as the angles between those same lines on the globe. This feature is enormously useful at sea because it means that observed angles between two landmarks correspond to angles measured on the map with a protractor, and also because small regions on such a map preserve their shape. Maps of that type are called conformal. The simplest way to produce a conformal map is by "stereographic projection." As the top illustration on the opposite page shows, a sphere is projected by straight lines from point *B* on the sphere's surface to a plane tangent to the sphere at a point opposite. The projection is called equatorial, polar or oblique depending respectively on whether the antipodes are on the equator, the poles or somewhere else. The price paid for conformality is a distortion of the scale factor that increases with distance from the center of the map.

If the projection to the tangent plane is made from the globe's center, it is a gnomonic projection, so called because it is related to the construction of a sundial with a gnomon. Every great circle on the globe becomes a straight line on a gnomonic map. The map is not conformal, but for navigators it has one merit that all other planar projections lack. A straight line between any two points on a gnomonic map corresponds to a great circle on the globe and therefore provides the geodesic: the shortest distance between the two points.

Since the point of projection can be at any spot inside or outside the globe, there is an endless variety of perspective projections. If the point is at infinity (all projecting lines parallel), the projection is orthographic. Our view of the moon or a view of the earth from the moon is essentially orthographic. Distance distortions are great at the edges of an orthographic map. The map preserves neither area nor angles, but if it is skillfully drawn, it gives a strong illusion of the earth's roundness. Perspective maps with the "eye" above the earth may be among the least accurate with respect to many properties, but they are the most accurate in matching our visual perceptions of a sphere.

Projections need not be made onto a plane. They can be made onto surrounding cylinders or cones that can then be cut and unrolled. Imagine the earth snugly fitted inside a cylinder. The projecting lines are parallel to the plane that cuts the great circle where the globe and the cylinder touch [*see bottom illustration on opposite page*]. The resulting map has the amazing property, highly desirable for many purposes, that areas are preserved: all closed curves have the same areas as their corresponding curves on the globe, and they have them in scale. If the cylinder touches the earth along the Equator, all meridians and parallels on the map become straight lines meeting at right angles.

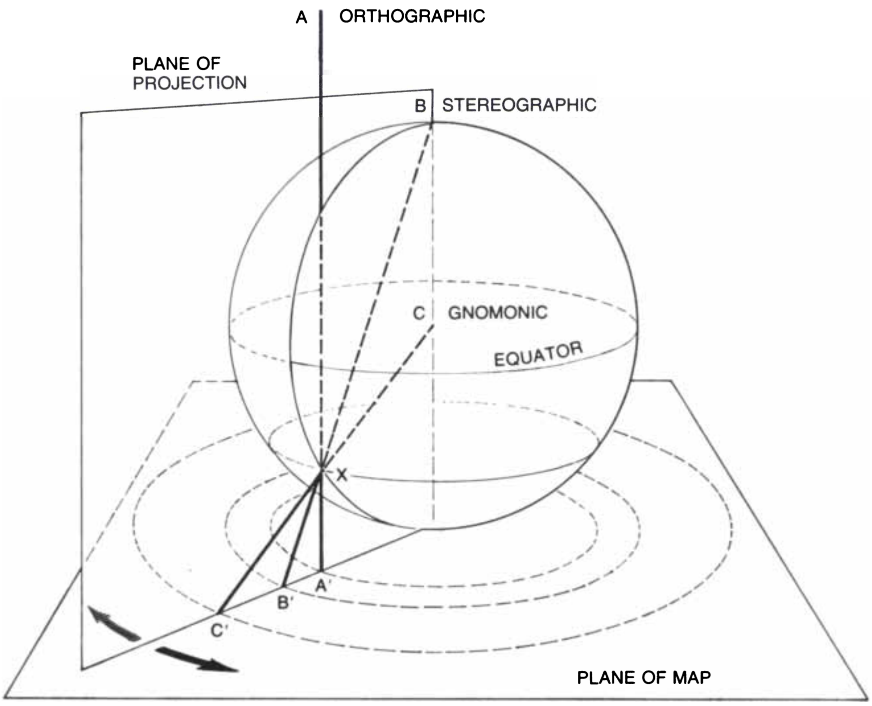
The equal-area cylindrical map is not conformal, and it severely distorts shapes and distances. Indeed, it is not hard to prove that no map can simultaneously be conformal and area-preserving. A great variety of other area-preserving maps have been devised. In modern atlases one of the most popular of equal-area world maps is an elliptical projection worked out by Karl B. Mollweide in 1805.

The cylindrical projection suggested to Gerhardus Mercator, a 16th-century Flemish geographer, the famous conformal map that bears his name. Imagine the earth's surface punctured at the poles, the two holes enlarged until the surface is a cylinder, the cylinder stretched along its length until the map is conformal, then cut along a meridian and unrolled. There is an enormous distortion of scale near the poles. As we learned in grade school, this familiar world map shows Greenland larger than South America when actually it is nine times smaller. (In order to minimize such scale variations modern atlases use a modification of the Mercator map called the Miller projection.) The Mercator projection has, however, one remarkable property that makes it invaluable to navigators. If you rule a straight line between any two points on the map, the line is a loxodrome, or rhumb line, connecting the two points. A loxodrome is a line that keeps a constant angle with parallels and meridians [*see top illustration on next page*]. Imagine a point on the globe that starts at the Equator and moves northward in any constant compass direction. The path will be a loxodrome that spirals toward the North Pole, finally strangling the pole after an infinity of turns around it. On a stereographic map (with its plane tangent to the North Pole) a loxodrome projects as a logarithmic spiral.

The loxodrome is not the shortest distance between two points, but for small distances it is reasonably close to a geodesic, and it has the practical value of being a path that does not require constant changes of bearing. For long distances navigators usually determine the geodesic from a gnomonic projection, then break it into shorter rhumb lines on a Mercator map to minimize changes in compass settings.

So much for the classic projections. Let us turn now to more radical distortions. For a two-point equidistant projection, points *A* and *B* are selected. The map is then drawn so that all distances from *A* and *B* to any other point on the map are in true scale. Such a map is useful to a person traveling from *A* to *B*. No matter how circuitous his route is he can at any time measure with a ruler exactly how far he is from both points. Suppose the two points on a two-point equidistant map of the world are the two poles. What will the map look like? Next month I shall give the answer.

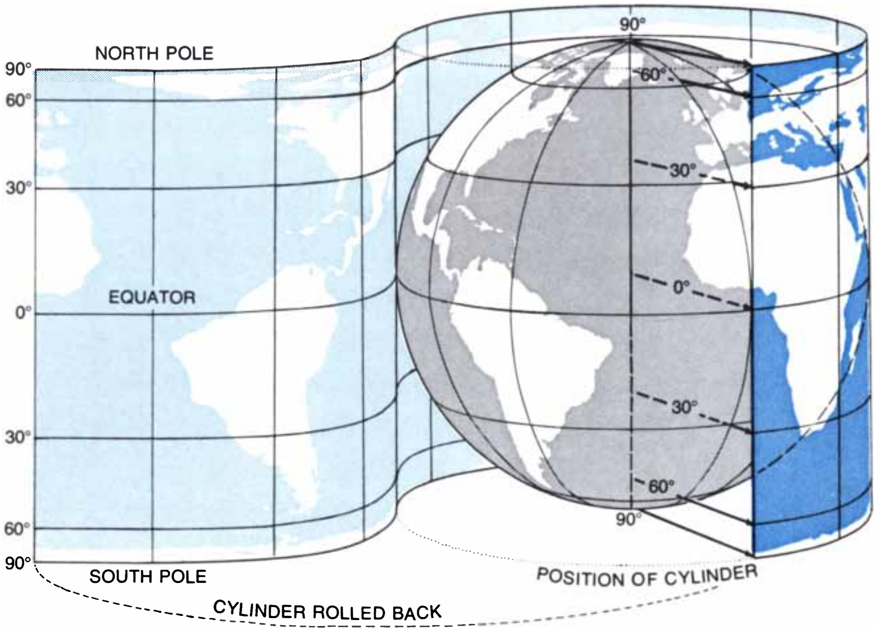
Another curious special-purpose map is the "Mecca map," which is designed to show at once to a Moslem the exact direction he must face when he prays at any spot on the globe. One way to draw such a map is to make a stereographic projection with Mecca at the plane's tangent point. Because the map is conformal Mecca's bearing angle can be determined by measuring the angle between a straight line to Mecca and a meridian. Unfortunately the meridians on such a map are curved, making it difficult to measure the angle exactly. One can, however, construct a Mecca map on which all meridians are straight lines, making it possible to measure



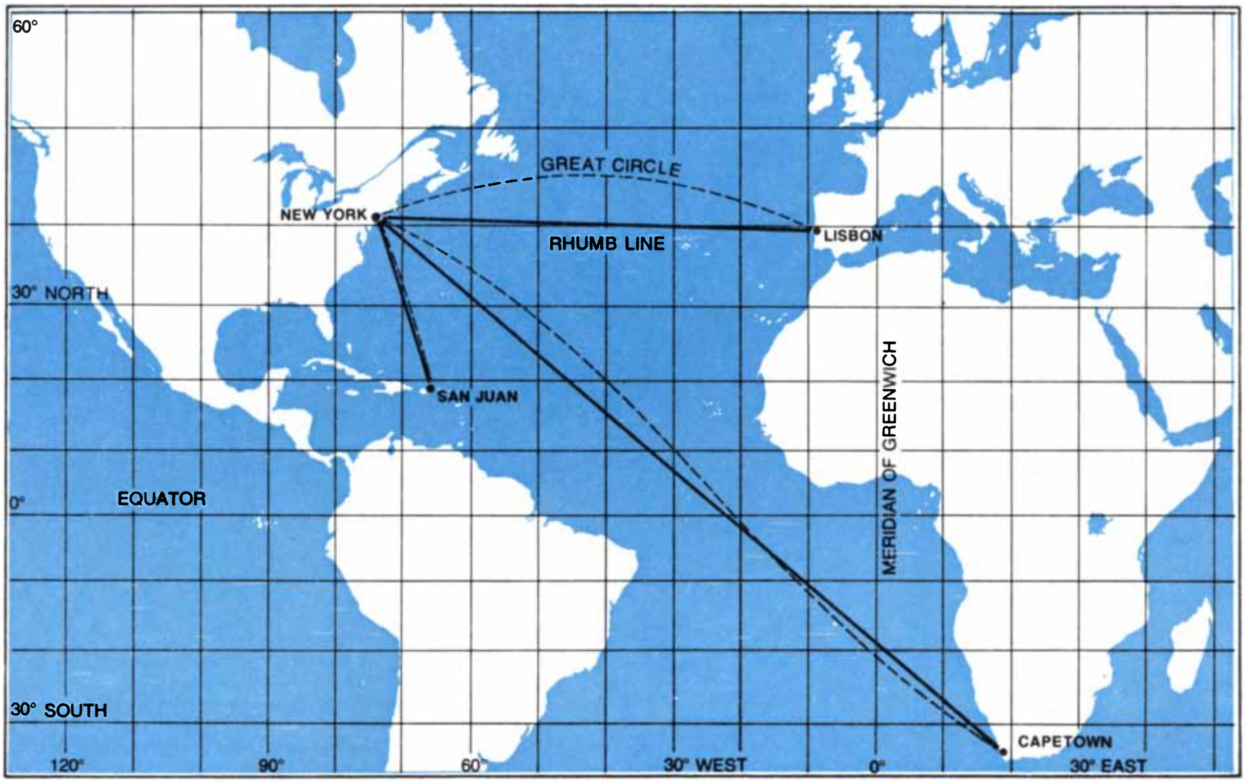
Three azimuthal projections: orthographic (A), *stereographic* (B) *and gnomonic* (C)

bearing angles with a protractor. Such a map, with Mecca replaced by another holy place, Wall Street, is given in an internal Bell Laboratories memorandum on map oddities written by Edgar N. Gilbert, a mathematician [*see bottom illustration on next page*]. The map's upper boundary is the North Pole.

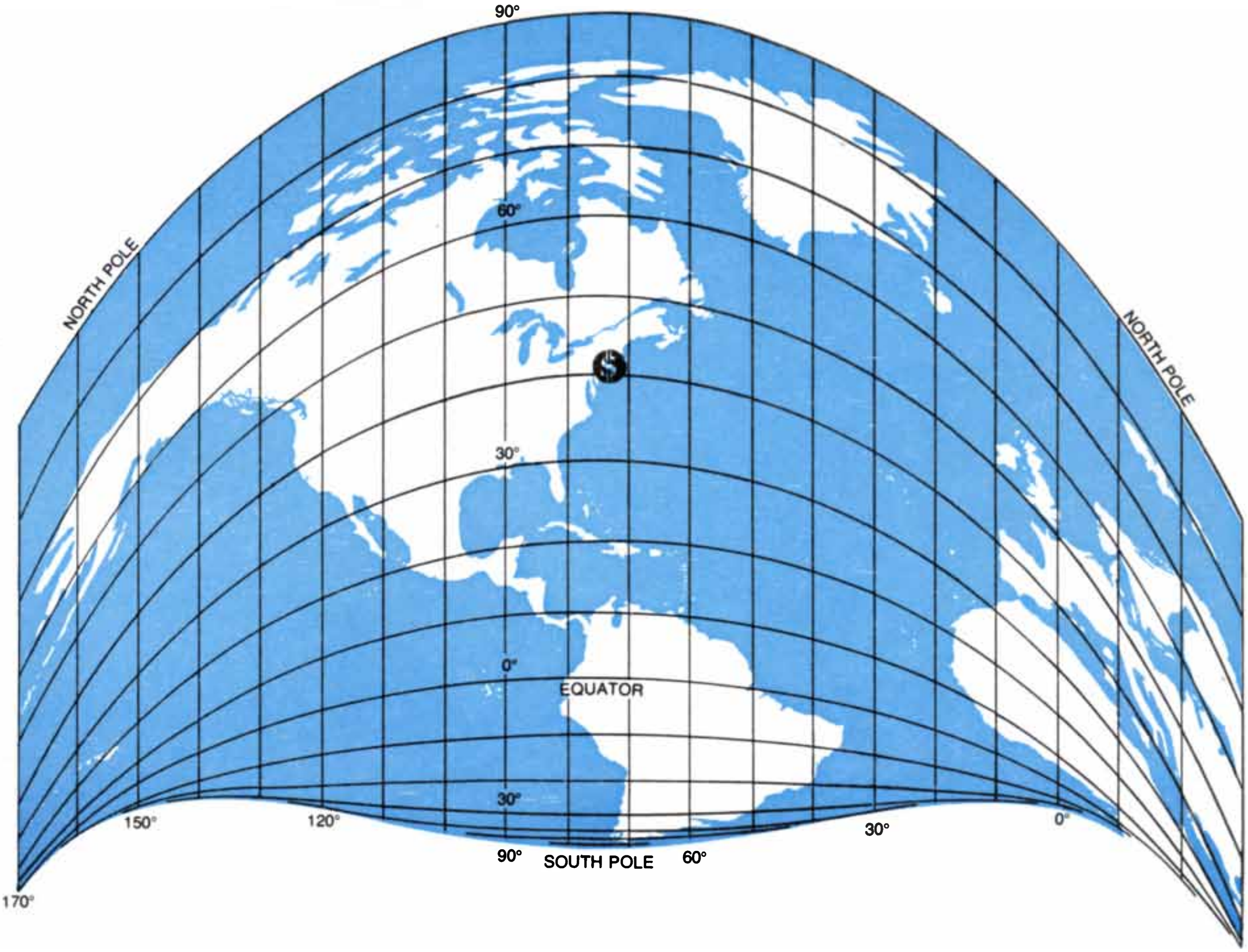
Gilbert's memorandum contains even stranger maps. One is a cordiform (heart-shaped) equal-area map invented by Johann Werner [*see illustration on page 123*]. It was popular in the 16th century but, writes Gilbert, has "now fallen into undeserved obscurity. The highly distorted parts of the map lie away from most major land masses. The curved latitude lines give the map a pleasing illusion of roundness.... The latitude lines are arcs of circles, evenly spaced and centered at the North Pole. The longitude lines are drawn to make



Cylindrical projection method for making an equal-area map



A Mercator conformal map with loxodromes, or rhumb lines, from New York

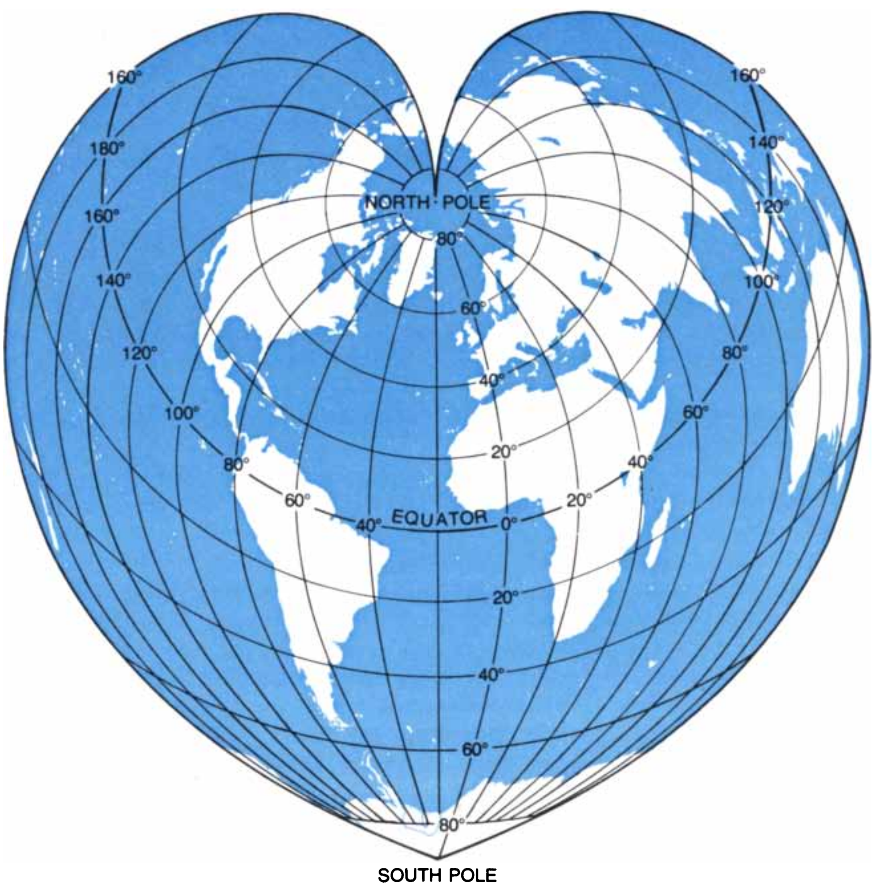


Edgar N. Gilbert's "Mecca map" for Wall Street distances along the latitude lines the same in the map as in the sphere."

The bunching of land masses on Werner's map reflects an actual bunching of land on the earth. The Pacific Ocean is so immense that if you look at a globe from a point above the English Channel, you will see more than 80 percent of the world's land, and the hemisphere opposite will be almost entirely water. Is such lopsidedness surprising? Gilbert came up with an answer by replacing the continents with small nonoverlapping circular caps. Assuming that N such caps are randomly distributed around the globe, what is the probability that the centers of all N circles lie in one hemisphere? In a paper titled "The Probability of Covering a Sphere with N Circular Caps" (*Biometrika,* Vol. 52, 1965, page 323) Gilbert shows that the probability is $2^{-N}(N^2 - N + 2)$. With $N = 7$ continents the probability is 11/32, proving that the earth's lopsided land distribution is not at all remarkable. The reader may enjoy testing the formula by determining the probability that the centers of one, two or three continents all lie in one hemisphere.

The strangest of all Gilbert's maps was made by taking a conformal world map based on a conical projection and projecting it back onto a sphere to produce a conformal "two-world map." The illustration on the next page shows how the globe looks in perspective from a point about five radii from its center. When people visit Gilbert's office, he likes to ask them what is wrong with his globe. If the visitor cannot see what is wrong, Gilbert gives the globe one slow, complete turn. "Even this hint," he writes, "does not always succeed." Actually every spot on the globe has a duplicate on the other side! Unless you are an experienced geographer, however, it is not easy to realize that you are seeing much more of the world than can normally be seen on one hemisphere.

With the aid of computer-generated graphics it is now possible to write programs that distort a map so that the areas express some desired value such as annual rainfall, retail sales and so on. Regions on the map still retain recognizable shapes in spite of the distortions. The joke map showing a New Yorker's idea of the U.S., although it goes back (in many variants) to precomputer days, is a familiar example of such maps. One of the top experts on such special-purpose map projections is Waldo R. Tobler, a geographer at the University of Michigan. In a paper titled "A Continuous Transformation Useful in Districting" (*Annals of the New York Academy of Sciences,* Vol. 119, 1973, pages 215–220) he explains a computer program that distorts a map to show relative populations of regions by their relative sizes, and he shows how such a technique could be a valuable aid in planning voting districts. In 1973 a geographer with a fondness for floppy bow ties was given a presentation award that consisted of a framed world map distorted to the shape of his tie. Drafting the map was no problem for Tobler's Calcomp 763 program.



Johann Werner's cordiform (heart-shaped) equal-area map

Dissection maps (as I call them) are world maps projected on a pattern of squares, triangles or polygonal tiles of other shapes. The tiles can be fitted together to make an "interrupted map" (a map with discontinuities) of any portion of the globe. The philosopher and mathematician Charles Sanders Peirce designed such a conformal map. It is a projection on eight isosceles right triangles that may be regarded as the faces of an octahedron that has been flattened until a space diagonal is zero. The vertexes of the zero diagonal are the North and South poles of Peirce's map (see "Quincuncial Projection of the Sphere," *American Journal of Mathematics,* Vol. 2, 1879, pages 394–396; reprinted in Peirce's *Collected Papers,* Vol. 7).

B. J. S. Cahill of Oakland, Calif., patented his butterfly map in 1913, and it enjoyed a considerable vogue in the 1930's. The world is projected onto the eight equilateral-triangular faces of a regular octahedron. Cahill had several versions of his map, based on different projections, but they all consisted of eight triangular tiles that could be fitted together as one pleased [*see top illustration on page 125*].

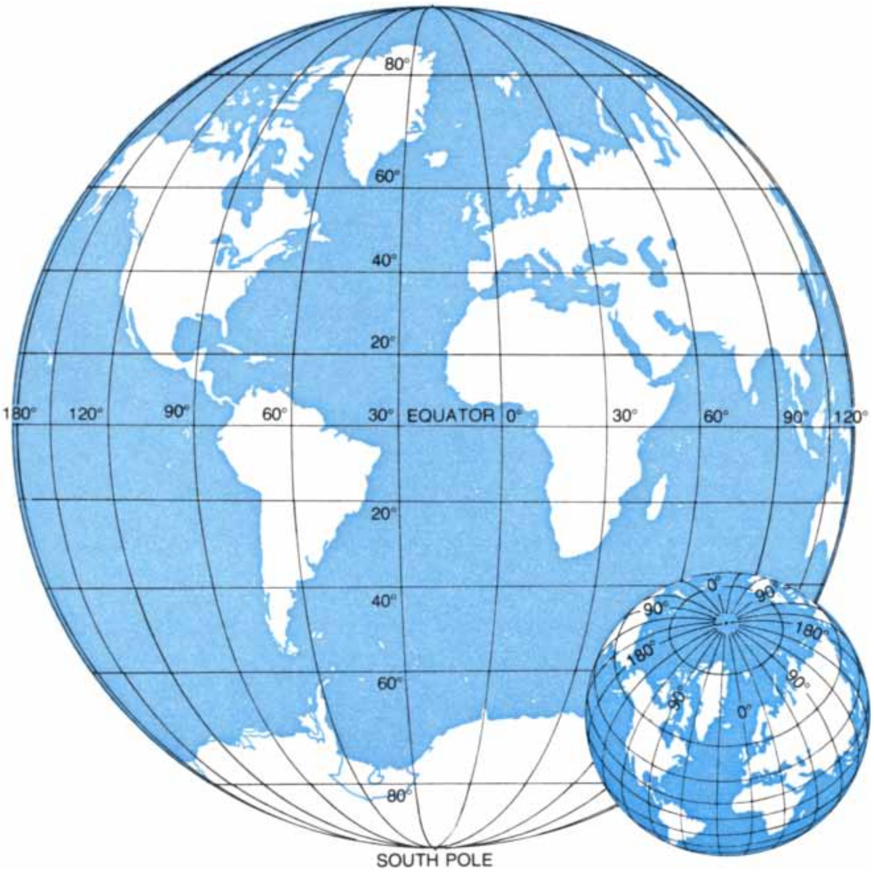
R. Buckminster Fuller's first Dymaxion map was a projection of the world onto the 14 faces (six squares and eight equilateral triangles) of a cuboctahedron. Gerard Piel, who was then science editor of *Life,* was so intrigued by it that he asked the cartographer Richard Edes Harrison to draw an unfolded net of the solid. Staff artists completed a color map of the drawing, and it was published on cover-stock paper in the March 1, 1943, issue of *Life.* It was a great success. All over the country, in homes and laboratories, one would see the little cuboctahedron hanging on a cord and rotating with currents of air.

At about the same time Irving Fisher, a distinguished Yale economist, thought of a similar idea: a gnomonic projection of the world to the 20 triangular faces of an icosahedron. The bottom illustration on page 125 shows an unfolded net of this Platonic solid as it appears inside the jacket of *World Maps and Globes,* an entertaining introduction to cartography by Fisher and O. M. Miller, published by Essential Books in 1944. Harrison was the map's cartographer.

The reader may enjoy pasting Fisher's gnomonic projection on heavy paper and folding it into the icosahedral globe or cutting out the triangles and fitting them together in different ways. If the poles are put at opposite vertexes, the Equator becomes a straight line. Fisher marketed several versions of his Likaglobe, as he called it, and wrote an article about its merits in *Geographical Review* for October, 1943. The globe can of course be projected on other regular solids, but distortion is exaggerated on the cube and the tetrahedron. Although the dodecahedron makes a handsome pseudoglobe, its faces cannot be used as tiles because regular pentagons will not tessellate the plane.

The icosahedron seems to be the ideal polyhedron for a dissection map, and Fuller himself has now adopted it. In 1954 he copyrighted his Dymaxion Sky-ocean Projection World Map, drawn by Shoji Sadao. It differs from Fisher's Likaglobe in having the North and South poles on opposite faces at points slightly off center. For $1.25 (which includes postage and handling) readers can obtain a punch-out version printed in four colors on heavy stock. It folds into a beautiful icosahedron that rests on a cardboard stand that comes with the map. Orders should go to R. Buckminster Fuller, Dymaxion Maps, 3500 Market Street, Philadelphia, Pa., 19104. Also available are several wall-map versions of the unfolded icosahedron and an explanation sheet on which Fuller discusses the philosophy and mechanics of the projection.

We have considered only a small portion of the many curious special-purpose maps designed by ingenious cartographers. Harrison once drew a world map consisting of nothing but ocean shipping lanes. From a distance you could see the continents clearly, but from close up you discovered that the map contained not a single shoreline. At the Christian Science Publishing House in Boston you can step inside a world globe 30 feet in diameter. If such a globe were transparent and viewed from the outside, everything would be mirror-reversed. From the inside the land masses and oceans appear normal.



Gilbert's "two-world" globe and an oblique view of the globe

The major tenet of a bizarre little sect called Koreshanity, which flourished in America late in the 19th century, was that we actually live on the under surface of such a hollow earth, with the entire cosmos inverted and compressed to fill the interior. The sect was founded by a baptist named Cyrus Reed Teed of Utica, N.Y. You will find Teed's inside-out maps in his great scientific work *The Cellular Cosmogony* (1870) and in the pages of *The Flaming Sword,* a periodical that the cult kept going until 1949. How did Teed explain the fact that we cannot point a telescope straight up and see the earth's other side? Well, he took care of it with special optical laws that stated light travels in bent paths. Teed's cosmos has not been sufficiently appreciated by philosophers of science. By performing an inversion operation on the universe, then inventing new laws of physics, one can obtain an inside-out cosmology that is not easy to refute except by slashing it with Occam's razor.

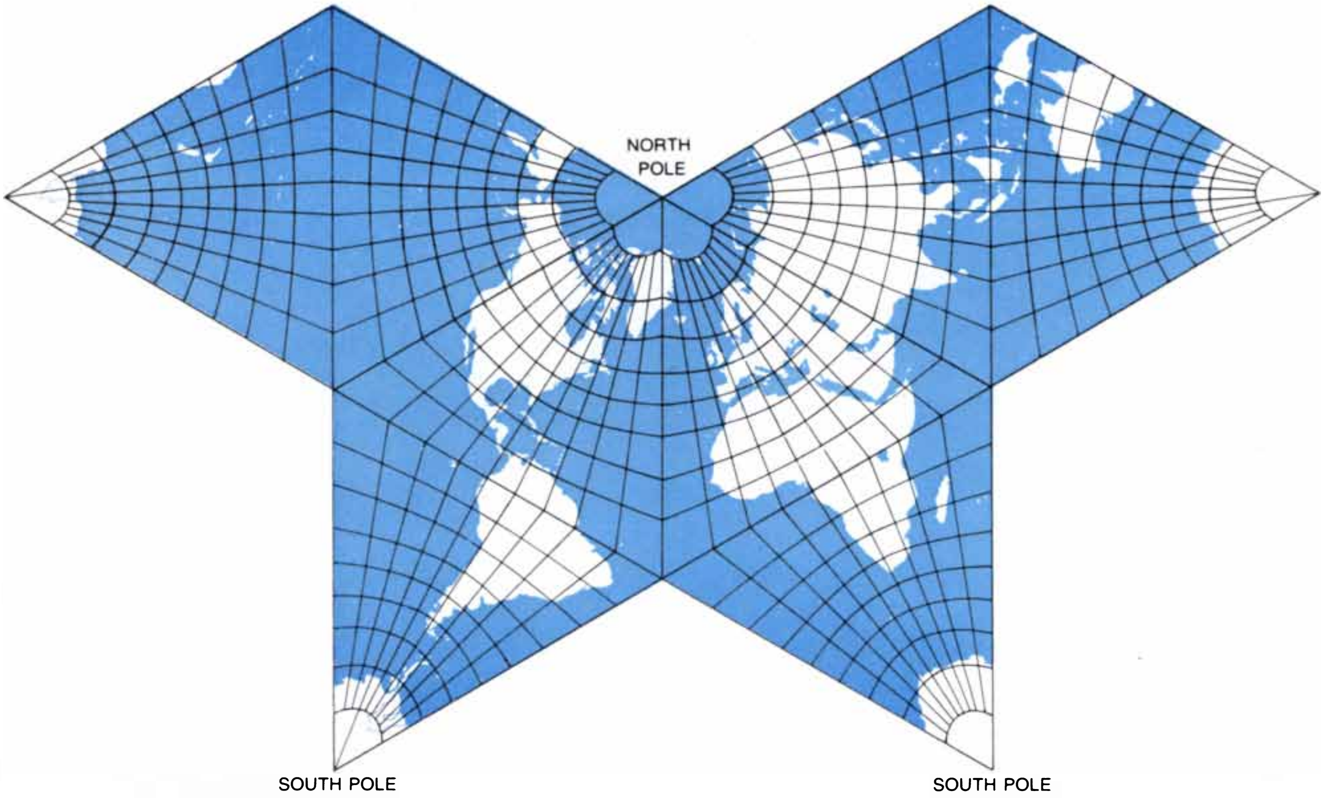
A large class of eccentric maps that we have not considered are the maps of imaginary regions. "Might-have-been" maps show what the world might have looked like if major wars had ended differently. Fantasy maps depict Oz, Hell, Eden, Poictesme, Narnia, Barsoom, Middle-Earth, Atlantis and other fanciful realms. J. B. Post has collected 98 of them in his beautiful *Atlas of Fantasy,* published in 1973 by Mirage Press, Box 7687, Baltimore, Md. And we must not forget the Bellman's ocean chart in Fit 2 of Lewis Carroll's *Hunting of the Snark:*

He had bought a large map representing
the sea,
Without the least vestige of land:
And the crew were much pleased when
they found it to be
A map they could all understand.

"What's the good of Mercator's North
Poles and Equators,
Tropics, Zones, and Meridian Lines?"
So the Bellman would cry: and the crew
would reply,
"They are merely conventional signs!

"Other maps are such shapes, with their
islands and capes!
But we've got our brave Captain
to thank"
(So the crew would protest) "that he's
bought *us* the best—
A perfect and absolute blank!"

Fisher's book on world maps closes with 36 excellent questions and answers about world geography. How well can



Butterfly map by B. J. S. Cahill

the reader do (without consulting a map) on the following selection?

1. You are on a ship five miles from an entrance to the Panama Canal and sailing due west toward it. In what body of water is your ship?

2. Flying due south from Detroit, what foreign country do you reach first?

3. Which is nearer Miami, California or Brazil?

4. Which is farther north, Venice or Halifax?

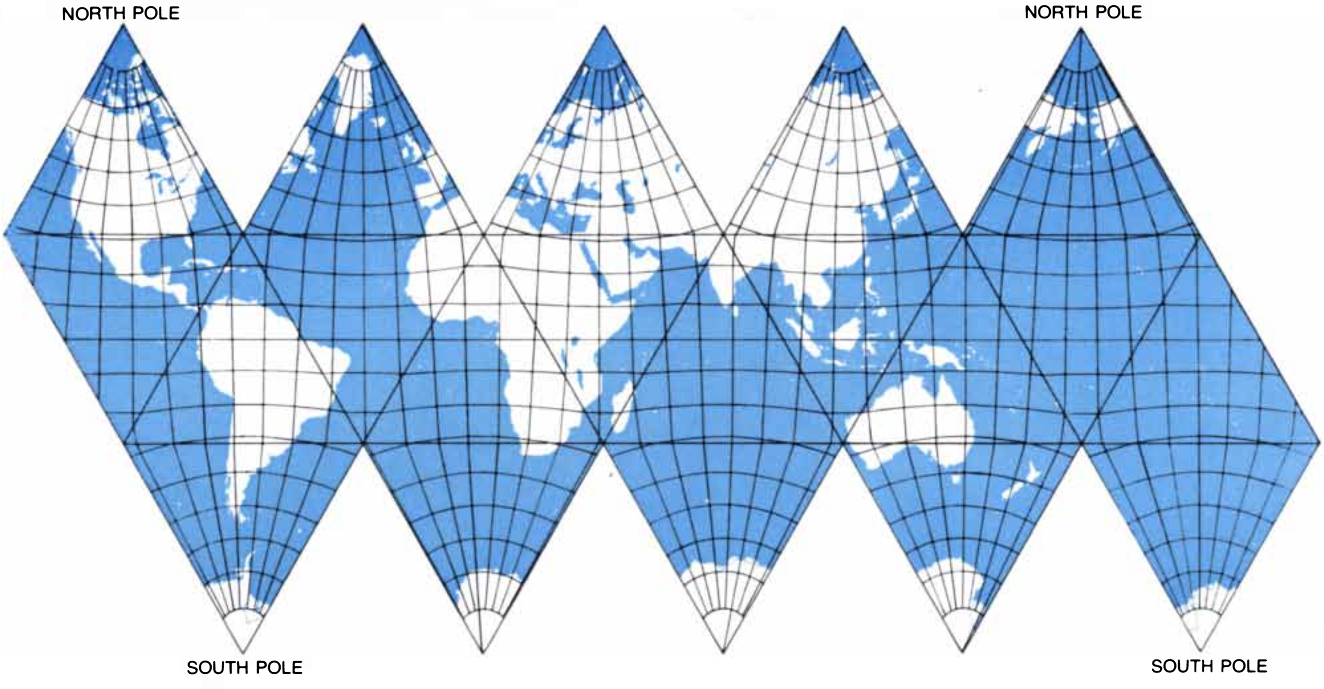
5. Which is farther south, Venice or Vladivostok?

6. Which is larger, Japan or Great Britain?

7. What four states in the U.S. touch at one point?

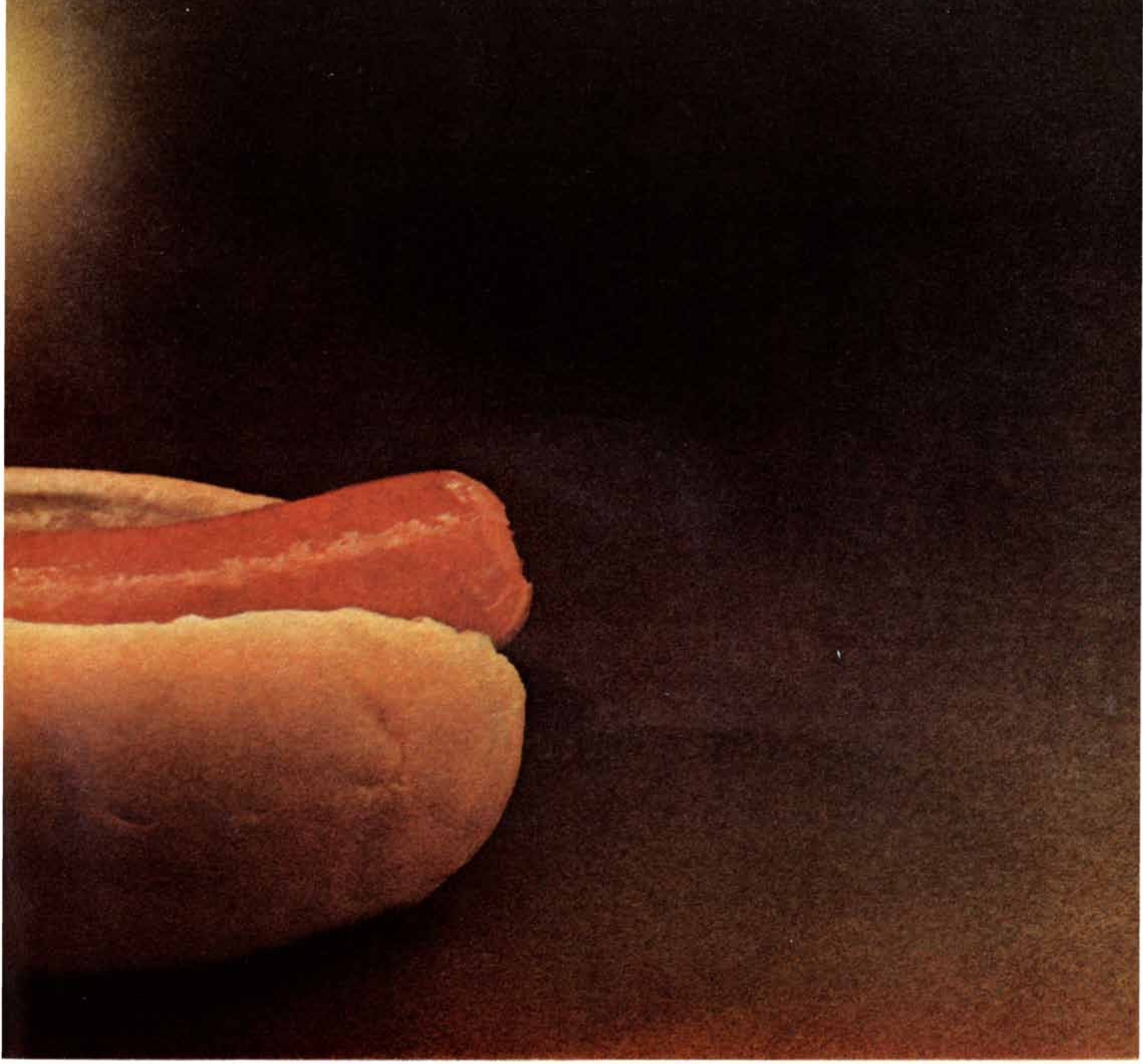
8. Does a geodesic (great circle) from Tokyo to the Panama Canal pass east or west of San Francisco?

The illustrations for this article are by Richard Edes Harrison. He also gave me invaluable assistance on the text.



Irving Fisher's Likaglobe, which folds into an icosahedron

Happy Birthday, Frank!



THE AMATEUR SCIENTIST

How to build and work with a Skinner box for the training of small animals

Conducted by C. L. Stong

B. F. Skinner, the Harvard psychologist, once conducted an experiment to see if two cats, left to themselves, would share a task essential to their joint survival. His apparatus consisted of a cage with a small pedal at one end; pushing the pedal caused a pellet of food to drop into a cup at the other end. The apparatus, which he had originally designed for an earlier experiment, has come to be known as a Skinner box.

The cats were taught separately to feed themselves by operating the pedal. Skinner next deprived the animals of food for 24 hours and then put them in the cage. One of the two cats spent its time alternately working the pedal and vainly running to the empty cup. The reason the cup was empty was that the other cat simply stood by the cup and ate the pellets as they came out. When it had eaten its fill, it curled up for a nap. The first cat then fed itself.

Scores of other behavior patterns have since been investigated by both specialists and amateurs using modified Skinner boxes. For example, a Skinner box made with lighted windows of various hues can disclose the range of color that an animal can perceive [see "The Amateur Scientist," SCIENTIFIC AMERICAN, October, 1970]. The learning that results from these procedures is called operant conditioning because the animal actively operates some aspect of its environment. Skinner boxes of the appropriate proportions can disclose variations in the sense perceptions and social behavior of animals ranging from mice and birds to dogs and pigs.

The first steps in doing such an experiment are acquiring a Skinner box and mastering the simple art of teaching the animal to use it. The box can be built by anyone reasonably skilled in the use of hand tools. Alternatively, inexpensive Skinner boxes designed for easy modification and use with animals the size of mice and rats are commercially available in the form of kits from Life Science Associates (P.O. Box 500, 1 Fenimore Road, Bayport, N.Y. 11705). The company also distributes a paperback manual titled *Experiments in Operant Conditioning*, by John M. Christiano, which presents in detail the technique of teaching animals to use the apparatus and of evaluating the results. Christiano also designed the apparatus distributed by the company.

The experimenter must provide at least one easily cleaned container for housing the animals. The containers are kept at room temperature between tests. To ensure that the experimenter is in compliance with the Expanded Animal Welfare Act of 1970 and with the policies of the National Institutes of Health on the humane treatment of animals in the laboratory, experimenters are urged to obtain a copy of the booklet *Guide for the Care and Use of Laboratory Animals*. It is available for 70 cents from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402.

The cost of the experiments varies with the size of the experimental animals, but the results of experiments with small animals are no less interesting than those of experiments with large ones. Moreover, teaching a mouse to perform an operation it would not undertake naturally can be just as satisfying as teaching a similar "trick" to a dog. Suppose you decide to work with white mice.

Live mice that are known to be in good health can be bought from companies that specialize in biological supplies, although most of these establishments do not fill small orders. The beginner should therefore ask for advice on procurement from an experienced person, perhaps a biology teacher in a local school or someone who makes a hobby of breeding mice. Not all pet shops are reliable sources of healthy animals. Mice bought from a pet shop should be checked by a local veterinarian for a few days. Frequently one can obtain a breeding pair from a local educational institution. With a pair or two from such a source you can soon raise a small laboratory colony from stock known to be healthy.

If breeding is undertaken, arrange adequate housing to keep the sexes and even the individuals apart as desired. A single pair of mice, if left to themselves, can produce 100 offspring per year. Each generation can begin to reproduce itself in from five to eight weeks. The unchecked colony could multiply within a year to more than 50,000 individuals. Humane considerations require that the size of the colony be limited to the minimum number of animals needed for the experiments. Mice, like other animals raised in captivity, lose many of their defenses as well as their food-gathering ability. Any such mice released into the natural environment are almost certainly doomed to starvation or to death by predation.

Christiano's manual on operant conditioning urges the experimenter to handle each animal briefly every day for a few days before teaching it to operate the Skinner box. Never chase the animal around its home cage with your hand, an action that reinforces the creature's natural tendency to run. Instead dangle your hand in the cage. Have patience. The animal will soon venture near, often with its tail conveniently outstretched. Pick it up by the middle of the tail.

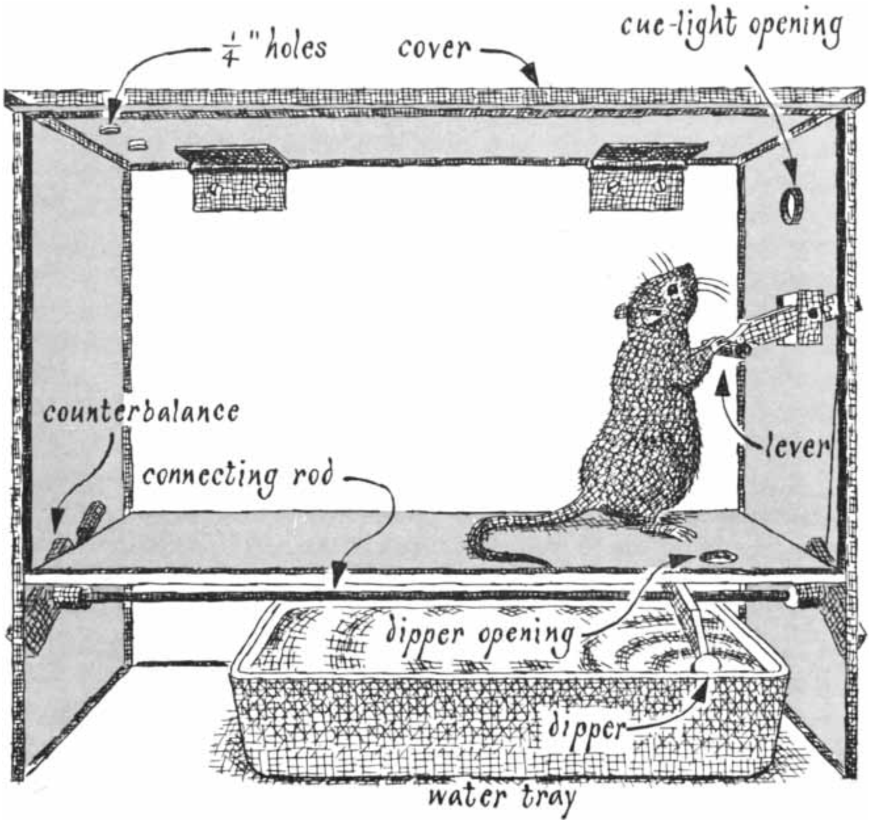
Christiano also suggests that you let the mouse walk on your bare hand. You should wear a sleeved shirt that can be easily washed. Pick up the mouse by its tail. Lower it onto your sleeve. By gentle nudging induce it to walk from your forearm onto your hand and thence onto the cage floor. Repeat the procedure several times. Finally, put your hand flat on the cage floor. Wait patiently until the mouse climbs onto your hand. Keep your fingertips buried in the bedding of the cage. This stratagem will help you to avoid getting a slight nip. You will be nipped only if the animal is frightened. If you are bitten, wash the small wound with soap and water and consult your physician. Laboratory mice are unlikely to carry a disease. They are more likely to catch something from you.

Having learned to work with mice, familiarize yourself with the Skinner box. The device consists of two essential elements: a housing that more or less confines the animal and a mechanism that presents simultaneously to the animal a reward such as food or water and a signal such as a sound or a light. The experimenter trains the animal by giving it a reward every time the animal happens to behave in the desired way or makes a move that tends to lead to the desired behavior.

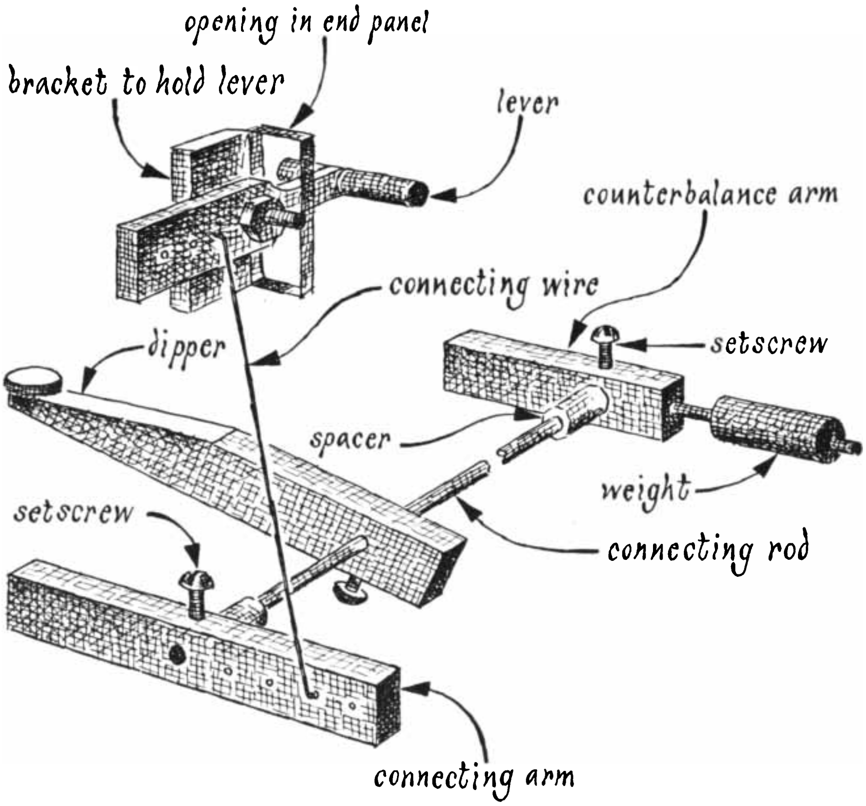
The reward, called a reinforcement by psychologists, is most effective only if it is presented simultaneously with the desired behavior or with some tendency toward that behavior. A delay of even a second between the behavior and the presentation of reinforcement can destroy much of the desired effect. In practice it may be difficult to synchronize the behavior and the reinforcement. For example, the animal may be in a remote part of the Skinner box and facing away from the source of food or water.

Operant-conditioning experimenters solve that problem by applying a principle from I. P. Pavlov's classic experiments with dogs. Pavlov demonstrated that when a neutral stimulus such as a bell is presented at the same time as food, the bell acquires some of the conditioning properties of the food. In a Skinner box that presents water as a reinforcement the water dipper makes a characteristic sound as it comes up. This sound becomes a secondary reinforcement and can be used to instantly reinforce the desired behavior. These techniques, based on the experiments of Pavlov and Skinner, enable an experimenter to condition in a matter of minutes a pattern of behavior that might take a professional animal trainer using traditional methods days or weeks to teach.

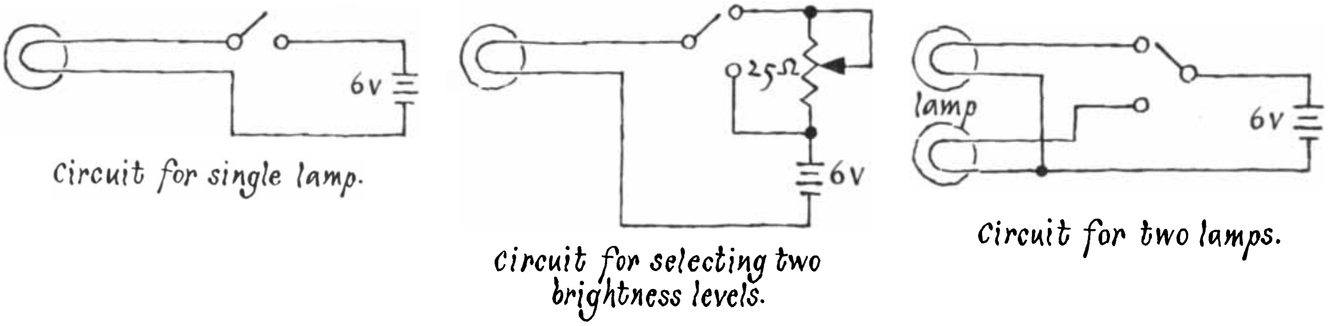
A simple but highly effective Skinner box has been designed for mouse experiments with water as the reinforcement [*see top illustration at right*]. The box has a front and back of transparent plastic; the ends, the floor and a hinged top are opaque. A mechanism for presenting the water and the signal simultaneously runs the length of the box just under the middle of the floor. It includes a slender metal shaft that is free to turn in holes



John M. Christiano's version of the Skinner box



Details of the reinforcement mechanism



Circuitry of the lamps

serving as bearings in the ends of the box.

One end of the shaft supports a counterbalancing lever arm. Clamped to the other end is a similar lever that serves to turn the shaft. This lever can be coupled to a smaller one by a piece of stiff wire. The smaller lever extends through the end of the box. The inner end of the smaller lever carries a pedal that a mouse sitting on its haunches can depress with its front paws.

The shaft also carries another lever, the outer end of which terminates in a horizontal disk about eight millimeters in diameter. When the shaft is turned to one extreme of its excursion, the disk is submerged in a small container of water under the floor. The counterweight is adjusted to the position where it automatically rotates the shaft to its opposite extreme, thus lifting the wet disk through the hole in the floor to a position that is accessible to the thirsty mouse. The upward excursion is stopped when the supporting lever strikes the bottom of the hole in the floor. The resulting sound, a distinct click, is heard at the instant the wet disk becomes available to the mouse.

A mouse that has been deprived of water for 24 hours will usually lick the film of water off the disk. More water could of course be delivered during each operation of the lever by substituting a shallow cup for the disk. Then, however, the animal might quench its thirst before the experimenter could complete a series of conditioning exercises.

Why not use food as the reinforcing agent? Food works as well as water for certain investigations and better than water for others. On the other hand, it is not needed for an experiment intended primarily to introduce the experimenter to the technique of teaching an animal how to use the Skinner box.

If food is used, the experimenter must first establish how much the animal normally eats and weighs when it is given unrestricted access to food and water. When these data have been compiled over a period of time, the food supply is reduced by 25 percent until the animal's weight has been reduced by 20 percent. The experimenter can then work out a feeding schedule that maintains a healthy but hungry animal.

Set up the initial experiment in a quiet, normally lighted room. Put the Skinner box on a table. Elevate the wet dipper and put in the box a mouse that is accustomed to being handled and that has been deprived of water for about 24 hours. (The precise period of time depends on the relative humidity and the moisture content of the animal's food.)

The mouse will soon begin to explore its surroundings. Put your hand on the control lever of the dipper assembly but do not make any further movements or otherwise distract the mouse. Soon, usually within a minute or less, the mouse will approach the disk and lick off the water.

Gently but quickly lower the disk into the water container and immediately let it rise to make a click. (Time the action to avoid hitting the nose of the animal with the dipper!) The mouse will again lick the disk. Repeat the procedure about six times.

Lower the dipper into the container and let it stay there. Wait until the animal turns its head away. The mouse will have begun to associate the click with the presentation of water. Depress and release the lever. At the sound of the click the mouse will turn its head to the dipper and again take water. Repeat the routine several times. Then wait until the mouse turns its body partly or fully away from the dipper before repeating the exercise.

The training routine requires some patience on the part of both the animal and the experimenter. Never touch the mouse, nudge it, speak to it or otherwise influence its behavior except by manipulating the apparatus. Limit the initial training period to 10 minutes. Return the animal to its home cage with a normal supply of food and water for 24 hours. Then deprive it of water for 24 hours and proceed with the second 10-minute training session.

Begin the second session by placing the mouse in the box with the dipper immersed in the water. After the animal has explored the box for a few minutes release the lever so that the dipper rises and sounds a click. The probability is high that the conditioned mouse will immediately go to the dipper and lick off the water. If it does not, let it find the dipper by exploration. Then rewet the dipper only once before the mouse turns away. This single operation will be sufficient to restore its memory.

The next step is to reinforce the desired behavior only when the mouse turns in the direction of the small lever that has the pedal on its inner end. After this behavior has been reinforced three or four times withhold the reinforcement until the mouse takes a step in the direction of the pedal. Then, having repeated the exercise two or three times, reinforce only after the mouse has taken two steps. By extending the routine you can teach the mouse to walk (even scurry) from the dipper to a position immediately under the pedal. You should be able to achieve this result by the end of the second training session.

Again allow the animal to eat and drink normally for 24 hours in its home cage. Then, following another 24 hours when the animal is deprived of water, undertake the third training session. Begin it exactly as you did the second, reminding the mouse of what it has learned as may be necessary. Thereafter when the mouse arrives at a position under the pedal, reinforce only an upward movement of its head. Soon it will hurry to the dipper and lift its head high.

When you have "shaped" this pattern of behavior, reinforce only after the animal has begun to sit upright on its haunches. By now, if you have been careful to reinforce only desired behavior, you should be able to induce the mouse to assume the upright position within a maximum period of five minutes. The remainder of the exercise is obvious. Shape the behavior of the mouse so that it puts one paw or both paws on the pedal and pushes it down.

Connect the wire link between the small lever and the large lever that you have been depressing. Thereafter the mouse will operate the mechanism until it is no longer thirsty. It might be interesting to repeat Skinner's investigation of work sharing by substituting a pair of trained and thirsty mice for his hungry cats.

The range of investigations that can be undertaken with the Skinner box is limited primarily by the ingenuity of the experimenter. Remember, however, that (as Christiano's manual points out) your timing in applying reinforcement is crucial to the success of all behavior shaping. Otherwise you may reinforce not the behavior you wanted but whatever the animal was doing at the exact moment the click sounded. Tail chewing, nose scratching, toe picking, circle turning, head bobbing and any other conceivable behavior might actually have been reinforced. Be sure to reinforce only the exact behavior that you want repeated.

It is possible to extinguish a pattern of reinforced behavior. Put the conditioned animal in the Skinner box for the usual conditioning period but remove the wire link between the pedal and the mechanism that operates the dipper. The animal will of course operate the pedal in an attempt to provide itself with water. In the absence of reinforcement the number of pedal presses per minute will decrease with time. Return the animal to its home cage for an hour or so (without water) and repeat the procedure. A smaller number of pedal presses will be observed, although usually an initial spurt of responding known as "spontaneous recovery" will occur. If this routine is followed for a few cycles, the animal's initial conditioning will be extinguished.

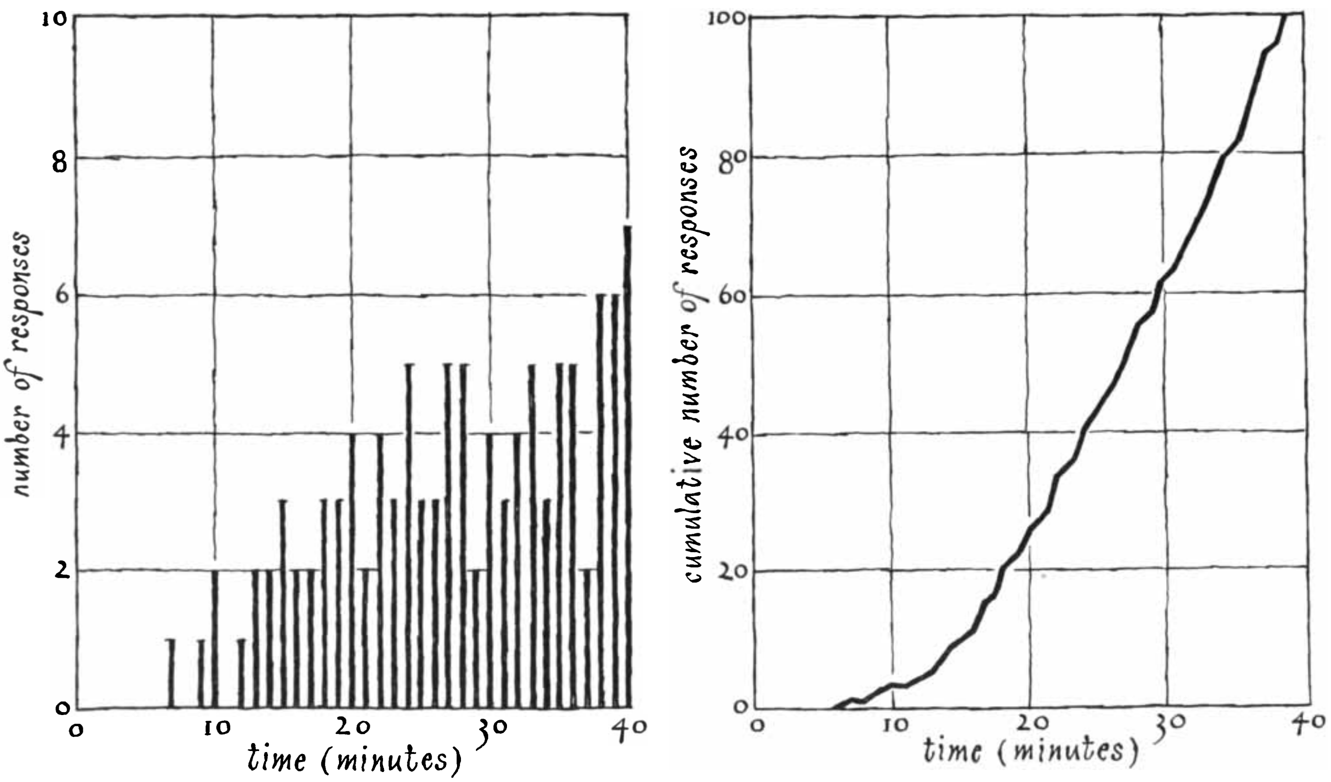
You may observe other patterns of behavior during extinction. For example, the mouse may strike the pedal sharply, bite it, resort to jumping or quick jerky motions, defecate, urinate or wash its face excessively. It is useful and interesting during any conditioning experiment to keep a detailed log of all behavior of every animal, including the number of times it pushes the pedal during each conditioning session.

The mice can be marked for identification by placing a tiny spot of distinctively colored lacquer, or coded combinations of colored spots, on the back of every animal. Coat only the tips of a few hairs. Keep and chart records of all responses.

Mice, like other animals, can learn to respond to more than one signal. For example, the Skinner box can be fitted with a miniature electric lamp placed, say, near the pedal. The experimenter can turn the lamp on and off by an external switch. An animal that has been conditioned to receive water at the sound of a click can then be conditioned to respond only when the lamp is on or, alternatively, when it is off.

The conditioning procedure is simple in concept. First the animal is placed in the Skinner box with the wet dipper and the lighted lamp. Remove the wire link and operate the dipper by hand. After the mouse has pressed the pedal perhaps six times turn off the lamp. Turn it on again just before the mouse operates the pedal. Gradually lengthen the time that the lamp is off, but do not make the mistake of letting the mouse successfully operate the mechanism and receive water while the lamp is off. Eventually you will succeed in extinguishing the established behavior pattern when the lamp is off. The animal will learn to operate the pedal only when the lamp is on, not when it is off.

Established patterns can be selectively extinguished by a number of alternative



Graphs of the response rates

Our promise
the LIFE Library of Photography will
you need to match

THE CAMERA . . .
. . . covers everything you need to know about the different kinds of cameras and lenses you can choose from — what each can and cannot do, and how to find the equipment that's best for *your* kinds of pictures . . .

PHOTOGRAPHING CHILDREN . . . tells you the right way to handle one of the most popular photographic subjects . . . from taking posed pictures that seem to tell the whole story of a child's fears, hopes, dreams . . . to candids that truly capture life's most magic and spontaneous moments . . .

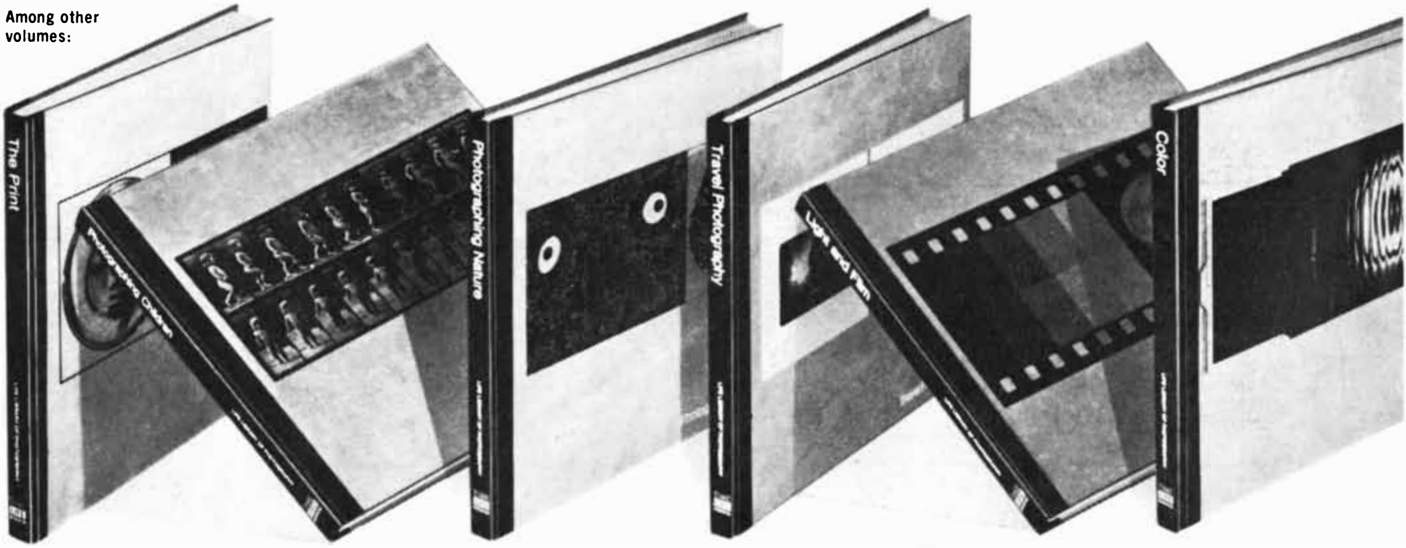
PHOTOGRAPHING NATURE . . . how to bring back nature pictures that have the feel and impact of the real thing . . . how to shoot at any time of day and in all kinds of weather . . . how to take ultra-close-ups of flowers or panoramas that go from horizon to horizon . . .

TRAVEL PHOTOGRAPHY . . .
. . . gives you the professional techniques for taking pictures that capture the real moods and feelings of the people and places you visit . . . and how to shoot it *all*, from the glare of a desert in harsh noon to the candlelit cellar of an old church . . .

LIGHT AND FILM . . .
. . . trains you to "see" in advance how the different types of film respond to different subject matter and lighting . . . teaches you how to get the right exposure and when to deliberately overexpose and underexpose . . .

COLOR . . . tells you how to try any thing and everything with color . . . how to achieve the most evocative moods or the wildest abstractions . . . how to shoot great color indoors and out, in rain or shine, early or late . . .

Among other volumes:



The LIFE Library of Photography
Invites you to accept THE CAMERA for 10 days free use

Do you enjoy photography now? Then think how much *more* it will mean to you when you have the technical skills to match your creative ability . . . when you can shoot confidently in every kind of situation! When you can shoot in mood-evoking mist or the harshness of high noon . . . by candlelight or footlight . . . for absolute realism or creative abstractions. When *you* control exactly what the final result will be!

But how do you learn what you need? The best answer yet is the LIFE Library of Photography.

The LIFE Library of Photography starts with the very basics . . . tells you what you should know about cameras, lenses, filters, lighting, the darkroom. You'll be shown just what to do to get the effects you want, with clear-step-by-step instructions and illustrations. And, as you progress, you'll find that the LIFE Library of Photography takes you just as far as you want or dare to go—with much you arn't likely to find anywhere else, including the "trade secrets" of the greatest team of photographers and photo technicians that ever lived!

Examine a Volume FREE

It's all here . . . from the basics of choosing the right lens to the far-out effects of solarization. If you enjoy photography, shouldn't you see a volume of the famous LIFE Library of Photography yourself? Do it now by sending for the volume, *The Camera,* to examine for ten days free. Details of this offer are on the coupon; read it and mail it now.

techniques. For example, you can let the animal operate the pedal successfully for a stated period of, say, 30 seconds with the lamp lighted. Then turn the lamp off and simultaneously disable the apparatus for the same interval. After a sufficient number of conditioning cycles the animal will operate the pedal less frequently when the lamp is off, and finally it will not operate the pedal at all. Whatever technique you employ, the objective is to establish a sequential link (as reflected by the behavior of the experimental animal) between the two signals, the click and the lamp.

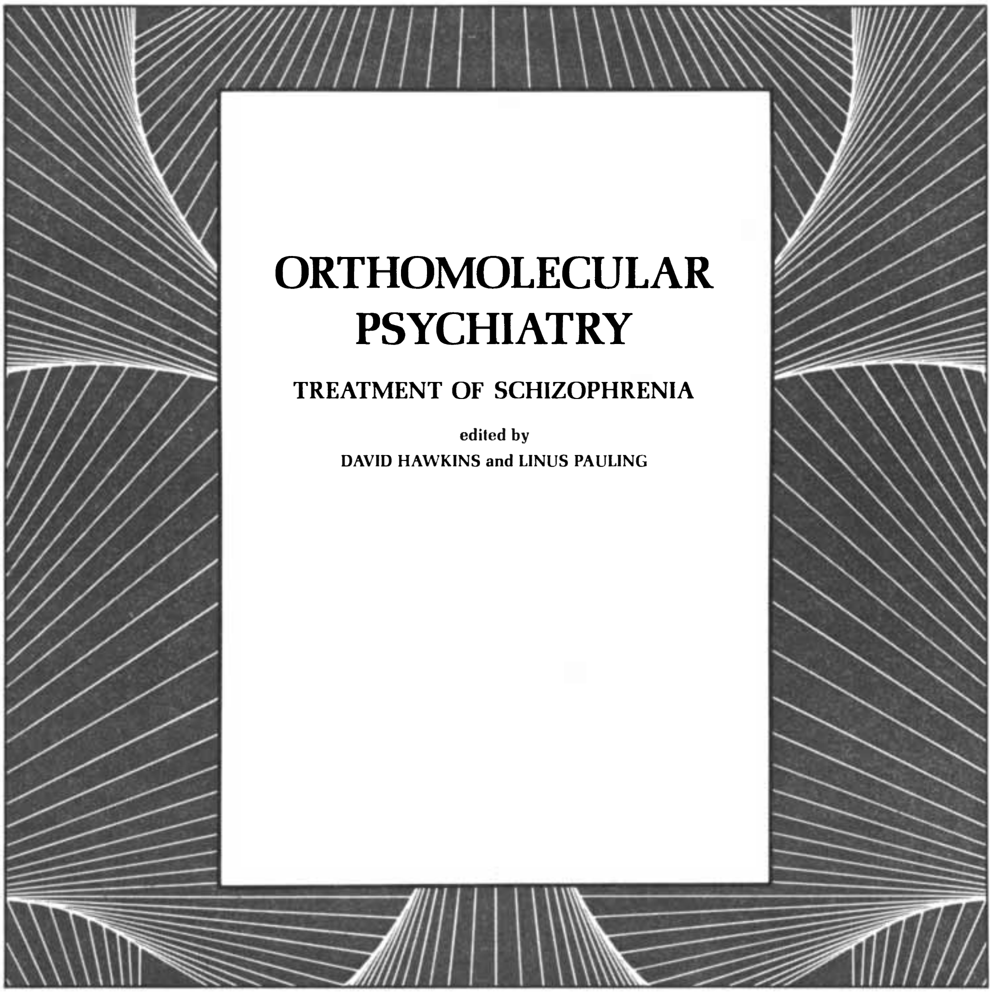
Can an animal be conditioned with a Skinner box to perform a prescribed set of operations serially? The answer is yes if the experimenter masters the art of linking (as evidenced by the animal's behavior) a set of sequential signals and responses. The series of responses can be quite complex.

Some years ago a commercial device based on that principle was displayed in the Chinatown section of New York. A sign on a wire cage invited the passerby to drop a coin in a mechanized collection box to see an "educated chicken" perform. When the coin dropped, a bell rang. A chicken in a corner of the cage immediately ran to the center and pulled a chain suspended from the top of the cage. That caused a mechanism to deliver a picture postcard of the chicken to the customer. The chicken then did a pirouette and pecked a pedal. The pedal caused a kernel of corn to drop into a cup. After snatching up the corn the chicken ended the show by retiring to its home corner.

Christiano suggests that behavior patterns of this complexity be initiated with a mouse that has been thoroughly conditioned to operate the pedal only when the light is on. Hang a looped key chain on the panel near the lamp and pedal. Put the conditioned mouse in the box with the lamp off. At first reinforce any casual or accidental touch of the chain by immediately turning on the light, which has by now become a secondary reinforcer. Maintain the lighted condition until the animal presses the pedal and obtains the primary reinforcer (the water). Turn off the lamp in exact synchrony with the presentation of the primary reinforcement. Thereafter, as in the other experiments, shape the chain-touching behavior by the method of successive approximations, always withholding the secondary reinforcement (the light) until the sequential pattern of behavior that you are attempting to shape has been observed.

"The pieces of a vast puzzle are coming together. If they can be successfully assembled, they will constitute a medical revolution . . . *orthomolecular medicine*."

— Human Behavior



From Walter C. Alvarez, M.D.:
"Many thanks for sending me the magnificent book by Drs. Hawkins and Linus Pauling. I am delighted to see it because I am a believer in the work of Drs. Hoffer and Osmond, who years ago started this idea of treatment for schizophrenia. What has bothered me for years has been the tendency of nearly all physicians to say that the teatment is no good, when they have never tried it on patients.

"Now I hope that more doctors will try it."

From *Human Behavior* [May 1973] :
"The overwhelming weight of evidence seems to point in the same direction: Orthomolecular medicine is rational, and it works. . . .[This book is] a huge collection of research articles and clinical reports concerning this biochemical approach to mental illness. . . .fascinating. . . . this is all the stuff of scientific discovery."

"I join my coeditor, Dr. Hawkins, and the other contributors to this book in expressing the hope that it will be found useful not only by scientists and physicians but also by those who suffer from schizophrenia and by their families." —Linus Pauling

ORTHOMOLECULAR PSYCHIATRY: Treatment of Schizophrenia
Edited by DAVID HAWKINS, The North Nassau Mental Health Center, and
LINUS PAULING, Stanford University
1973, 697 pages, 92 illustrations, 79 tables, $17.00

W. H. FREEMAN AND COMPANY
660 Market Street, San Francisco, California 94104
58 Kings Road, Reading, England RG1 3AA

BOOKS

The double helix, trilobites, infrared detectors and plant chemical products

by Philip Morrison

THE PATH TO THE DOUBLE HELIX, by Robert Olby. Foreword by Francis Crick. University of Washington Press ($23.50). James Watson a decade ago wrote his narrative of the final push, a tale inward-looking, provocative and brilliant. Like the commando hero he saw in himself, Honest Jim told of anxious pursuit, with Old Alpha Helix cunningly at work across the sea while paper models and smiling popsies filled Cambridge days about equally with tantalizing promises. But battle journals, however honest, are not quite history; they concentrate on the smaller scene, and might misinterpret the hint a dark lady gave when asked for directions at a crossroads. The historians describe it all later—the entire campaign, not just the fight. It is they who first sort the pieces of paper that document the weapons, the battle order and the long struggle for position.

This volume is a "full scholarly account of how the structure of DNA was discovered, set against its proper historical background." So justly writes that other hero of the double helix, who supplied the foreword. Robert Olby is concerned with the entire problem, from the nature of the support of scientific institutions and the rise of schools of research to the logical development of concept and experiment. A professional historian of science at the University of Leeds, he clearly has sound judgment and critical understanding. His sources include not only the published literature (plus theses and notebooks) in the substantive field and the considerable body of memoir and critique that has by now appeared but also a good deal of contemporary private correspondence and many interviews with principals and spear carriers alike. He retains a lively interest in strong personality and incident—and there are plenty along the way—to add color to a coherent narrative of intellectual struggle.

The story is a 20th-century one. True, there is a prehistory of nucleic acid biochemistry, but it left little legacy. We are in the world that emerged after the rediscovery of Mendel, the development of polypeptide synthesis by Emil Fischer and the birth of X-ray crystal analysis before World War I. It would be invidious to try to summarize all the rich texture of this history, but one may outline the chronicle by describing the five stages of the campaign.

The first question (here one sees how the hot issues of one decade become the unexamined "self-evident" axioms of the next) is that of long-chain polymers. James Clerk Maxwell thought the average molecule in living microscopic beings contained about 50 atoms. The term "macromolecule" was introduced about 1925 by Hermann Staudinger. He was ridiculed by his fellow chemists for his ideas, which revived with quantitative support the classical hunches of Kekulé. The work carried him into the "unpleasant and poorly defined" rubber and polystyrene. It was not easy for X-ray crystallographers to rid themselves of the idea that the molecule could not be larger than the unit cell they saw. Nevertheless, by 1939 the X-ray analysts in Berlin and the remarkable laboratory at Leeds under W. T. Astbury, an enthusiast and an original ("I am alpha and omega, the beginning and the end of the whole thing," he said), had demonstrated long fibrous biomolecules and had laid the foundations of a sophisticated method of X-ray analysis via models—although their first results were seductively misleading.

Next one had to part tape from cassette. By 1931 the chemists had found nucleic acids to be rather simple, with only four nucleotides, and had turned their attention to the more complex protein. Here is the story of those infants with air-darkened urine, which led by way of Garrod and Beadle to the genetics of molds and to the clear recognition that genes somehow code for enzyme proteins. Even a direct look at the protein rods of tobacco mosaic virus did not, "after all, reveal the chemical identity of self-reproducing units." The denatured virus left a purified RNA residue, but in 1942 they did not check carefully to see if that portion could be infectious by itself. One reason was that the preparation method had damaged the RNA tapes; another was that the investigators were in a "Kuhnian box": no one was prepared for the revolutionary idea of genetic information without protein.

That way was of course found by the medical immunochemists at the Rockefeller. They devoted themselves to the curious transformations of the pneumococcus as a key to immunization against bacterial lobar pneumonia. The bombshell of induced transformation, with its disturbing whiff of Lamarck, was detonated in 1928 by a "quiet and retiring medical officer... a civil servant and proud of it," a man who could not be persuaded to leave his bench to attend a meeting or to read a paper. (Once his friends had to bundle him into a taxi to get him to a relevant talk at an international congress in London.) That bacteriologist, Frederick Griffith, was the other side—and fully as English—of Astbury's fine presentation of a puffy John Bull. The answer was found and secured by 1943. Macfarlane Burnet visited Oswald Avery in New York and wrote home of "an extremely exciting discovery which, put rather crudely, is nothing less than the isolation of a pure gene in the form of desoxyribonucleic acid." The final step in preparation for Watson-Crick was the enigma of the base ratios: the one-to-one match of the amounts of adenine and thymine and of guanine and cytosine in DNA over a wide variation in their separate fractions. This had been made clear by the early 1950's by a number of people.

Here the theme changes. The issue is no longer model paradigms, molecule size or protein versus nucleic acid but the rise of complementary schools. One was the informational school of Luria and Delbrück and their friends: physical biologists who worked out the genetics of bacterial viruses, suspicious of biochemists as people who were analyzing the brass in the gears rather than how the

What's new in concrete math... the lives of bees...symmetry... archaeology...data analysis... quantum physics...astronomy...

For your own reading pleasure. Or as gifts.

(Publishers' Prices shown)

42325. THE DISCOVERY OF OUR GALAXY. *Charles A. Whitney.* "...a clear and exciting account of some of the history behind modern extragalactic astronomy...Charles Whitney's account is easily understandable, rapid moving, remarkably free of jargon, and requires no mathematical background."—*American Scientist.* **$10.00**

49340. FUNDAMENTAL ASTRONOMY: Solar System and Beyond. *Franklyn W. Cole.* Traces our intellectual assault on the mysteries of the universe from man's earliest sensory-oriented glimmerings to the very frontiers of modern astronomical thought. **$16.50**

81820. SYMMETRY IN SCIENCE AND ART. *A. V. Shubnikov and V.A. Koptsik.* This new edition of a classic work on symmetry explores every facet of the subject. *Counts as 2 of your 3 books.* **$35.00**

50220. THE GENTLE TASADAY. *John Nance.* A veteran journalist's first-hand account of how these primitive food-gathering people were found, what scientists have discovered about their culture, and what the future holds for them. **$15.00**

52190. HANDBOOK OF APPLIED MATHEMATICS. *Edited by Carl Pearson.* Over 20 contributing author-specialists provide over 1,000 pages of vital and practical techniques for use of mathematics as a tool. *Counts as 2 of your 3 books.* **$37.50**

87855. THE WORLD OF MEASUREMENTS. *H. Arthur Klein.* Wide-ranging overview of meteorology gives a fascinating history of the development of weights and measures. **$14.95**

39745. COMPANION TO CONCRETE MATHEMATICS. *Z.A. Melzak.* "A delightful collection of problems organized into over sixty sections, together with historical notes, solutions, observations, and remarks. Mathematicians at all levels will enjoy sampling this work."—*American Mathematical Monthly.* **$16.50**

68750. PHYSICAL THOUGHT FROM THE PRESOCRATICS TO THE QUANTUM PHYSICISTS. *Samuel Sambursky.* 500-plus page anthology covers the most important events in the development of the physical sciences. Includes excerpts of great scientists from Aristotle to Pauli. **$20.00**

34310. ARCHAEOLOGY BENEATH THE SEA. *George F. Bass.* Exciting account of how the underwater exploration of a Syrian trading ship which sank off the coast of Turkey in 1200 B.C. paved the way for the science of underwater archaeology. **$12.50**

How the Club Operates • The Book Club **News**, describing the coming Main Selection and Alternate Selections, will be sent to you 15 times a year at 3-4 week intervals. • If you wish to purchase the Main Selection, do nothing and it will be shipped to you automatically. • If you prefer one of the Alternates, or no books at all simply indicate your decision on the reply form always enclosed with the **News** and mail it so that we receive it by the date specified. • The **News** is mailed in time to allow you at least 10 days to decide if you want the coming Main Selection. If, because of late mail delivery of the **News**, you should ever receive a Main Selection without having had the 10 day consideration period, that Selection may be returned at Club expense. • Continued membership past the trial period brings you the added benefits of our Bonus Book Plan—an important way to save even more, at least 70% off publishers' prices.

MEMBERSHIP APPLICATION

Take any 3 books for only $3.95

if you will join now for a trial period and take only 3 more books—at member discount prices—over the next two years.

The Library of Science

Riverside, New Jersey 08075

Please accept my application for membership and send me the three volumes indicated, billing me only $3.95 for all three. I agree to purchase at least three additional Selections or Alternates during the first two years I am a member, under the club plan described in this ad. Savings range up to 30% and occasionally even more. My membership is cancelable any time after I buy these three books. A shipping and handling charge is added to all shipments. **Send no money. Members are billed when books arrive.**

3 books for $3.95

Indicate by number the 3 books you want.

[ ] [ ] [ ]

A few expensive books (noted in book descriptions) count as 2 choices.

Name ______________________

Address ______________________

City ____________ State ______ Zip ________

(Offer good in Continental U.S. and Canada only. Prices slightly higher in Canada.) Books purchased for professional purposes may be a tax-deductible expense. 2-2AD

37195. BY THE EVIDENCE. *L.S.B. Leakey.* Fascinating reading of Leakey's early years in Africa and his development into one of the most respected scientists of the twentieth century. **$9.95**

68735. PHYSICS OF THE EARTH AND PLANETS. *Alan H. Cook.* What forces shaped the earth? What is the mechanism that prompts continental drift? These questions—and dozens more—are answered by one of the world's leading authorities on physical meteorology. *Counts as 2 of your 3 books.* **$24.75**

48630. THE FIRST GREAT CIVILIZATIONS. *Jacquetta Hawkes.* Extraordinary history of the great civilizations of Mesopotamia, the Indus Valley, and Egypt. Big and lavishly illustrated. **$12.50**

62630. THE MILKY WAY. Fourth Edition. *Bart Bok and Priscilla Bok.* A thoroughly updated and expanded edition of the most popular book on our galaxy. Outsized volume, filled with over 125 pictures, charts and diagrams, includes the latest in scientific research. **$15.00**

44861. THE ENCYCLOPEDIA OF CHEMISTRY. Third Edition. *Edited by Clifford A. Hampel and Gessner G. Hawley.* Completely revised and greatly expanded new edition of a classic encyclopedia. 600 leading chemists thoroughly explore over 800 theoretical and applied aspects of chemistry and chemical technology. Over 1200 pages. *Counts as 2 of your 3 books.* **$39.50**

52650. THE HERITAGE OF COPERNICUS. *Edited by Jerzy Neyman.* Sponsored by the National Academy of Sciences, this volume commemorates the 500th birthday of Copernicus by offering 25 highly readable essays on the quasi-Copernican revolutions that have occurred in the past five centuries. *Counts as 2 of your 3 books.* **$30.00**

44941. THE ENCYCLOPEDIA OF PHYSICS. Second Edition. *Edited by Robert M. Besancon.* In this updated and expanded edition of a modern classic, over 300 distinguished authorities in 344 articles offer incisive, authoritative information on every major topic in physics today. Outsized volume contains more than 1000 pages. *Counts as 2 of your 3 books.* **$37.50**

68690. THE PHILOSOPHY OF QUANTUM MECHANICS. *Max Jammer.* A distinguished scholar and physicist offers a truly impressive study of the interpretations of quantum mechanics. Explores virtually all of the major interpretations—from the early semi-classical to the modern stochastic and statistical interpretations. **$19.95**

64875. NUBIAN RESCUE. *Rex Keating.* Dramatic story of the race to save hundreds of Egyptian monuments from rising waters of the Assuan Dam. Fully illustrated. **$12.50**

73190. QUEUEING SYSTEMS. Volume I: Theory. *Leonard Kleinrock.* Kleinrock's long-awaited work. "...a beautifully written innovative book. If you study queues, own this book."—Ivan T. Frisch, Editor-in-Chief of *Networks.* **$19.95**

78820. THE SOCIAL BEHAVIOR OF THE BEES. *Charles D. Michener.* The first definitive work on group phenomena among these most amazing insects. A publishing classic. *Counts as 2 of your 3 books.* **$25.00**

73330. A RANDOM WALK IN SCIENCE. *Compiled by Robert L. Weber and Eric Mendoza.* A collection of jokes, short essays and poems on humor and humanity in physics and related sciences. "...it is the best anthology of its kind I have seen."—Martin Gardner. **$12.50**

44000. EINSTEIN AND THE GENERATIONS OF SCIENCE. *Lewis S. Feuer.* What really accounts for the shattering insights of Einstein, Bohr, de Broglie, Heisenberg, and other great scientists? A bold and controversial explanation. **$12.95**

67190. THE PATH TO THE DOUBLE HELIX. *Robert Olby.* A captivating history of twentieth century experimental biochemistry and a highly readable account of the physical and chemical nature of the gene. *Counts as 2 of your 3 books.* **$23.50**

watch worked. The second school, no less influenced by the successes and the methods of physics, was that of the British structuralists, dealing in complex X-ray crystallography. By training and tradition the two groups shared an appreciation of the subtlety of the chemical bond and even seemed to hanker for physically new specific forces between biological molecules. There were, it is true, two schools of X-ray structural studies. The British, under the Braggs, with Bernal and Perutz and Wilkins and Huggins, was the best in the world; its members were analysts in all detail, tireless and ingenious. But in Pasadena there was Linus Pauling, whose conviction that the bond lengths and angles of the quantum theorists were reliable guides strengthened the bold strategy—model-building synthesis—with which he outguessed the structures cautiously puzzled out of the spots of the diffraction diagrams. Pauling and his co-workers found and proved the protein helix in 1950 and 1951. By the end of 1951 Perutz was convinced that the Californians were right. The protein helix fit the data neatly and was deeply plausible on general group-theoretical grounds. The helix became fashionable. "You would be eccentric . . . if you didn't think DNA was helical," reports Crick.

Watson met Crick in 1951. Crick was then a confident extrovert, although a theorist without a Ph.D. at 35; Watson, a diffident "loner," was a prodigy, a former quiz kid with a bacteriophage Ph.D. at 22. The structuralists at their most thoughtful had met the informationalists at their most enterprising—and Peter Pauling lived upstairs in Cambridge, a hot line to Pasadena. (Luria and Linus Pauling could not themselves visit England then, because the Age of McCarthy kept those two men without passports. It may be that it was the U.S. Passport Office that fixed the outcome of the hunt for the double helix.)

The last 100 pages cover the final engagement culminating in May, 1953, and extend the tale briefly to 1973. The meticulous study seems to do justice to all the participants in a tangled skein of ideas, personalities and interchanges by design and by chance, but the historian and not the reviewer must be read for that detailed story. The discovery would have been made, "we may be fairly confident," in London or Pasadena a year or two later if Dr. Watson had never come to Mr. Crick.

One important theme is slighted: the computer metaphor supplied by the mathematicians, from Turing's machine to the clever von Neumann theorem on self-reproduction as no paradox. To a physicist the molecular realization of DNA still seems less important than the generalities of its functional structure. The few nucleotide symbols with low interaction, the great length, the helical coiling, complementary replication with the alternation of double helix and single helix—all these seem more general than any particular bases and bonds. It was Crick himself who saw in 1955 that what a single DNA base coil provided was not a set of holes for a long protein template but a spatial pattern of hydrogen bonds. He *predicted* the little adapter molecules that transcribe the code, a byte at a time. It was a glowing fusion of structure and function, and it opened the "tumultuous phase" of molecular biology that is now a richly elaborated scheme, neither mysteriously quantum-mechanical nor naïvely reductionist—nor yet all-powerful in the study of life.

TRILOBITES: A PHOTOGRAPHIC ATLAS, by Riccardo Levi-Setti. University of Chicago Press ($27.50). "The earliest description of a trilobite was given by Karl von Linné" himself. The fossil was 11 inches long, a giant; no other example of the same species was found in the two centuries since Linnaeus had found his in the splendid mineral cabinet of wealthy Count Tessin. In 1974 the author of this slender square book of some 150 large and striking black-and-white photographs may have at long last come across the same species, in the Middle Cambrian shale beds at Manuels in Newfoundland: "giant trilobites, up to one foot long and colored in bright yellow, orange and red (the colors of iron oxides coating the carapace)." His find was shared by his nine-year-old son Emile.

No paper atlas can quite come up to the romance of such a find in the old muds, now shale, that continental drift had long ago parted from the Swedish strata of the same epoch, but the crisp, detailed macrophotographs displayed here impart a measure of the same excitement to anyone who turns the pages, even if he is not already a fancier of trilobites. Most of the pictures are magnified a few times and display the markings, lobes, hinges, eyes and doubled appendages of these ancient marine arthropods. It is remarkable how much we know of these sculptured creatures that dominated the seas for half of all the time that multicellular life is known to have existed, from about 600 million to 230 million years ago, to the end of the Permian. They were filter feeders mostly. Their tribe first appears fullblown, pretty surely because they had suddenly found a way to mineralize the chitin layers of their carapaces with calcite or calcium phosphate and thus to leave hard, durable casts. Here are entire nurseries of trilobites, all sizes together, here the cast-off molted portions we mainly find, even one soft-shelled animal preserved after molting. A wonderful black layer at Rome, N.Y., classically studied at the end of the past century, showed the soft parts of 700 animals (usually we see only the hardened back portions) preserved by mineralization with tiny particles of pyrite. Soft-X-ray pictures of those specimens made in stereoscopic pairs in 1973 (by J. Cisne, working with the author) have revealed much of the internal structure of the species, its digestive parts and even its complex functional musculature. Some even show the cloud of food particles squeezed out by burial compression. Trilobites typically crawled, often flipped their old shell upside down as they wriggled out of it and very frequently rolled up to guard their soft underparts. Some orders had complex and effective compound eyes. Two eye forms are found that are divided into two layers split along an aspherical curve. Both boundaries were described mathematically in the 17th century, one by Huygens and one by Descartes—with no reference whatever to trilobites but as solutions to correct for spherical aberration. Ray tracing demonstrates the accuracy of the evolved optical design and tells us that the top layer was of properly oriented crystalline calcite and the lower one of chitin. Apparently, then, fully aplanatic, aspherical lenses are nothing new but are hundreds of millions of years old under the sunny waters.

Professor Levi-Setti is an Italian-trained nuclear physicist at the University of Chicago and an amateur at trilobites, although a devoted and learned one. It is striking that in his day Franco Rasetti, a partner in the famous Fermi group at Rome, came to work in nuclear physics at a rather provincial university on this side of the Atlantic just before the war. He found the equipment and conditions inadequate for his work with neutrons and radioactivity. And so he turned to the study (already dear to him) of trilobites, which he could split easily out of the road cuts around his new campus, and became a redoubtable trilobite authority. (The two Italian trilobite-loving nuclear physicists met several years ago in Rome.)

There is much to learn in the concise, semitechnical but personal text. Yet the work is not mainly for reading; it is above all to delight the eye. Few books of science in a year mark out so strong and beautiful a visual pattern.

INFRARED DETECTORS, edited by Richard D. Hudson, Jr., and Jacqueline Wordsworth Hudson. Dowden, Hutchinson & Ross, Inc. Distributed by Halsted Press ($26). Even in the darkest motion-picture theater there is an ambient flood of photons not much lower than that of a sunlit landscape. We cannot see them, however: they are infrared photons emitted by the warm bodies of the audience. Their wavelengths cluster around 10 micrometers, about 20 times as long as the most abundant photon of sunlight, and their energy is less by the same factor. The rods of the peripheral retina, most sensitive of the eye's light receptors, are no better able to detect such light than is the undifferentiated surface of the human skin. The threshold of vision drops by a factor of a million million as the incoming photon wavelength increases from that of green light by a factor of only two.

Plenty of people, notably the military, like to see by infrared. The broad band that stretches from the conventional one-millimeter wavelength, where techniques border on microwave ones, to one micrometer, where dye-enhanced photography works rather well, is filled with natural and artificial signals. But the infrared band is uncommonly noisy: the innermost black-lined tube of a telescope is a bright glow at 10 micrometers and longer. The key to meaningful detection is not sensitivity but a subtle understanding of thermal noise.

By the turn of the century infrared radiation in the atmosphere and in the laboratory was already a field of great interest. The best detector then available was, like the human skin, a thermal detector of infrared, converting a broad span of photon wavelengths indifferently to heat energy and noting the small temperature change. Samuel Langley bragged quite honestly that his bolometer of thin, blackened fibers of platinum could reveal by small changes in its electrical resistance the radiation from a cow a quarter of a mile away. We still use the resistive bolometer—now of doped thin-film germanium or of silicon cooled below liquid-helium temperature—and it is able to pick up not a resting cow but a warhead at silo temperature flying in the cool of orbit 1,000 kilometers away. Such elegant analogues to the temperature-sensitive skin are not the only sensitive detectors for such weak photons. Mainly since World War II we have found a wide variety of eye analogues as well, various detectors capable of response photon by photon throughout most of the infrared band.

The ultimate limits are no longer in doubt, and they are not far away. They are determined not by ingenuity but by that inescapable noise. The galvanometer spots trembled even in the old days; by 1925 it was plain that using more sensitive instruments would merely make the thermal fluctuations more noticeable. You had to understand the sources of noise. By the 1950's the total situation was clear: the sources of noise lay in the randomness of photons, both incoming and outgoing (every detector also radiates), and the fluctuations of heat flow from the surroundings to the detector and of the electric-circuit energies required for measurement. Any noise arising from the practical detector function itself is excess, and it can be reduced by understanding and hard work—often by cooling the detector.

Nowadays most well-developed commercially available detectors approach ultimate limits within a factor of 10, the best of them within a factor of two. Bandwidth, mechanical stability, available area and speed of response are the present bases for choice. This excellent anthology of 43 facsimile reprints, picked papers dating from 1946 to late 1973, sums up the field. There are several excellent review papers and many clear and important single contributions, ranging from work on the eye to engineering accounts of cooling techniques and of how to get as close as one can to making a truly black body (use a very sharp cone). Photon detectors (including some laser work) take the biggest slice of the pages, but thermal detectors are not slighted.

The photoelectric effect demands volts of energy to set an electron free from a surface—or from ordinary molecular bonds. Thus the eye, the television camera and the photographic emulsion work well in the visible, where a photon has a volt or two to spend. For longer waves the quantum jump must be a hop that is realized within the intricate lattices of some semiconductors. Carefully grown thin single-crystal sandwiches of such substances as lead-tin-tellurium respond well to photons in the important 10-micrometer range: it costs little energy to generate an electron-hole pair at the interface between a lattice with positive current carriers and one with negative carriers. For even longer waves there is the purest indium antimonide. In it some electrons, shielded by the electrical pseudovacuum of the perfectly responding lattice, act as though they were bound with a percent or so of normal atomic energies. Photons with wavelengths of up to a millimeter can be absorbed, not to release such electrons (which are already mainly free) but to give them a little more kinetic energy, detectable as an increase in "hot" electron mobility and hence as a change in alternating current-voltage relations.

Ingenuity and plenty of painstaking development are manifest in these papers, all of which are American or British in origin (although the art is reasonably cosmopolitan, as the augmented bibliography makes clear). Diamond windows, cryogenic refrigerators (which could fit nicely in a lunchbox and are able to cool a small detector to liquid-nitrogen temperatures on the same power as a light bulb) and elegant mirror optics that cool the detector they house by a kind of radiative jujitsu, making sure that it can see only the coldest and darkest reaches of outer space, are examples of the detail and depth here collected. All is from the open literature, and the closed doors are not hard to extrapolate beyond. The editors, who offer a number of useful summary sections, are well-informed experts in technical publication from the wilder suburbs of Los Angeles, where these mysteries are known to flourish.

PHYTOCHEMISTRY: VOLUME I, THE PROCESS AND PRODUCTS OF PHOTOSYNTHESIS. VOLUME II, ORGANIC METABOLITES. VOLUME III, INORGANIC ELEMENTS AND SPECIAL GROUPS OF CHEMICALS. Edited by Lawrence P. Miller. Van Nostrand Reinhold Company (Volume I, Volume II and Volume III $24.50 each). These leaf green, not very bulky volumes present a modestly encyclopedic but not frighteningly technical summary of the chemistry of plants, with emphasis on the green, higher plants, although the others are not absent. Most of the three dozen chapters treat some broad class of compounds. A few are more general, looking at such topics as molecular taxonomy or prescription drugs. The authors are all from North America or Western Europe; the pages offer text spotted with tables and spidery chains of structural formulas, with a few pages of electron micrographs. A general reader, such as this reviewer, will not read every page or even every chapter, but the work is an inviting reference source, surprisingly lively. Most of the authors have the happy ability to say what they know and do not know without obfuscation.

There are three classes of plant chemical products, which define the three

"This is the most clever geometry book I have ever seen!"

—Thomas E. Lepperd, Evanston Township High School, Evanston, Illinois

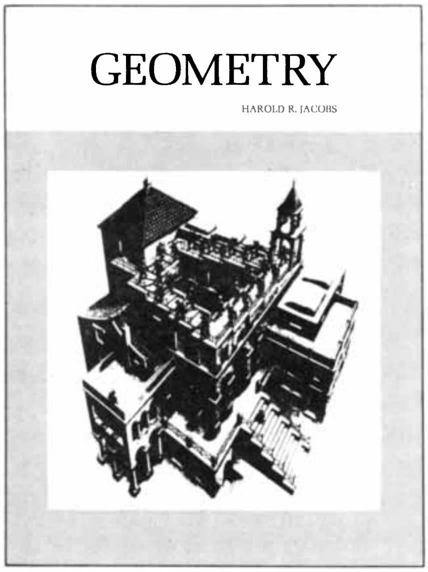
"Another typical Jacobs masterpiece!"

—Al Bigger, Auburn High School, Auburn, Alabama

"This book may single-handedly save geometry in the high schools."

—Patrick J. Boyle, Santa Rosa Junior College, California

Strong words about a strong book



GEOMETRY

1974, 701 pages, 1100 illus.,
ISBN 0-7167-0456-0, $9.00
Teacher's Guide $5.00
Transparency Masters $40.00
Test Masters $2.50

Also by Harold Jacobs
MATHEMATICS
A Human Endeavor
A Textbook for Those Who Think They Don't Like the Subject
Adopted by more than 2000 high schools
Foreword by Martin Gardner
1970, 529 pages, 634 illus.,
ISBN 0-7167-0439-0, $8.50
Teacher's Guide $5.00

"Einstein could have used Mr. Jacobs' book and still not have become so bogged down in the run-of-the-mill education that he could have missed the theory of relativity. Remember, Einstein said 'common sense is the collection of prejudices acquired by age 18'—no prejudices would be acquired by the geometry student and quite a few would be eradicated through this text."

—James J. Dolan, Chaminade Preparatory School, Canoga Park, California

"This is a book that belongs in every geometry class in the country. It has something to offer students at every level, combining wit and imagination with some very solid mathematical principles."

—Diana Brown, Wheeler High School, Marietta, Georgia

"This is a fascinating book with much more than the usual theorems. The cartoons and photographs make this a geometry book that is enjoyable reading. Most important, the problems are very good—they make one *want* to think and understand. . . . Congratulations, Freeman and Jacobs have done it again."

—Mark Saks, Community College of Philadelphia, Pennsylvania

The underlying theme of *Geometry* is a simple one: the principles of geometry may be more easily learned when they are shown to have immediate, practical applications in the student's world. *Geometry* is an exciting blend of humor, clear explanations, and an uncompromising attention to basic concepts—a blend that really does inspire students. We think that once you've examined this book, you'll have some strong words of praise too.

W. H. FREEMAN AND COMPANY
660 Market Street, San Francisco, California 94104
58 Kings Road, Reading, England RG1 3AA

volumes of the work. The first class is made up of bulk products that flow from green plants as converters of solar energy on the grand scale: food, fiber, oils and timber (to say nothing of the even more essential product oxygen). The second class includes such widely occurring families of compounds as alkaloids, steroids and terpenes. The third class holds the smaller-yield fine products: waxes, vitamins, plant hormones, sulfur compounds and the rich enzymatic repertory of the mitochondria.

The carbohydrates and their polymers store solar energy and also support the broad photosurfaces of the plant. Cellulose is the main structural material of the higher plants, a linear chain of simple glucose links that constitutes about half of a perennial plant and a third of an annual. Chitin, linked with protein, serves the same structural function for the fungi and the lower plants. Chitin is another linear chain of sugar rings, but with a modified amino group replacing one simpler decoration on each glucose ring. Similar chitin-protein complexes surprisingly form the exoskeleton of the lobster and the ladybug.

We attach significance to the integer 5 because we count on our fingers. The diverse plant compounds called terpenes give five and 10 a still deeper meaning. Extraordinarily diverse, all terpenes contain the units called isoprene: five carbon atoms linked in a particular way. Camphor and menthol and most such fragrant products of plants are typical members of the family; the blue haze above an evergreen forest in summer is pungent with droplets of terpenes. Such was the classical view. Nowadays the biochemists see multiples of five carbons almost everywhere. The carotenoids—the pigments and the vitamin precursors of colorful leaves and of carrot root—are 40-carbon terpenes. Powerful plant hormones such as the gibberellins and the sterols can be seen as terpene derivatives, from the C_{20} or the C_{40} clan. Even chlorophyll can be traced to a C_{20} precursor, phytol. Rubber, gutta-percha, chicle and jelutong (a similar latex that is a key ingredient of bubble gum) are all lengthy, inert polymers of isoprene-like units. There is power in five.

The phytochemists are far from the end of their discipline. The how is hard enough, and the why escapes us in many places. It seems clear that rubber latex does not serve to seal plant leaks against insect punctures, is not a reserve food supply or a product of a respiratory reaction chain. "There is no evidence that polyisoprenes have any biological function whatsoever.... [They are] inert end-products of metabolism." The same letdown is much repeated; nicotine and caffeine and morphine, however active for human beings, may be no more than "metabolic sludge" for their solar-powered synthesizers. Oxygen must long ago have deserved the same description.

Les Objets Fractals: Forme, Hasard et Dimension, by Benoît Mandelbrot. Flammarion, Editeur, Paris (44 francs). There is a short list of works that display for the mind's delight the profound role of geometric form in the fabric of the world. D'Arcy Wentworth Thompson's *On Growth and Form,* with its account of living creatures large and small, bridges and backbones, honeycombs, cracked mud and rhinoceros horns, heads the list. The geometry therein is Greek in topic and simplicity. Hermann Weyl's marvelous lectures on symmetry in nature follow closely, extending the mathematics to obey the demands of repetitive rigid displacement. The papers of Cyril Stanley Smith—summarized by him in this magazine some 15 years ago—first made clear how foams and tissues and polycrystalline metals evince the struggle between the Gibbsian play of local energies and the overall rule of topology, Eulerian in epoch. This French paperback is a fresh addition to that notable shelf, bringing to the knowing viewer of natural artificial form the unexpectedly diverse application of the theory of real variables. It is proper 20th-century mathematics nucleated around names such as Minkowski, Besicovitch and Kolmogorov.

The book and in some sense the ideas begin with Jean Perrin, whose study of quantitative Brownian motion visibly displayed the reality of the kinetic theory before World War I. Perrin reminds us that for mathematicians, with their nice analysis, "curves which have no tangent are the rule, and the well-regulated ones, such as the circle, are very interesting but highly special cases." These views *seemed* only an "intellectual exercise, without doubt ingenious, but... sterile.... Nature presents no such complications and does not even suggest the idea. *But the contrary is true,* and the logic of the mathematicians has kept them closer to reality than the practical representations of the physicists." Even the great analyst Charles Hermite once wrote to Thomas Jean Stieltjes of turning away "in fright and horror from the lamentable affliction of functions which have no derivative."

Most appreciators of mathematics have long understood this point on the molecular level. What Dr. Mandelbrot has shown is a world of visible form that is governed by the properties of Brownian motion, suitably generalized by the work of many hands, including his own original studies.

Take the analyst's "snowflake" curve (named here after its discoverer, H. von Koch), which is made in an elementary way from an equilateral triangle of unit side by replacing the middle third of each side with a "cape," itself the two jutting equal sides of a triangle a third as large as the original, and so on, repeating indefinitely. With each repetition the length increases by the factor 4/3. Such a continuous curve has in the limit no tangents, infinite length and zero area.

The mathematicians have presented a number of measures applicable to such a remarkable curve that can distinguish it overall from a circle, even though it too is merely a one-dimensional manifold. Consider two properties of the snowflake curve. It is self-similar: magnify any section and you see under the magnifier just the same form on a smaller scale. Step off its length with a pair of dividers; with dividers set at one unit distance the length you measure will be three units. Next try a one-third-unit divider setting. The length is now four units. The stepped-off length will increase continually as the divider setting grows smaller, following a simple power law, with the calculated exponent .26.

A straight line shares some properties with the snowflake. It too is self-similar, but its stepped-off length is the same for every divider opening. It has a dimension of unity, in our usual way of thinking. Here arises the book's title: it is now plausible to assign the snowflake curve a generalized dimension (applicable to self-similar curves) that is greater than that of its tame kindred, a straight line. The general formula sensibly assigns a generalized dimension 1 to the straight line and "dimension" $\log 4 / \log 3 = 1.26\ldots$ to the grander snowflake. Hence the book's title. Dr. Mandelbrot offers us a new word, fractal, suited to the tongue in both French and English, to convey this meaning of intermediate dimensionality. ("Fractional" itself would grate if it were applied to a number that is in general not rational.)

Where is the physics? Here lies the novelty and the bulk of the book. Of course we have no physical snowflake curves. Nature gives no infinities, not even within molecular collisions. There is a cutoff at the angstrom level. Still, surprises abound. The land-sea shorelines of the real world approximate the snowflake curve! Two modifications are needed. There is no such formal regularity in the coast of Brittany; it is self-similar in the statistical sense only. A map at a 10-kilometer scale shows some complex bays and peninsulas; so does a map at a .1-kilometer scale. On the average the forms do not differ (apart, of course, from any particular fishing cove we know). Empirically the stepped-off length of a coastline on maps at varying scales obeys a power law like the snowflake curve's, from a scale of hundreds of kilometers down to one of perhaps meters, where geography stops and pebbles begin. Only the exponent differs.

The book shows us simulated coasts, from which we form the impression that the dimension 1.1 is too small for the real map and 1.5 is clearly too large. It seems that mountain relief, islands, lakes, the holes in Appenzeller and Emmenthaler cheeses, the craters of the moon, the distribution of stars close to us in the galaxy and a good deal more can be described by the use of generalized Brownian motions and the idea of the fractal dimension. The cosmos itself may in the end prove to be no uniform domain but a random assembly in space, uniform only statistically, finite in mass and yet infinite in extent. Turbulent eddies, word frequencies, computer components, even the incomes of the population share some of these properties, formed out of the infinities of measure theory and the play of correlated probabilities. There is much more to be done, and it is not reasonable to assume that these results imply one mechanism.

The author has given us a delightful "macédoine" of a book, between monograph and popularization. It makes an irresistible mix, with lots of graphics, few formulas, careful and yet informal exposition, even biographical notes on the often eccentric but fascinating people who have been his predecessors in this new domain. One original philologist, a Harvard lecturer named George Zipf, published a law of word frequencies in 1949, but he "justified the Anschluss by a mathematical formula" in the same book. "One sees with him, in the clearest fashion—even in caricature—the extraordinary difficulties that attend any interdisciplinary approach."

Dr. Mandelbrot has overcome all such barriers; he is a well-known applied mathematician, trained in France but working chiefly in this country, where he is a fellow and scientific adviser to the director of research at IBM. His small volume is a good bet to become a classic of its deep yet elementary genre. Some American publisher ought to get it into English soon.

"I left my own practice for an Air Force Commission."

Elias A. Thomas, Jr.

"After 10 years of solo private practice, I found myself growing increasingly dissatisfied with what I was doing and with the direction of my life and profession.

"I did not feel as though I were fulfilling my ambitions, my hopes for a satisfying career, or my obligations to my family. The administration of my office with its ever mounting burdens of paper work and constant financial worries began to push me farther and farther away from my interests as a surgeon.

"In striving for the financial security that I wanted for my wife and children, I found it more and more difficult to spend time with them because of the amount of effort required to manage my work and office.

"I searched for a way to change this state of affairs without needing to begin all over again. In most specialties 5-10 years are required to establish an adequate practice. My children are growing up and will soon be thinking about their own careers and lives. I wanted to be able to help them as my hard-working parents had helped me.

Dr. Elias A. Thomas, Jr., Lt. Colonel, USAF. Graduated from Tufts University School of Medicine, Medford, Mass., in 1955. Surgical residency was at Boston City Hospital. Served in the U.S. Army Medical Corps. Discharged as a Major. Completed his residency at St. Elizabeth's Hospital, Brighton, Mass.

"That's where the Air Force entered the picture. The Air Force affords me the opportunity to earn a good, regular salary while practicing medicine and doing the work for which I was trained. It has relieved me of some of the administrative work load which I had in private practice and allowed me to concentrate on surgery. It has given me time and opportunity to share with my wife and children.

"There is no utopia on earth, and in the Air Force there are additional military duties and responsibilities, but these I have found to be challenging and maturing. One must always keep an open mind in all spheres of life if one is to survive and grow.

"The Air Force has many good medical facilities and an earned reputation for trying to keep physicians in clinical medicine providing direct care to patients, rather than placing them involuntarily in administrative positions as they progress to higher grades.

"The Air Force gave me that new start and zeal which I wanted for my career and the life of my family. My professional experience and prior military service count—I started at an excellent rank and salary. My initial assignment is at a fine hospital where I have the privilege of working with many stimulating and dedicated colleagues and personnel.

"There are countless benefits in the Air Force such as life insurance, housing, continuing education and advancement opportunities, travel, a 30-day paid leave every year, and medical experience difficult to match anywhere, not to mention retirement benefits.

"I am proud to say that I have already received specialized training at the Air Force's internationally respected School of Aerospace Medicine with some of the finest instructors and equipment in the world. It would be hard to duplicate this experience anywhere or in any way.

"Certainly, I had my own personal reasons for joining the Air Force. You may have some of your own. If you are considering a move to a more satisfying life and career, send in the coupon below."

For the location of the Air Force Medical Recruiting organization serving your area, fill in the coupon.

Health Care Opportunities
P.O. Box AF
Peoria, IL 61614

C-SA-115

Yes, I am interested in Air Force Health Care opportunities. I understand there is no obligation.

Name ________ (Please Print) ________ Sex (M)__ (F)__

Address ________

City ________

State ________ Zip ________ Phone ________

Soc. Sec. # ________ Date of Birth ________

Specialty ________

School ________ Year of Degree ________

Health Care at its best.
Air Force

BIBLIOGRAPHY

Readers interested in further reading on the subjects covered by articles in this issue may find the lists below helpful.

NUCLEAR-FREE ZONES

THE TREATY FOR THE PROHIBITION OF NUCLEAR WEAPONS IN LATIN AMERICA (TREATY OF TLATELOLCO). Alfonso García Robles in *SIPRI Yearbook of World Armaments and Disarmament 1969/70.* Stockholm International Peace Research Institute. Humanities Press, 1970.

SPECIAL REPORT OF THE CONFERENCE OF THE COMMITTEE ON DISARMAMENT TRANSMITTING A COMPREHENSIVE STUDY OF THE QUESTION OF NUCLEAR-WEAPONS-FREE-ZONES IN ALL OF ITS ASPECTS. Document No. A-10225 of the General Assembly, United Nations, 1975.

THE PROTEIN SWITCH OF MUSCLE CONTRACTION

CONTROL OF MUSCLE CONTRACTION. S. Ebashi, M. Endo and I. Ohtsuki in *Quarterly Review of Biophysics,* Vol. 2, No. 4, pages 351–384; 1969.

TROPOMYOSIN CRYSTAL DYNAMICS. C. Cohen, D. L. D. Caspar, D. A. D. Parry and R. M. Lucas in *Cold Spring Harbor Symposia on Quantitative Biology,* Vol. 36, pages 205–216; 1971.

STRUCTURAL CHANGES IN THE ACTIN- AND MYOSIN-CONTAINING FILAMENTS DURING CONTRACTION. H. E. Huxley in *Cold Spring Harbor Symposia on Quantitative Biology,* Vol. 37, pages 361–376; 1972.

TROPOMYOSIN-TROPONIN ASSEMBLY. C. Cohen, D. L. D. Caspar, J. P. Johnson, K. Nauss, S. S. Margossian and D. A. D. Parry in *Cold Spring Harbor Symposia on Quantitative Biology,* Vol. 37, pages 287–297; 1972.

TROPOMYOSIN: CORRELATION OF AMINO ACID SEQUENCE AND STRUCTURE. D. Stone, J. Sodek, P. Johnson and L. B. Smillie in *Proceedings of the Ninth Federation of Experimental Biological Sciences Meeting,* Vol. 31, 1974.

HIGH-GRADIENT MAGNETIC SEPARATION

MAGNETIC DESULFURIZATION OF COAL. Sergio C. Trindade and Henry H. Kolm in *IEEE Transactions on Magnetics,* Vol. MAG-9, No. 3, pages 310–313; September, 1973.

MAGNETIC SEPARATION IN WATER POLLUTION CONTROL. Christopher de Latour in *IEEE Transactions on Magnetics,* Vol. MAG-9, No. 3, pages 314–316; September, 1973.

MAGNETIC SEPARATION: A REVIEW OF PRINCIPLES, DEVICES, AND APPLICATIONS. John A. Oberteuffer in *IEEE Transactions on Magnetics,* Vol. MAG-10, No. 2, pages 223–238; June, 1974.

THE CANCER PROBLEM

END RESULTS IN CANCER: REPORT NO. 4. Edited by L. M. Axtell, S. J. Cutler and M. H. Myers. Publication No. (NIH) 73–272, Department of Health, Education, and Welfare, 1972.

FREQUENCY OF SKIN CANCER AND SOLAR KERATOSES IN A RURAL SOUTHERN COUNTY AS DETERMINED BY POPULATION SAMPLING. Z. W. Zagula-Mally, E. W. Rosenberg and M. Kashgarian in *Cancer,* Vol. 34, pages 345–349; 1974.

IMPACT OF CANCER: YEARS OF LIFE LOST DUE TO CANCER. J. L. Murray and L. M. Axtell in *The Journal of the National Cancer Institute,* Vol. 52, pages 3–7; 1974.

MUTATION, SELECTION AND CANCER. J. Cairns in *Nature,* Vol. 255, pages 197–200; 1975.

UNUSUAL MECHANISMS FOR THE GENERATION OF LIFT IN FLYING ANIMALS

QUICK ESTIMATES OF FLIGHT FITNESS IN HOVERING ANIMALS, INCLUDING NOVEL MECHANISMS FOR LIFT PRODUCTION. Torkel Weis-Fogh in *Journal of Experimental Biology,* Vol. 59, pages 169–230; 1973.

ON THE WEIS-FOGH MECHANISM OF LIFT GENERATION. M. J. Lighthill in *Journal of Fluid Mechanics,* Vol. 60, Part 1, pages 1–17; August 21, 1973.

ENERGETICS AND AERODYNAMICS OF FLAPPING FLIGHT: A SYNTHESIS. Torkel Weis-Fogh in *Insect Flight: Royal Entomological Society Symposium 7,* edited by R. C. Rainey. Blackwell Scientific Publications, 1975.

THE SUBDUCTION OF THE LITHOSPHERE

SEISMOLOGY AND THE NEW GLOBAL TECTONICS. Bryan Isacks, Jack Oliver and Lynn R. Sykes in *Journal of Geophysical Research,* Vol. 73, No. 18, pages 5855–5899; September 15, 1968.

MOUNTAIN BELTS AND THE NEW GLOBAL TECTONICS. John F. Dewey and John M. Bird in *Journal of Geophysical Research,* Vol. 75, No. 14, pages 2625–2647; May 10, 1970.

TEMPERATURE FIELD AND GEOPHYSICAL EFFECTS OF A DOWNGOING SLAB. M. Nafi Toksöz, John W. Minear and Bruce R. Julian in *Journal of Geophysical Research,* Vol. 76, No. 5, pages 1113–1138; February 10, 1971.

EVOLUTION OF THE DOWNGOING LITHOSPHERE AND THE MECHANISMS OF DEEP FOCUS EARTHQUAKES. M. Nafi Toksöz, Norman H. Sleep and Albert T. Smith in *The Geophysical Journal of the Royal Astronomical Society,* Vol. 35, Nos. 1–3, pages 285–310; December, 1973.

THE SYNTHESIS OF DIAMOND AT LOW PRESSURE

DIAMOND SYNTHESIS AND THE BEHAVIOR OF CARBON AT VERY HIGH PRESSURES AND TEMPERATURES. F. P. Bundy in *Annals of the New York Academy of Sciences,* Vol. 105, Art. 17, pages 951–982; September 10, 1964.

SYNTHESIS OF GEM DIAMONDS. General Electric Research and Development Center, 1970.

THE ROLE OF PUPIL SIZE IN COMMUNICATION

ATTITUDE AND PUPIL SIZE. Eckhard H. Hess in *Scientific American,* Vol. 212, No. 4, pages 46–54; April, 1965.

PSYCHOLOGICAL SIGNIFICANCE OF PUPILLARY MOVEMENTS. Bram C. Goldwater in *Psychological Bulletin,* Vol. 77, No. 5, pages 340–355; May, 1972.

PUPILLOMETRICS: A METHOD OF STUDYING MENTAL, EMOTIONAL AND SENSORY PROCESSES. Eckhard H. Hess in *Handbook of Psychophysiology,* edited by N. S. Greenfield and R. A. Sternbach. Holt, Rinehart and Winston, 1973.

MATHEMATICAL GAMES

WORLD MAPS AND GLOBES. Irving Fisher and O. M. Miller. Essential Books, 1944.

DISTORTION IN MAPS. E. N. Gilbert in *SIAM Review,* Vol. 16, No. 1, pages 47–62; January, 1974.

THE AMATEUR SCIENTIST

SCHEDULES OF REINFORCEMENT. C. B. Serster and B. F. Skinner. Appleton-Century Crofts, 1957.

THE ANALYSIS OF BEHAVIOR. James Holland and B. F. Skinner. McGraw-Hill Book Company, 1961.

We built this Cutlass Salon for Dick Cullen who wanted a car in the grand touring tradition that wasn't traditional.

So maybe Dick didn't have a lot of money to spend for his new car. But can you blame him for wanting the feel of a grand touring car? He found one with a reasonable price tag, and

looks that were anything but traditional. The 1976 Cutlass Salon.

Dick likes to take off and drive. He settles into the reclining bucket seat, lets the front and rear stabilizer bars reduce the lean in corners, feels the steel-belted radials dig into the turns—and goes. It's the confident, comfortable performance of a grand touring car.

Dick also liked the no-cost option of choosing the new 5-speed overdrive or an automatic transmission to go with the standard Rocket 260 V8. He picked the 5-speed to put a little more guts into his driving.

Dick likes the good feeling of having an Olds Cutlass Salon around him. It makes even short drives in the country a grand touring experience. We built it that way.

1976 CUTLASS SALON

Oldsmobile

Can we build one for you?


HE WALL STREET
What's News—
12
CHIVAS REGAL
FOUNDED
1801
BLENDED SCOTCH
WHISKY
PRODUCE OF SCOTLAND
100% SCOTCH WHISKIES
86 PROOF
CHIVAS BROTHERS LTD.
ABERDEEN SCOTLAND