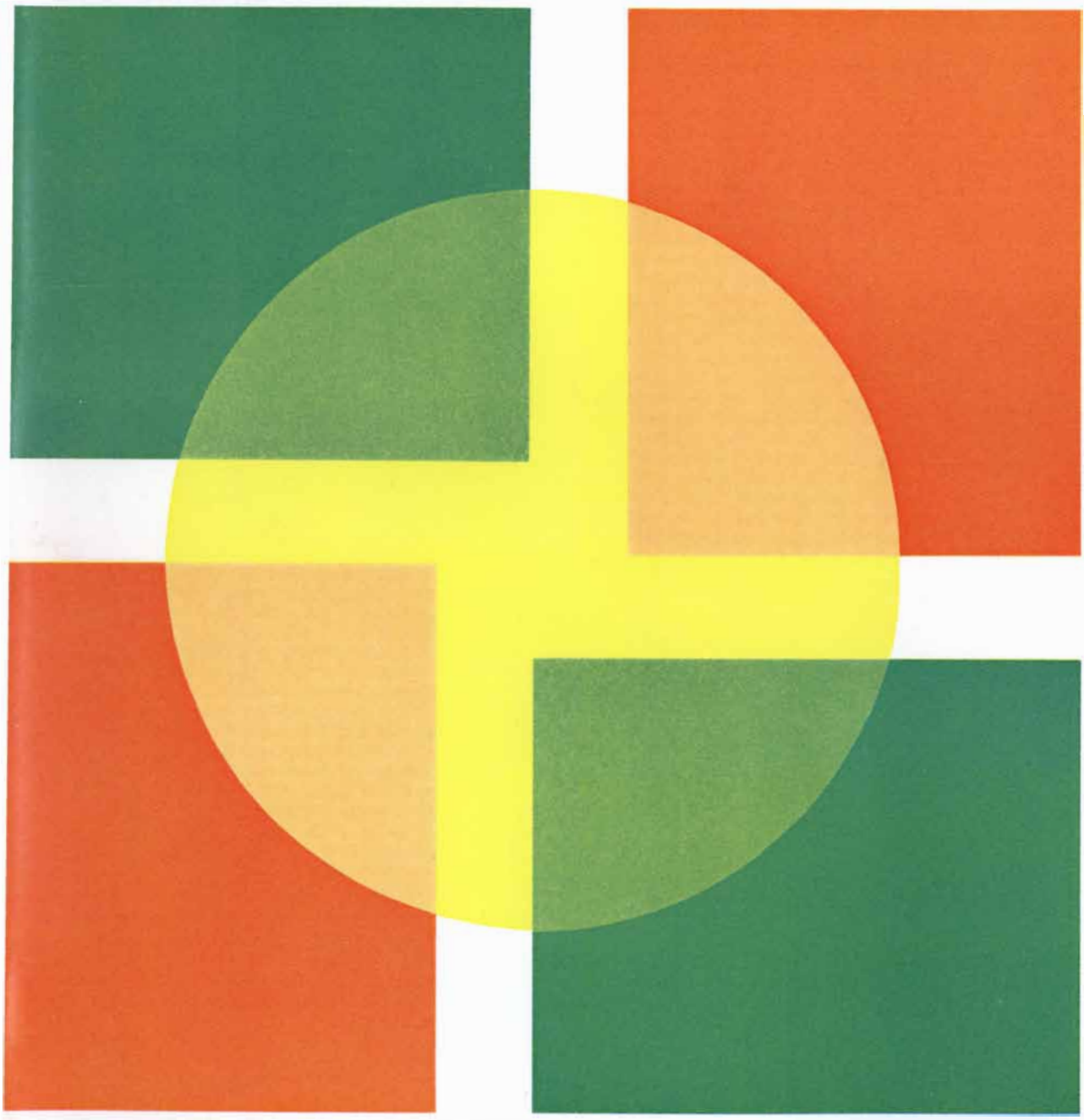
SCIENTIFIC AMERICAN



PERCEPTION OF TRANSPARENCY

ONE DOLLAR

April 1974

It's as easy to live with on the road as it is standing still.

Which is why a GMC MotorHome looks the way it does. Aerodynamic. And low—lower than any other motorhome. So it can handle crosswinds. And corners.

That's also why we've provided it with front-wheel drive, a 455-cubic-inch V-8 with automatic transmission, and tandem rear wheels with air suspension. So it can handle highway speeds. And off-highway bumps.

Not that we've ignored living comfort. In fact, 15 floor plans, in four different color-coordinated decors, are available.

If you want specifics, send for our 28-page color catalog. Simply write: GMC MotorHome Headquarters, Drawer Y, Dept. 311, Lansing, Mich. 48909.

Better yet, see your local GMC MotorHome dealer.

The GMC MotorHome. It doesn't look like a box. It doesn't ride like a truck.

GMC

The MotorHome from General Motors.

Conversation Pieces

Technically intriguing items from TRW, guaranteed to add luster to your conversation and amaze your friends.

Stars and Atoms The molecules and atoms we have here on earth are made up of large volumes of space flecked with tiny dots of matter. For example, if you were sitting in the top row of a large football stadium (e.g., the Los Angeles Coliseum), the upper tier of seats would represent the orbit of a marble-sized electron. The atom's nucleus would be a BB sitting on the fifty-yard line. Everything in between the two would be empty space.

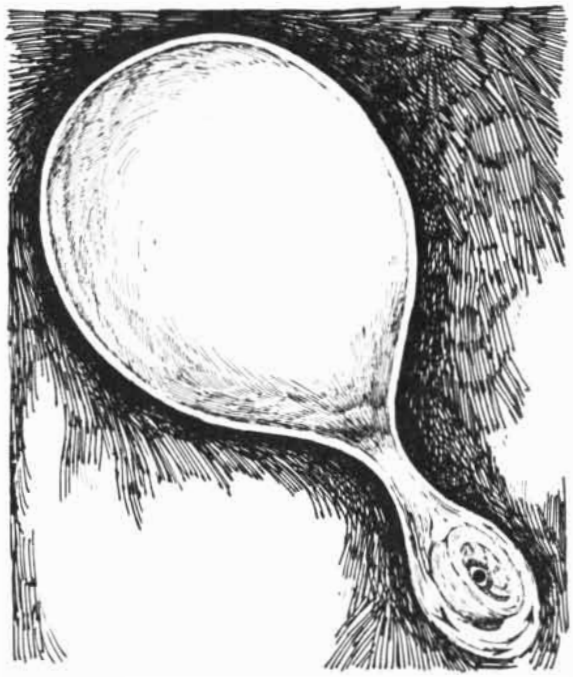
In the interior of stars, matter can be much more dense. For example, when a large star runs out of hydrogen fuel, the immense forces of gravity which have been held at bay by thermonuclear burning within the star suddenly become dominant. As Fred Hoyle puts it, the star has to pay all of its back gravitational taxes at once. The forces of this violent, almost instantaneous collapse are sometimes so great that the electron whizzing around the stadium is driven into the BB sized nucleus on the fifty-yard line. The two opposite charges cancel one another to form a neutron. Then, under the crushing force of gravity, the entire stadium fills up with marble-like neutrons.

Matter of this density exists in the heavens in the form of neutron stars or pulsars. In effect, pulsars are giant atomic nuclei in which the interatomic spaces of matter here on earth have been spectacularly reduced by gravitational collapse. We can learn about the structure of matter in such stars from the high energy radiation they emit.

Imagine now a stellar collapse so violent that the marble-like neutrons themselves are smashed together by the gravitational crush. The matter produced by such a collapse is unimaginably dense. The gravitational field of the resulting stellar object is so intense that no light (or any other kind of radiation) can escape its surface. Hence it is called a black hole. If you shined a flashlight directly on a black hole, you would see nothing for the photons of light would be sucked down its gravitational drain, never to return to your eye.

While black holes cannot be observed directly, their effects on stars unfortunate enough to be near them can be seen. Cygnus X-1 (see illustration) contains the first black hole tentatively identified. The hole is an invisible but dominant component of a binary pair of stars. It is sucking the material of its visible companion into a rotating disk. The violence of the transfer and shredding action heats up the atoms being sucked out of the visible star until they emit x-rays near the black hole, thus indirectly revealing its presence.

Today many physicists are interested in astronomy because much that we have to learn about the fundamentals of matter and energy can only be learned from the stars. That is why TRW Systems is building the High Energy Astronomy Observatory (HEAO) for NASA. The information this observatory will gather beginning in 1977 may well cause us to revise major portions of contemporary physics.



Cygnus X-1. Kip Thorne of the California Institute of Technology performed calculations leading to this model of the black hole. Our illustration is based on a painting of his model by Lois Cohen of the Griffith Observatory.

For further information, write on your company letterhead to:

TRW
SYSTEMS GROUP

Attention: Marketing Communications, E2/9043
One Space Park, Redondo Beach, California 90278

Our Green Lite℠* Service satisfied the world's toughest critic.

Not even Don Rickles can find fault with our Green Lite* Service.

It means your rental agreement is filled out and waiting when you pick up the car. All you do is show your driver's license, authorized credit card and sign.

If you're in a hurry, when you return the car, you don't have to wait in line. Charge customers can take advantage of our express check-in box.†

We also feature GM cars, and give S&H Green Stamps on all U.S. rentals. And at most locations you get two extra hours free with our 26 hour check-in day.

Sign up for National's Green Lite Service. It's so good, we left Rickles speechless.

†Available at participating locations. © National Car Rental System, Inc. 1974 (in Canada it's Tilden Rent-a-car; in Europe, Africa and the Middle East it's Europcar.) *Green Lite and the design of the semaphore are service marks of National Car Rental System, Inc.

NATIONAL CAR RENTAL

When it comes to service we mean business.

Fill out questions 1 through 20 providing all information requested (please print).
Mail to National Car Rental, 5501 Green Valley Drive, Minneapolis, Minnesota 55437.

1 FIRST NAME AND INITIAL

2 LAST NAME

3 HOME ADDRESS

4 CITY (AND PROVINCE)

5 STATE (OR COUNTRY) 6 ZIP

7 COMPANY NAME

8 COMPANY ADDRESS

9 CITY (AND PROVINCE)

10 STATE (OR COUNTRY) 11 ZIP

12 IF WE BILL YOU DIRECTLY WHERE SHOULD WE SEND THE BILL? HOME ☐ OFFICE ☐

13 DRIVERS LICENSE NUMBER

14 ISSUING STATE

15 CAR PREFERENCE

☐ LUXURY (CADILLAC OR EQUIVALENT)
☐ STANDARD (IMPALA OR EQUIVALENT)
☐ COMPACT (NOVA OR EQUIVALENT)

16 DO YOU NORMALLY PARTICIPATE IN COLLISION DAMAGE WAIVER? YES ☐ NO ☐

17 DO YOU NORMALLY PARTICIPATE IN PERSONAL ACCIDENT INSURANCE? YES ☐ NO ☐

PLEASE INDICATE BELOW WHICH CREDIT CARD HONORED BY NATIONAL YOU PREFER TO USE. SELECT ONLY ONE AND INDICATE CARD NUMBER.

		NUMBER (INCLUDE ALL LETTERS AND NUMBERS)
19 NATIONAL CAR RENTAL CARD NUMBER		
20 AMERICAN EXPRESS	AX	
CARTE BLANCHE	BL	
DINERS / AMERICAN CLUB / TORCH CLUB	DC	
BANKAMERICARD	BA	
MASTER CHARGE	MC	
TWA GETAWAY	TG	
AIR TRAVEL	AT	
TILDEN/HERTZ/AVIS	THA	
OTHER	OT	Name

Number

DATE

SIGNATURE

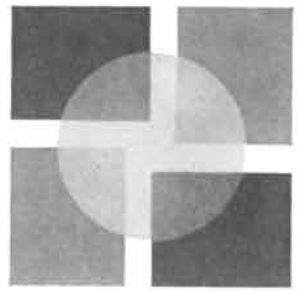
4—SA

Established 1845 **SCIENTIFIC AMERICAN** April 1974 Volume 230 Number 4

ARTICLES

DEPARTMENTS

PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND ATADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $12 PER YEAR, U.S., POSSESSIONS AND CANADA; $15 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 is designed to create an impression of transparency even though no part of the illustration is actually transparent in the physical sense of the term. The yellow circle is generally perceived as being transparent and on top of the four colored rectangles. The conditions that give rise to the phenomenon of perceptual transparency are discussed by Fabio Metelli in "The Perception of Transparency" (*page 90*).

THE ILLUSTRATIONS

Cover illustration by Jerome Kuhl

Page	Source
20–21	Victor W. Sidel and Ruth Sidel, Montefiore Hospital and Medical Center and Albert Einstein College of Medicine
22–26	Graphic Presentation Services, Inc.
27	Victor W. Sidel and Ruth Sidel, Montefiore Hospital and Medical Center and Albert Einstein College of Medicine
28	Gabor Kiss
29	F. W. Goro
30–35	Gabor Kiss
37	Department of Embryology, Carnegie Institution of Washington
38–46	Carol Woike
49	Jet Propulsion Laboratory, California Institute of Technology
53	R. A. Parham, Institute of Paper Chemistry
54–55	International Paper Company (*top left and center, bottom right*), Maurice R. Castagne (*top right, bottom left and center*)
56–58	Lorelle M. Raboni
59–61	R. A. Parham, Institute of Paper Chemistry
66	R. H. Sanders, National Radio Astronomy Observatory; G. T. Wrixon, University College, Cork
67	Allen Beechel
68–69	Hale Observatories
70	Lick Observatory
71–77	Allen Beechel
81–89	Ilil Arbel
90–98	Jerome Kuhl
100	Metropolitan Museum of Art
102	D. F. Waterhouse, Commonwealth Scientific and Industrial Research Organization
103–104	Tom Prentiss
105	D. F. Waterhouse, Commonwealth Scientific and Industrial Research Organization
106	Tom Prentiss
107–108	D. F. Waterhouse, Commonwealth Scientific and Industrial Research Organization
110–114	Ilil Arbel
116	Ewing Lusk
117–120	Jerome Kuhl

"Our new BMW is a unique combination of luxury, performance and handling. And it's amazingly easy on fuel."

Mr. William Ficker, internationally prominent yachtsman and architect of Newport Beach, California, successfully defended the America's Cup with the yacht Intrepid. His words express the feelings many BMW owners have for their cars.

"The BMW is designed from the inside out, just as any fine yacht or building program dictates. Mechanical considerations and passenger comfort are uncompromised, and then packaged in a graceful, refined body. With so many other automobiles contributing to visual pollution and highway congestion, the BMW is a refreshing relief.

"Now we particularly appreciate BMW's efficiency, as we regularly get over 20 miles to the gallon.

"Simply put, I think BMW is the best car we have ever owned."

At BMW, we build cars that above all are more *efficient.*

Road & Track calls BMW's new 3-liter engine the most sophisticated in-line six in the world. Smooth, dependable, and brilliantly responsive, yet remarkably easy on fuel.

Our fully independent four-wheel suspension gives BMW superior roadholding capabilities.

Extraordinary handling characteristics are but one of BMW's own demanding safety criteria. BMW safety features are built into the car's design, not just tacked on.

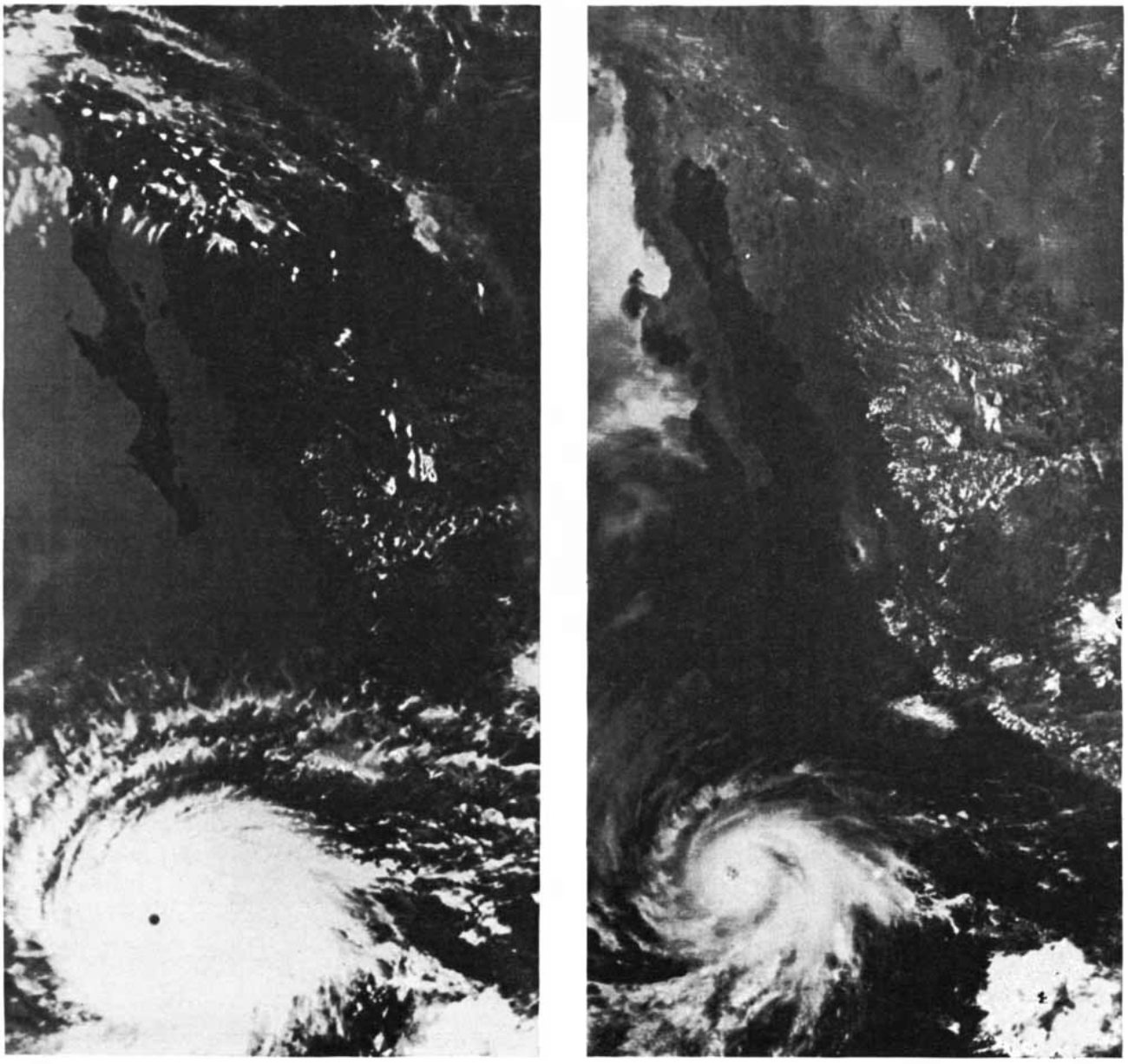
The fast, rugged, spirited 3-liter BMW is designed to achieve technological superiority in every aspect of automotive construction. In the highest tradition of the Bavarian Motor Works, it is more than a fine car. It is what luxury *should* be—the luxury of great engineering.

BMW automobiles are available in a full range of two and four door models.
Sole U.S. Importer: **Hoffman Motors Corporation.**
375 Park Avenue, New York, N.Y. 10022;
12541 Beatrice Street, Los Angeles, Calif. 90066.
Dealers Coast to Coast.
Advantageous European Delivery Program.
Write for free brochure.

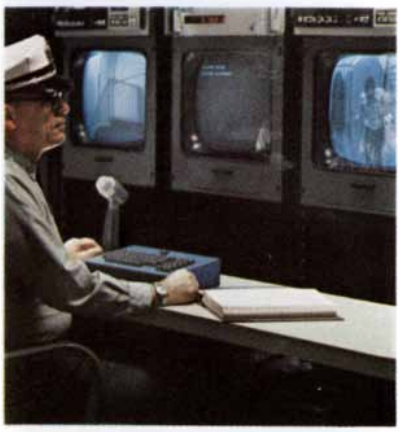
Canada: East: BMW Distributors Eastern Canada, Ltd. Ontario.
West BMW Distributors (Western) Co.
British Columbia

Bavarian Motor Works

Somebody has to provide the technology before life gets safer



Safer from killer storms.
You're looking at a tropical storm south of the Baja Peninsula from an altitude of 450 miles. The left picture, from a 2-mile resolution infrared sensor, shows temperature of the scene. The right picture, obtained simultaneously with the same resolution, shows reflected visual light. The combined information from both pictures gives the type, height, and vertical contour of the clouds. This Westinghouse multisensor system also includes a low light-level sensor that sees under moonlight conditions. Now, unequaled pictures of the earth give better warning of hurricanes, tornadoes, heavy rains and snows—even fog.



Safer from saboteurs and vandals.

Our integrated security systems are designed to protect key facilities and valuable assets against theft, fire, and malicious destruction. They combine perimeter and internal intrusion detection, with surveillance cameras that "see in the dark." We tie these systems together through a centralized command and control center. We install and service equipment, and train your employees, too.

Safer in the air.

We're building some of the world's most advanced air traffic control equipment. Westinghouse provides display equipment, surveillance radars, instrument landing systems, and beacons for use throughout the Federal Aviation Administration's control network. This technology, coupled with skilled air traffic controllers, makes your flight safer from takeoff to touchdown.

You can be sure if it's Westinghouse. For our brochure, write Marketing Manager, Defense, Westinghouse Electric Corporation, Pittsburgh, Pennsylvania 15222. DH-2-73

Westinghouse helps make it happen

Plymouth Gold Duster.
You buy the car. We'll cover the roof.

The Plymouth Gold Duster is one of the most popular cars we build. No wonder. Gold Duster is a compact, but it has room enough to seat five. And its trunk is really a whopper.

It's a "keeper."
It comes equipped with vinyl seats, thick carpeting, special sound insulation, wheel covers, whitewall tires, and special trim inside and out; the things people usually want on a car anyway.

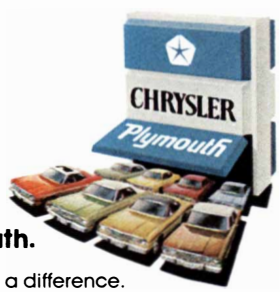
The roof is on the house.
When you order a Gold Duster with all its equipment, we don't charge your dealer for the vinyl roof. So he doesn't have to charge you.

Electronic ignition.
Something else you'll like about Duster is its Electronic Ignition System. There are no points or condenser to replace. And it helps spark plugs last up to 18,000 miles. So it saves maintenance money.

Canopy vinyl roof—no extra charge.

It goes miles and miles on a gallon of gas.
In mileage tests conducted by the United States Auto Club in January, Duster with a "Slant-Six" engine and the new standard, fuel-saving 2.76 axle got substantially better mileage than Nova and Maverick. In city driving alone, it even beat Mustang II and did a little better than Vega—both 4-cylinder cars.

You'll find a lot of things to like about Gold Duster.
Just go to your local Chrysler-Plymouth dealer, and look under the free vinyl roof.



Gold Duster.
One of eight great small car buys from Plymouth.

Extra care in engineering...it makes a difference.

Tune in NHL Hockey on NBC-TV. Brought to you by Chrysler Corporation.

LETTERS

Sirs:

It was disappointing to find no mention of the energy potential of organic municipal, farm and forestry wastes in David J. Rose's "Energy Policy in the U.S." [SCIENTIFIC AMERICAN, January]. Certainly these wastes are a real, though neglected, source of power. Waste disposal is a problem too, as is typified by the rapidly worsening conditions with New York City's ocean disposal. Perhaps using wastes to generate power is a potential solution for both problems.

The city of Los Angeles sewage plant generates all its electric power from methane gas obtained from its treatment tanks, and the surplus power is fed into the Los Angeles power net. More power could be generated with better equipment and processes....

Science has recently (December 28, 1973) run an article on the generation of methanol from a wide variety of fuels, including wastes and plant material. The article shows not only that methanol can be produced economically from these sources but also that it can be used to extend or replace our gasoline supplies.

In spite of the apparent potential of wastes to supply energy, there appears to be very little consideration of this source on a national scale. Could Professor Rose provide an estimate of the energy potential of such wastes? Neglecting economics for the moment, could a significant proportion of our energy needs be filled from wastes? Finally, what are the economics, including consideration that the wastes being used would otherwise have to be disposed of at some cost?

HAMILTON BARHYDT

Playa del Rey, Calif.

Sirs:

I read David J. Rose's article with great interest. He ignores one of the most promising developments that can contribute toward a solution of medium-term energy problems. At Pickering, just outside Toronto, there is now operating what I think could modestly be called the world's most successful nuclear power plant. It employs Canadian-designed and Canadian-built heavy-water-moderated, heavy-water-cooled, natural-uranium-fueled reactors called CANDU. There are four units generating 540 megawatts each in the Pickering plant. Not only is it at this time the highest-power nuclear-electric generating station in operation in the world; even more important, each of the four units has an outstanding record of availability. The four units were brought into service in succession over the past two and a half years. The fourth unit came to full power last May only 12 days after its startup. By an interesting coincidence, during the peak power demand in New York City last August the amount of power exported by Ontario Hydro to New York reached a peak of just under 2,000 megawatts and was very close to the output of the Pickering Nuclear Plant. The Pickering reactors are of a pressure-tube type rather than a pressure-vessel one and are believed by many engineers to have overall advantages over the U.S. designs. The design, moreover, allows all the advantages of on-power refueling, which is routine at Pickering.

The economics of the CANDU reactor have been a source of argument among experts for years. Its capital cost is higher than that of the U.S. light-water reactors but its fuel costs are very much lower. The marginal cost of operating the Pickering reactors is less than one mill per kilowatt hour. Most calculations show that the 30-year cost of owning and operating a CANDU reactor is lower than that of a comparable U.S. reactor.

Whether the CANDU is a little cheaper or a little more expensive, however, is not the point at issue. The main point is the simple fact that for every pound of uranium mined and put through a CANDU reactor there will appear twice as many kilowatt-hours of electricity as there would if the same uranium had been put through the U.S. reactor system. In addition the CANDU system allows the introduction of thorium into the core with further great increases in efficiency. In fact, a CANDU reactor containing thorium could be made to breed, although this would not be the most efficient way of using it.

Canada has at Pinawa, near Winnipeg, an operating prototype of a heavy-water-moderated, organic-liquid-cooled reactor with a record of outstanding performance over more than five years. This design promises lower capital cost, higher operating efficiency and greatly reduced radiation exposure of operating and maintenance personnel. Canada has recently decided not to proceed at the present time with a larger version of this organic-liquid-cooled reactor. There is probably no more promising venture for the U.S. than to take up this technology and to build an organic-cooled CANDU reactor with thorium in the core. Such a reactor would have most of the advantages of the breeders and requires far less development to achieve early operation. Any adoption of the CANDU designs in place of conventional light-water reactors would result in more efficient use of uranium and would also reduce the load on enrichment facilities.

OMOND M. SOLANDT

Toronto, Ont.

Sirs:

...Rose leaves an important point stranded. Who is to make the necessary decisions in the U.S. energy policy? He hints at a national debate to clear the energy strategies. This would imply a thorough knowledge of alternatives, and surely that does not exist. How can the average American make a decision on the relative merits of different types of fission reactors when he has no foundation on which to base a decision? And are politicians any more qualified?...

JON J. KOS

University of Missouri
Rolla, Mo.

Sirs:

In one article of modest length I could not include all the alternatives in a topic as complex as energy, at least not without merely presenting unarticulated lists. Many people who have written bring up topics that would have appeared in a longer account.

Producing energy from organic wastes is an attractive proposition to an increasing number of cities, as Mr. Barhydt and other correspondents point out. The total available energy is not large, as a simple calculation shows: household refuse amounts to about five pounds per person per day, with an energy equivalent of (about) 18 cubic feet of natural gas. These wastes could be processed in a chemical digester to yield the gas, which would augment the natural supply. If all such refuse were treated (an obvious impossibility), the natural-gas augmentation would be about 7 percent, or 2 percent of total national energy needs. Methanating animal wastes from feedlots in the Middle West is probably the best application of this idea, because the wastes are already gathered at relatively few sites. The principal attraction of the scheme, however, is not the additional

energy but the reduced environmental pollution; the wastes *are* objectionable pollutants precisely because they contain excess free energy, in both the thermodynamic and the economic sense. The wastes become relatively inert after their energy is extracted.

Dr. Solandt describes a series of successful Canadian nuclear reactors. The United Kingdom now seriously considers adopting them in preference to either the American light-water types or the indigenous gas-cooled ones. I had seriously considered extending the fission-reactor assessment to include the Canadian species, but then to preserve balance I would have had to enlarge the breeder-reactor discussion to include more on the gas-cooled fast reactor, the molten-salt reactor and the European prototypes. There was no room for that much, so that I arbitrarily limited the reactor part to types for which strong economic champions existed in the U.S. I should in fact like to see a good and fresh comparative assessment made among the principal contending reactor types, including Canadian ones. Then the U.S. and other countries will have a clearer idea of what they are doing or what they have been committed to.

Professor Kos raises a basic and difficult question. He says, by implication, that the U.S. Atomic Energy Commission's assessment of nuclear power cannot be taken entirely at face value, and that the same is true for other groups, both public and private, on other options. I agree. Some improvement should come from agencies being forced to consider broader issues. More will come from having at least some mechanism for making broad decisions about energy rather than no mechanism at all (William E. Simon's office v. nothing, for example). A more serious impediment to balanced decision-making has been the lack of proper technological advice to Congress. Advice it has aplenty, too often from self-serving sources. I have great hopes for improvement through the operation of the new Congressional Office of Technological Assessment; it should serve a very important role.

DAVID J. ROSE

Massachusetts Institute of Technology
Cambridge

Sirs:

In the introduction to our article on laser-induced narrow optical resonances ["Laser Spectroscopy," by M. S. Feld and V. S. Letokhov; SCIENTIFIC AMER-

ICAN, December, 1973] we erred in stating that laser spectroscopy using molecular beams is limited in resolution to about 1/50 the Doppler width. In recent laser-molecular-beam experiments Shaoul Ezekiel of the Massachusetts Institute of Technology, Michael M. Hercher of the University of Rochester, H. Walther of the University of Cologne and others have obtained high-resolution spectra, in one case (Ezekiel) with better than 10^{-3} the Doppler width. This shows that molecular beams are also a promising tool for high-resolution laser spectroscopy.

For completeness we should also like to mention that the early observations of the Lamb dip in the helium-neon laser were done by Ross A. McFarlane, W. R. Bennett, Jr., and Willis E. Lamb, Jr., at Bell Laboratories and by Ali Javan and Abraham Szöke at M.I.T.

M. S. FELD

Cambridge, Mass.

V. S. LETOKHOV

Moscow, U.S.S.R.

Scientific American, April, 1974; Vol. 230, No. 4. Published monthly by Scientific American, Inc., 415 Madison Avenue, New York N.Y. 10017; Gerard Piel, president; Dennis Flanagan, vice-president; Donald H. Miller, Jr., vice-president and secretary; George S. Conn, treasurer; Arlene Wright, assistant treasurer.

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

Advertising correspondence should be addressed to Harry T. Morris, 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 25 cents.

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

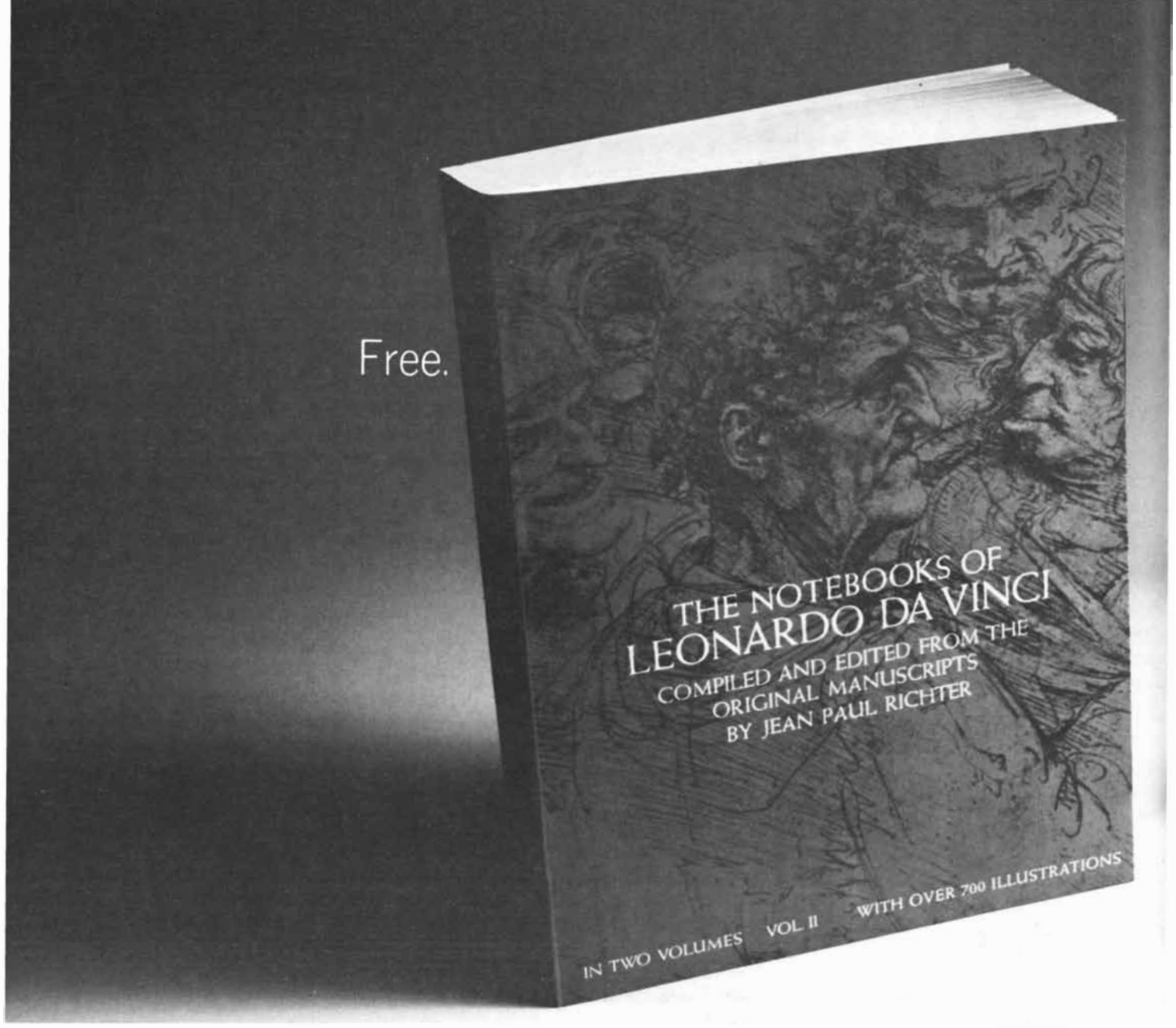
☐ Please enter change of address.

☐ Please enter subscription.
U.S. and Canada:
☐ 1 year, $12 ☐ 2 years, $22 ☐ 3 years, $30
All other countries:
☐ 1 year, $15 ☐ 2 years, $27 ☐ 3 years, $37

NAME

NEW ADDRESS

OLD ADDRESS



Monroe presents 500 for 30 minutes of

Introducing the totally new 1920.

In order to introduce a scientific and engineering electronic calculator so new and different that it defies description, we've come up with a unique kind of offer. One of the most fascinating printed works of the century. A 500 page volume of the long lost notes of Leonardo da Vinci.

Long available only to serious scholars and meticulously penned in a mysterious mirror writing, this newly translated volume contains a wealth of information. Leonardo's voluminous notes on sculpture, architecture, astronomy, anatomy, physical geography, engineering, and mechanics showing him to be centuries ahead of his time. Sold in fine bookstores for five dollars. And it's yours absolutely free.

In return for this gift, we want to show you a new electronic computing device so unique that it must be tried to be believed.

A device that allows you to work in algebraic form. And that's only the beginning. With the 1920 you work with the simplest keyboard, so comfortable that it can become second-nature.

Simpler than any you've ever tried, but with all the power you'll ever need. Plus a display you can read even from across a room.

And in 30 minutes we can show you what we mean.

And for your time, with no further obligation, we will present you with a card you return to us. As soon as we receive it, we'll eonardo Volume by return mail.

pages of DaVinci your time.

Monroe, The Calculator Company
550 Central Avenue, Orange, New Jersey 07051

MONROE
Litton

☐ Yes, I'd like to see the 1920. Please call for an appointment.
☐ Yes, I'm interested, but send further information on your offer by return mail.

SA

Name_____
Title_____
Company_____
Address_____
Phone_____
City_____ State_____ Zip_____

Our supply of the Notebooks is limited. So to be sure to get yours early, act now.

To reserve your copy, and schedule a demonstration please return the coupon or call your local Monroe office. We're sure that both our product and the Leonardo Notebook will amaze you. And provide for time well spent.

MONROE
The Calculator Company.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

APRIL, 1924: "The now famous Teapot Dome oil reserve was set aside for the Navy in 1915. It is situated about 30 miles north of the town of Casper, Wyo. The region consists of the Teapot Dome fields themselves and the adjacent Salt Creek fields, which were being privately worked. A very important question in its bearing on the present controversy is whether or not Teapot Dome was threatened by wells being drilled in the Salt Creek property. The ideal form for an oil-bearing bed or set of beds is dome-shaped, like a basin or a set of basins turned upside down. In the Salt Creek and Teapot Dome fields we have two domes with a depression between. The Naval Reserve includes not only all of Teapot Dome but also the south end of the Salt Creek anticline, or arch. Because the boundary of the reserve is above the saddle or depression between the domes, wells driven north of the boundary would drain oil from the Salt Creek anticline and not from Teapot Dome."

"It seemed to the pioneers who first sighted the bacteria that life could not live in narrower confines. Yet today we hear of living creatures far smaller than most bacteria but of even greater importance in the havoc they work on the human race. To this group of living morsels has been given the name 'filter passers.' Among the diseases believed to be caused by them are influenza, smallpox, hydrophobia, scarlet fever and measles. Even the plant world suffers from them, for mosaic disease in tobacco plants is due to them. The size of a filter-passer is emphasized by the property from which it has derived its name—the power of passing through a filter. Every filter passer can pass through a filter of compressed diatoms; some can negotiate a filter made of unglazed porcelain, which has much smaller pores. They are tiny specks; that is all that can be said of them."

"One of the most notable triumphs of modern astronomy has been the development of the new photometric methods of determining the distances of the stars. If we can find out, in any reliable way, how bright a star is, it is a very simple matter to figure out how far off it is. One test of the real brightness of stars depends on the fact that among the variable stars of a certain type (usually called the Cepheid variables) there is a very close relation between the period and the actual brightness. A dozen or more variable stars of the Cepheid class lie near enough to us so that we can use the drift caused by the sun's motion to find their distance. It appears that a star with a period of eight days is about 1,000 times as bright as the sun, and the brightest stars, with periods of 100 days, run up to the enormous figure of 50,000 times the sun's light. With this established it becomes a straightforward matter to calculate the distance of any variable star of the Cepheid type, or of any cluster of stars in which such variables can be found."

"Who has not burned toast? The electric toaster is the greatest offender in this matter, but now the problem has been solved for the housewife, if she can pay the price for an automatic toaster. Every slice is a rich golden brown. There is no watching or turning, for when the toast is finished, the racks raise it out of the oven and the current is shut off. This seeming marvel is accomplished by a timing device."

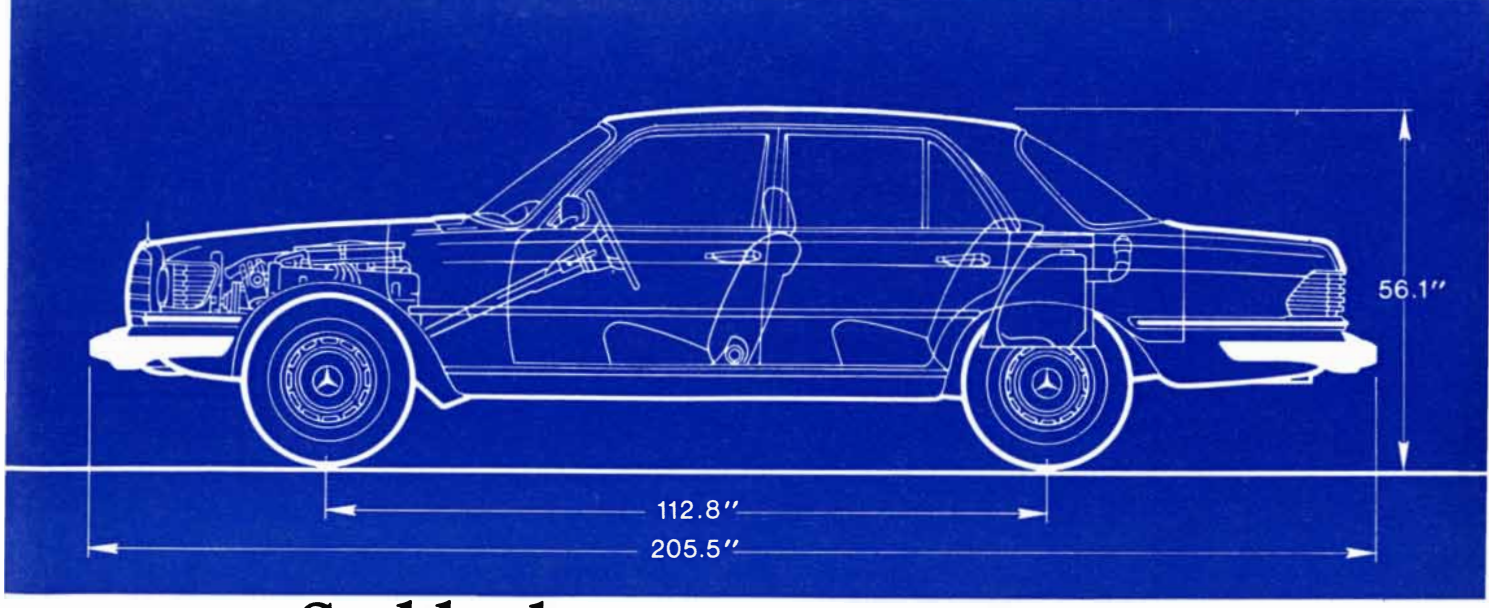
SCIENTIFIC AMERICAN

APRIL, 1874: "Spectral analysis, confirmed by the chemical analysis of meteorites, has familiarized us with the idea that all the bodies of nature, the planets of our system as well as the suns most distant from us, are composed of the same elements and animated by the same physical forces, even to the most delicate and minute details. Since therefore these same forces act under our eyes as essential agents of life, we are naturally led to consider the conditions of organic existence on our globe as applicable to the circumstances of other species. If our earth is inhabited, why not the other orbs that fill up space, seeing that the same matter is everywhere present? The eminent French astronomer Faye, in considering this question, concludes that although the chemical elements necessary to life are largely extended throughout the universe, the requirements of temperature, water and atmosphere exclude almost all heavenly bodies, so that we are barely able to cite two planets of our system, Mars and Venus, where the conditions of life have any shadow of probability of existing, and on the only globe of which we can speak with certainty, the moon, we know them to be utterly absent."

"Soundings were made in the Pacific Ocean during last year by Captain George T. Belknap of the steamer *Tuscarora,* with reference to the projected laying of a telegraphic cable from this coast to Japan. Remarkable depths, which have no parallels in the plateaus of the Atlantic, have been discovered. A sudden descent to a great depth is continuous down the entire coast, varying from 20 to 70 miles out. In the latitude of San Francisco Bay the bottom suddenly descends to a depth of two miles. Thirty miles off the Golden Gate the bottom is reached at 100 fathoms; at 55 miles it has descended to 1,700 fathoms, and 100 miles out the enormous depth of 2,548 fathoms has been measured without reaching bottom."

"The specialized structure of the small group of animals of which the horse is the chief member used to form one of the strongest supports to the theory of specific creation. Curiously, the same order of animals is now among the best supporters of the theory of evolution. When Darwin wrote his *Origin of Species* enough was known to justify, to his mind, the hypothesis that the peculiar legs and feet of those animals had been produced by a long course of variations from the less specialized forms of former periods. By his opponents this confident belief in what no one had ever seen was taken as evidence only of his abandonment to theory. Increasing knowledge has proved Darwin right and his critics wrong. One by one the predicted connecting links have been found, in later years particularly in our own country, where the line of descent appears to be more direct and the record more complete than has been found in the Old World. It now reaches clearly to the Orohippus of the Eocene, an animal about the size of a fox."

"A device has been designed by Mr. Stewart Harrison of London, England, for the purpose of automatically flooding with water the interior of a building attacked by fire. A valve is held in its seating, so as to prevent the efflux of water, by an annular wedge of fusible metal. In case of a fire occurring by accident, as soon as the boiling point of water is attained the fusible plug melts, the valve falls and water in continuous streams is at once discharged."



Suddenly, everyone wants a car with sensible size, reasonable weight, good mileage and safety. The Mercedes-Benz.

For some automobile manufacturers, a car with all of these qualities is years away. Every Mercedes-Benz Sedan has them all right now.

Mercedes-Benz has engineered its products with efficiency in mind for decades.

Sensible size, reasonable weight

An example is the 450SE Sedan. It's smaller than America's best-selling sedan outside, but *inside,* big car room.

What about gasoline mileage? A 450SE gets better mileage than any domestic luxury sedan. A combination of reasons tells you why:

The 450SE has a unique engine. An overhead camshaft configuration, this V-8 is smaller than any domestic example. Its electronic fuel injection measures out fuel only as it's needed.

The engine has an easier job. The 450SE weighs nearly half a ton less than the lightest domestic luxury car. Its shape was developed in the wind tunnel to slice through the air with low air resistance.

A smaller, more efficient engine, reasonable weight and a low drag shape. The reasons why a Mercedes-Benz 450SE gets up to 25% better mileage than domestic luxury sedans.

Safety

The Mercedes-Benz 450SE Sedan combines over fifty separate and significant safety elements, all engineered to complement one another.

No other manufacturer makes safety a more important element in car design. The results of this commitment are to be found in every model Mercedes-Benz builds.

The best resale value in America

Advanced engineering and fuel economy. Two reasons why, after three years, Mercedes-Benz automobiles have a better resale value than any car made in America. And not just better—*the best.* Who says so? Leasing sources—and the automobile industry itself.

All of a sudden everyone wants design efficiency in an automobile. For that matter, these days everyone *needs* it. Mercedes-Benz has had it for years. Perhaps it's time you experienced it. Mercedes-Benz automobiles.

Engineered for today's needs.

Mercedes-Benz

THE AUTHORS

VICTOR W. SIDEL and RUTH SIDEL ("The Delivery of Medical Care in China") are husband and wife; Victor Sidel is chief of the department of social medicine at Montefiore Hospital and Medical Center in New York and professor of community health and chairman of the department of community health at the Albert Einstein College of Medicine, and Ruth Sidel is a psychiatric social worker. Victor Sidel was graduated from Princeton University and the Harvard Medical School and received postgraduate training in internal medicine and in preventive and social medicine. From 1964 to 1969 he was head of the community medicine unit at the Massachusetts General Hospital and a member of the faculty in preventive medicine and in medicine at the Harvard Medical School. Ruth Sidel was graduated from Wellesley College and the Boston University School of Social Work. She has worked for several years in the care of emotionally disturbed preschool children, most recently as social work supervisor in the comprehensive child-care project associated with the Albert Einstein College of Medicine.

P. K. TIEN ("Integrated Optics") is head of the department of electron physics research at Bell Laboratories. Born in China, he received his bachelor's degree in electrical engineering at the National Central University in Nanking in 1942. After his graduation he joined several friends in organizing a chemical plant producing soy sauce in Shanghai; he became vice-president and chief accountant of the company. In 1946 he came to the U.S. for graduate study, receiving his master's degree and his Ph.D. at Stanford University in 1948 and 1951 respectively. He joined the electronics research laboratory of Bell Laboratories in 1952. Tien writes that he and his two college-age daughters, one studying applied mathematics and the other physics, "often discuss mathematics and physics together and are not satisfied until a simple explanation of the problem is found." One of his hobbies is drawing; he has drawn most of the illustrations that he displays with his numerous talks and lectures on optoelectronics.

ALAN E. BEER and RUPERT E. BILLINGHAM ("The Embryo as a Transplant") are at the Southwestern Medical School of the University of Texas; Beer is associate professor of cell biology and of obstetrics and gynecology, and Billingham is professor of cell biology and chairman of the department of cell biology. Beer obtained his M.D. at the Indiana University School of Medicine and then took specialty training in obstetrics and gynecology at the University of Pennsylvania School of Medicine. Billingham, who was born in England, received his education at the University of Oxford. Moving to the U.S. in 1957, he spent 14 years at the University of Pennsylvania. In 1961 he was elected a Fellow of the Royal Society.

F. KEITH HALL ("Wood Pulp") is director of primary processing and nonwovens research at the Corporate Research Center of the International Paper Company. Born in England, he was graduated from the University of Manchester in 1951, obtaining his doctorate at the University of Leeds in 1954. In 1956 he joined Courtaulds of Canada as a research chemist. After two years he was transferred to the manufacturing side of the business, eventually becoming deputy plant manager. "I am a keen amateur naturalist," he writes, "and a fanatical fisherman. Athletically I suppose I play a fair game of tennis (I used to play on the university teams long years ago), although of late years the quarterfinals or semifinals of the local club championship are about as far as I go. I remain an omnivorous reader and enjoy classical music very much."

R. H. SANDERS and G. T. WRIXON ("The Center of the Galaxy") are respectively a member of the staff of the National Radio Astronomy Observatory and a member of the faculty at University College, Cork. Sanders was graduated from Rice University in 1966 and obtained his Ph.D. (in astrophysics) from Princeton University in 1970. "My spare-time activities," he writes, "include walking along the Blue Ridge, going to movies (preferably foreign), folk dancing and reading history (usually about the Russian Revolution or the American Civil War)." Wrixon, who was born in Ireland, was graduated from University College, Cork, in 1961. His master's degree is from the California Institute of Technology and his Ph.D. from the University of California at Berkeley. For five years until returning to Ireland recently he was with Bell Laboratories. "I am married to a girl from Long Island," he writes, "and we have two little boys who demand, but do not receive, all our attention. We enjoy sailing and walking in the country. At present we are busy refurbishing an old Georgian house that we have bought in the countryside."

JAMES W. VALENTINE and ELDRIDGE M. MOORES ("Plate Tectonics and the History of Life in the Oceans") are respectively professor and associate professor of geology at the University of California at Davis. Valentine, who obtained his bachelor's degree at Phillips University in 1951 and his master's degree and doctorate from the University of California at Los Angeles in 1954 and 1958 respectively, taught at the University of Missouri for six years before going to the University of California at Davis in 1964. "I collect the works of Charles Darwin, all issues in all languages," he writes, "and am interested in (and collect when possible) traditional African art." Moores received his bachelor's degree at the California Institute of Technology in 1959 and his Ph.D. from Princeton University in 1963. At Davis, where he began in 1966, he serves as chairman of the department of geology. He enjoys playing the cello and the piano and also describes himself as interested in "history, politics, skiing and hiking."

FABIO METELLI ("The Perception of Transparency") is professor of psychology at the University of Padua, where he works as an investigator in the university's Institute of Experimental Psychology.

D. F. WATERHOUSE ("The Biological Control of Dung") is chief of the division of entomology of the Commonwealth Scientific and Industrial Research Organization in Australia, with which he has been associated since 1938. "My principal research over the years," he writes, "has been in various fields of insect physiology, but I have little time for active research nowadays. In recent years I have been heavily concerned with encouraging the development of pest management in Australia, and particularly in stimulating a vastly increased effort in biological control as an important component of such management. My principal hobby is Gyotaku, the ancient Japanese art of fish printing. Several of my prints have been published recently in Japan in a volume containing works of the modern masters of this art. I am one of the few foreign members of the exclusive Tokyo Gyotaku-no-kai (Friends of the Fish Print)." Waterhouse's degrees (bachelor, master and doctor of science) are from the University of Sydney. He is a Fellow of the Royal Society.

The Pace Company developed a national energy model...

Three years ago the Pace Company, a consulting and engineering firm in Houston, predicted an energy crisis. They called on McDonnell Douglas Automation Company (MCAUTO™) for computer power to run a sophisticated mathematical model that simulates U.S. energy systems and international petroleum balances.

Objective: To provide major suppliers and users of energy with information they need to find solutions to the crisis.

To run the model, Pace uses MCAUTO's 4-million byte computer through a teleprocessing terminal located in their offices.

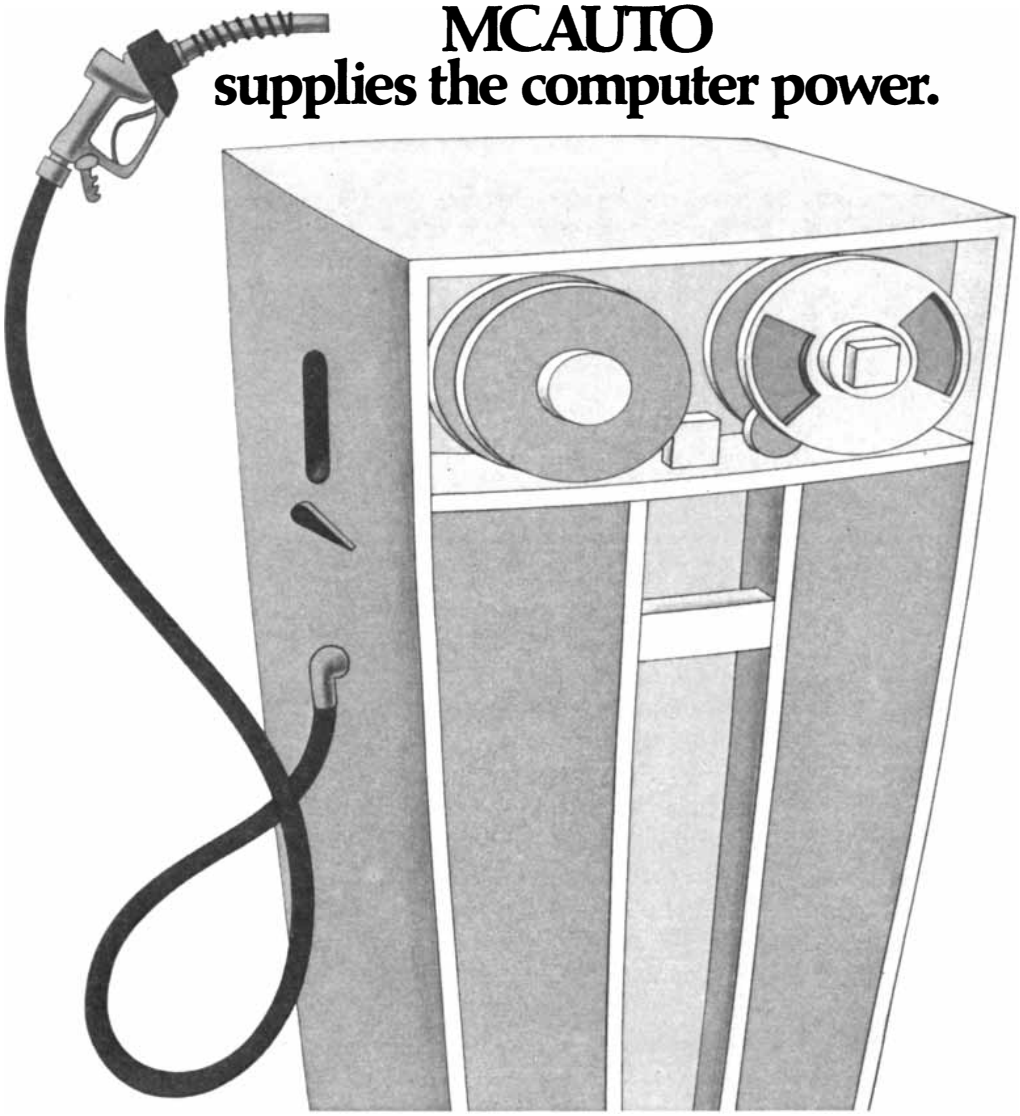
Result: Pace is providing its clients with fast answers to "What if?" questions posed to the computer. Answers are based on rapidly changing information on energy conditions.

Hundreds of companies – in business, science and engineering – with more than 2,000 terminals use MCAUTO's nationwide teleprocessing network to reach our large-scale computers. They get quick solutions to complex problems. And they pay for only what they use. Without upgrading their computers and supporting larger staffs.

In 1973, we served over 1300 clients with a broad range of data utility services from five major computer centers. We can do the same for you.

Put 14 years of computer experience to work on your problems. Start by writing for more information to: W. R. Vickroy, McDonnell Douglas Automation Company, Dept. KR-655, Box 516, St. Louis, MO. 63166. Or call (314) 232-8021.

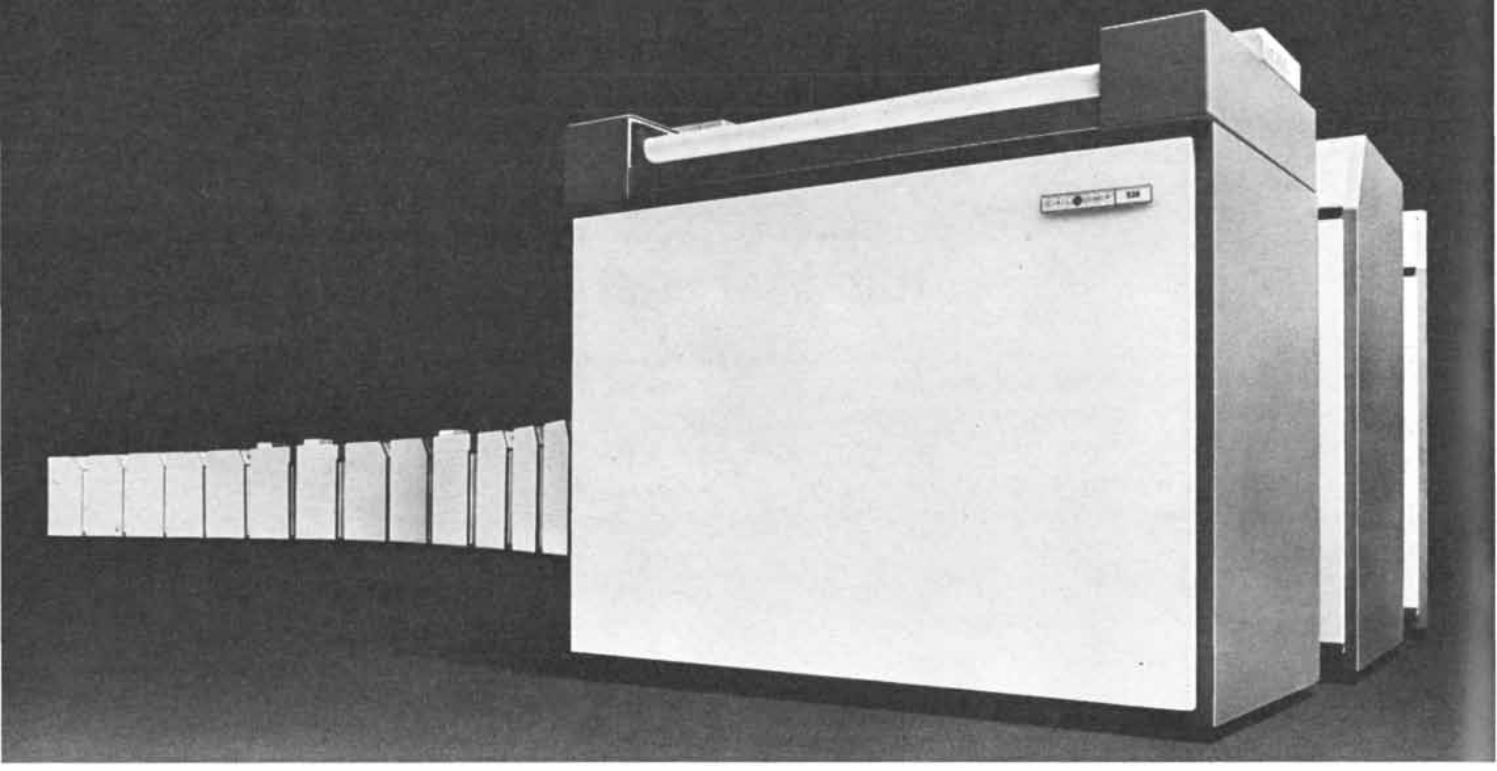
MCAUTO supplies the computer power.



MCDONNELL DOUGLAS AUTOMATION COMPANY

ST LOUIS CHICAGO CLEVELAND DENVER DETROIT HOUSTON LOS ANGELES NEW YORK PHILADELPHIA PITTSBURGH SAN FRANCISCO SEATTLE SIOUX CITY WASHINGTON, D C

A DIVISION OF MCDONNELL DOUGLAS



When you're the leader, how do you follow the leader?

We introduced digital plotting.

For the last ten years, our drum plotters have set the standards of their industry.

Our 565, and the models we've built around it, have made us the largest manufacturer of drum plotters in the world.

But we've known for a long time that someone would come along with something new one day. What we've been working on, is making certain that the new leader would still be us.

Starting now, you'll measure drum plotters by our two new models.

First, the 936. It's faster and it has greater plotting quality than the unit it replaces.

And, it costs less.

Next, our top of the line 1036. It's almost twice as fast as the 936. And again, its plotting quality is higher and its price is lower than the unit it replaces.

We've increased performance and decreased the price of the drum plotter. And that's going to be a hard act to follow.

CALCOMP

Call or write California Computer Products, Inc., SA-M4-74, 2411 West La Palma Avenue, Anaheim, California 92801. (714) 821-2011.

Established 1845 SCIENTIFIC AMERICAN April 1974 Volume 230 Number 4

The Delivery of Medical Care in China

The main feature of the Chinese system is an integrated network of neighborhood stations that serve the functions of preventive medicine, primary medical care and referral to larger centers

by Victor W. Sidel and Ruth Sidel

The health of the Chinese people has changed greatly over the past few decades. The change is apparent in many ways, some that are reflected in the anecdotes that returning visitors tell and others that are evident in the few available statistics. In the second category perhaps the most dramatic is the reported change in the principal causes of death. In China during the 1930's and 1940's the leading causes of death were on the one hand infectious and parasitic diseases and on the other complications of malnutrition (in many cases a euphemism for starvation). Today it appears that the leading causes of death, at least in a large city such as Shanghai, are the same as those in the developed nations of the West: cancer, stroke and heart disease.

On the surface this change might seem merely to substitute one termination of life for another. Its significance, however, is inescapable: it is evidence that the people of China are dying at progressively higher ages. Other statistics point to the same conclusion. For example, in the 1930's the life expectancy of a newborn child in Shanghai was some 40 years. Data compiled by the city's Bureau of Public Health in 1972 suggest that today the life expectancy is more than 70 years. The data are all the more remarkable when one considers that they are from what is one of the poorer and technologically less developed nations in the world.

In some ways the anecdotal evidence is even more convincing than the statistical, much of which is fragmentary and unconfirmed. Almost all visitors from abroad who travel the urban lanes and country paths they knew 25 years or more ago comment on the change in appearances. Where sick children and ailing adults were once a commonplace sight, today they are rarely seen; both children and adults appear to be in excellent health.

How have the Chinese, with their limited technical resources, managed to do this? The answer is that their revolution of 25 years ago gave rise to many changes in the Chinese way of life, including changes in the methods of delivering health care. To understand the nature of these changes and to see today's pattern of Chinese medical and social services in context, it is necessary to know something about former conditions.

In 1949 the population of China was estimated to be 540 million, some 85 percent of which was rural. With respect to the practice of what the Chinese call "Western medicine," there were then in China 40,000 Western-style physicians at most and perhaps 90,000 beds in Western-style hospitals. If these medical resources had been evenly distributed, the ratio of physicians to potential patients would have been one to every 13,000 and of beds per patient one to every 6,000. Instead, of course, most of the resources were concentrated in a few cities, and even there most of the population depended on practitioners of traditional Chinese medicine for such care as they received.

Beginning in 1949 China's new government confronted this deficiency in health resources by initiating a dual program. Some of the strategies adopted by the new Ministry of Health were unique to China; others were the same strategies that have been adopted by many other technologically underdeveloped nations. In the first category were innovative efforts to involve the bulk of the population in "mass movements." These were aimed primarily at improving public health and sanitation. A further innovation was an attempt to enlist the practitioners of traditional medicine in overall health programs.

In the second category were programs that emphasized the training of large numbers of new health workers. The principal efforts to increase the numbers of health personnel were directed on the one hand toward the training of "middle level" health workers and on the other toward the establishment of "centers of excellence." These centers were urban training facilities that were expected to pioneer new medical techniques and also to provide a flow of skilled personnel to areas of special need.

Following Russian models, the Ministry of Health set up a number of "middle medical schools." Students who had reached the intermediate level of the secondary school system were sent to

middle medical schools for a three-year course that prepared them to work as "assistant doctors." This is a category comparable to the Russian feldsher, a physician's assistant who is expected to act as a physician when necessary. At the same time the middle medical schools trained other personnel as nurses, midwives, technicians and pharmacists.

The Ministry of Health also expanded the existing program of "higher" medical education. Some medical schools were moved from the coastal cities to the interior and some new schools were founded. Here again Russian models were followed. Separate faculties were responsible for pediatrics, for general medicine, for stomatology (dentistry and other treatment of diseases of the mouth) and for public health. The period of study at the higher medical schools was five or six years; the students were recruited from those who had completed the senior level of secondary school. At the pinnacle of this higher-educational system was the China Medical College in Peking. Here an eight-year curriculum was offered, its objective being to train teachers and research personnel.

By 1965 the ambitious program had produced more than 100,000 new physicians and some 170,000 assistant doctors. In the same period, however, China's population had grown from 540 million to about 725 million. Even though the physician-to-patient ratio of one to 5,000 was substantially better than it had been in 1949, it was still far from the ratio of one to 1,000 or better that is typical of richer nations. The progress of the Health Ministry's innovative programs also had been good but large gaps remained. Even though some traditional methods had been adopted, practitioners of traditional medicine were still looked on as second-class physicians. By the same token the mass-movement programs had successfully attacked a number of public-health problems, but professional or, as the Chinese say, "expert" health workers continued to dominate Chinese medicine. Perhaps most important of all, the center of gravity of medical care remained in China's urban areas. The inadequate level of health care delivered to the rural Chinese who are the vast majority of the nation's population led to criticism of the Ministry of Health. This was climaxed in 1965 by an action that proved to be a forerunner of the "Cultural Revolution" of 1966–1969: publication of what is now known as Chairman Mao's June 26 Directive. "In medical and health work," the directive ordered, "put the stress on the rural areas."

Both of us work in the field of medical and social services in the U.S. and have studied the delivery of these services in a number of other countries. We thus felt ourselves fortunate to be among the first Americans to be invited to visit China by the Chinese Medical Association soon after the "Ping-Pong breakthrough" of 1971. By then the Cultural Revolution had wound down and we were eager to see what its effects had

BIRTH-CONTROL POSTER shows a woman "barefoot doctor," her medical kit slung over her shoulder, holding a book titled *Late Marriage and Plan for Birth Control.* She points to the first of six vignettes; the captions, somewhat abbreviated, read "To study and apply Chairman Mao's Thought," "To consolidate the dictatorship of the proletariat," "To prepare against war or national disaster and for the people," "To support world revolution," "To raise successors to the revolution" and "To grasp revolution, promote production and prepare against war." The title of the poster is "Plan for Good Birth Control for Revolution"; the five characters on the lid of the medical kit repeat a slogan enunciated at every level of Chinese health care: "Wei ren min fu wu," or "Serve the people."

been. We were able to observe several aspects of contemporary Chinese medicine, including the delivery of medical care, in September and October of 1971 and then again in greater depth in September and October of 1972. We traveled with representatives of the Chinese Medical Association to the nation's two largest cities, Peking and Shanghai, to a number of provincial cities and towns and to rural areas in both the densely populated coastal regions and the less crowded interior.

A foreigner thinks of Shanghai and Peking as being purely urban areas. In reality, of the almost 11 million Chinese who live in the "independent municipality" of Shanghai five million inhabit the 10 rural counties that surround the city proper. The rural five million are the population of some 200 "communes": self-governing political and economic units that are each divided into between 10 and 30 "production brigades." The production brigades are subdivided into "production teams," each several hundred strong. In all, the rural population of Shanghai incorporates some 2,700 production brigades and nearly 28,000 production teams.

One example of rural health care that we saw in 1971 was provided by the Ma Chiao commune outside Shanghai proper. We visited one brigade of the commune, the Sing Sing brigade, with a total population of 1,850 subdivided into 12 production teams. The brigade health station, where we interviewed the staff, was served by four "barefoot doctors" and a midwife. Each production team in the brigade, we learned, had one to three additional health aides.

It is worth mentioning here that the term barefoot doctor is a literal translation of the Chinese appellation (*chijiao yisheng*) that has been given these health workers. We never met a barefoot doctor who was barefoot. Moreover, these workers are not addressed as "doctor" (*yisheng* or, more honorifically, *dafu*) by their patients but are called "comrade" (*tongzhi*). The barefoot doctors, who generally receive three to six months' initial training, followed by continuing on-the-job education, evidently think of themselves not as "expert" health workers but as peasants who do some medical work.

The eldest of the four barefoot doctors at the brigade health station, Ho Shi-chang, was 30. Before his medical training his education had consisted of six years of primary school, completed at age 13. Between then and the start of his medical training in 1964 Ho worked as a farmer; in 1971 he still spent about half of his working time farming. The year before Mao's June 26 Directive was issued Ho had been one of 274 students who spent three months at the county hospital receiving basic training from the 13 health workers on the hospital staff. He later spent three additional months at the commune hospital for practical training; he still spends one day a week there as part of his continuing practical education.

RURAL HEALTH STATION outside Shanghai is furnished plainly; here four barefoot doctors and one midwife provide primary health care for the 1,850 members of the Sing Sing Brigade. The midwife, with her back to camera, attends two or three deliveries monthly.

Like his colleagues, Ho is responsible for treating the "light diseases" of his fellow brigade members: minor injuries, gastrointestinal illness, colds and bronchitis. He also administers immunization against diphtheria, tetanus, whooping cough, measles, smallpox, poliomyelitis, Japanese *B* encephalitis and meningococcal meningitis. Another of his public-health duties is to supervise the collection, treatment and storage of human excreta for utilization as fertilizer. The actual work may be overseen by the production-team health aides, but the responsibility is his, as are continuing campaigns against pests such as flies, fleas, cockroaches and snails (the last is the intermediate host of the organism that causes schistosomiasis).

The contents of Ho's medical bag provide a good measure of his capacity to treat various diseases [*see illustration on next page*]. In the U.S. 39 of these medications are in use; only nine of them can be obtained without a physician's prescription. When Ho encounters a problem beyond his scope, he refers the patient to the commune hospital. (The 200 or so commune hospitals in rural Shanghai have an average of 30 beds each.) If the matter is more serious, the patient is referred to the county hospital.

Ho and his colleagues divide their time about equally between farmwork and duty at the brigade health station. As members of the commune they share in the produce the commune raises. They also share in the commune's cash income, which is divided among the commune's members on the basis of the "work points" each earns. They receive no extra income for their work as barefoot doctors, but their health-station service earns them work points, so that they lose no income either.

At the time of our visit in 1971 Ho was one of 7,700 barefoot doctors working in the communes of rural Shanghai. This represented a new high in the level of health care. In the late 1950's, as part of the initial effort to train health auxiliaries, health teams from the city proper had trained some 3,900 rural residents in health care. During the retrenchment of 1961–1965, however, training was largely discontinued and many rural health aides returned to full-time farmwork. The program that has produced most of the barefoot doctors

in China today began only after Mao's June 26 Directive of 1965. By 1968 in the Shanghai area some 4,500 barefoot doctors had been trained, and the new trainees had themselves trained an additional 29,000 health aides to provide part-time health care at the production-team level.

Like Shanghai, Peking has a substantial rural population. The city's nine urban districts have four million residents, but three million more live in the nine counties outside the city proper. During our 1971 visit we found that the delivery of medical care in this rural area differed very little in pattern from the system in rural Shanghai. The commune we visited, the Shuang Chiao commune, had a population of 38,000, subdivided into six production brigades and 77 production teams. The commune hospital had a clinical laboratory and X-ray facilities but had no beds and served outpatients only; any patient requiring hospitalization was referred to the county hospital. At this commune we visited the aid station of one production team, the Shuang Pei team, which was staffed by a single barefoot doctor. Liu Yu-cheng, age 23, had completed the lower level of secondary school. When Mao's June 26 Directive was issued, the other members of the Shuang Pei team nominated Liu for medical training. His basic training had come from a mobile team of physicians who normally worked in Peking proper but were then assigned to the Shuang Chiao commune hospital; the instruction period lasted for three months. Since that time Liu has been given short leaves of absence for further study. Not long before our visit he had spent three months in the city studying traditional Chinese medicine.

During our visit in 1972 we saw one of the commune hospitals in the rural area outside Peking. The hospital had opened in May, 1971. It has 30 beds, and its staff of 48 serves a commune with a total population of 46,000. Seven of the staff members are physicians; five are Western-trained and two are traditional practitioners. All seven are on loan from the staff of a hospital in Peking proper. The balance of the staff consists of 15 nurses, 25 health auxiliaries, 11 technicians and administrators and one cook. In the 17 months between opening day and our visit in October, 1972, some 8,000 members of the commune had received care as outpatients (an average of 500 per month) and 500 more had entered the hospital for treatment (an average of one patient per bed per month).

Medical care in urban China follows the same pattern of decentralization that we observed in the rural areas. In Peking proper, for example, each of the nine urban districts has an average population of about 400,000. The city's municipal medical services include four specialized research hospitals and 23 general hospitals; 10 of the general hospitals have more than 500 beds.

Each of the nine districts is subdivided into "neighborhoods." In the West District of Peking, which we visited, there were nine such neighborhoods; the one that was our host, the Fengsheng Neighborhood, has a population of 53,000. One municipal hospital in the West District, the People's Hospital, is where any Fengsheng Neighborhood patients are sent if care at one of the specialized hospitals is not required. Within the neighborhood itself the only hospital-like facility is exclusively for outpatient care. Its staff of 90, however, provides the nucleus of the neighborhood health-care apparatus. The staff includes seven Western-trained physicians and 20 practitioners of traditional medicine, 31 nurses and technicians, 18 administrative and other personnel and 14 trainees. The public-health department is responsible for supervising the urban equivalent of the rural production-team aid stations: a total of 25 health stations operated by "Lane Committees."

The Fengsheng Neighborhood has 132 lanes in all, so that each of the 25 Lane Committees represents the residents of five or six lanes, or some 400 families. Each Lane Committee health station is staffed morning and afternoon by local housewives who have the title Red Medical Worker. The daily hours are from 8:00 to 11:00 A.M. and from 1:00 to 5:30 P.M.

The health station we visited was maintained by the Wu Ting Lane Committee. Its plain single room was fur-

MEDICATIONS
- *Adona (cardiac stimulant) ampules
- Adrenalin ampules
- Aminophyllin tablets and ampules
- Ammonium chloride tablets and solution
- *Analgin tablets and ampules
- Aspirin-phenacetin-caffeine tablets
- Atropine tablets
- Belladonna extract tablets
- *Berberine tablets
- Brown's mixture tablets and liquid
- Caffeine sodium benzoate ampules
- Chloromycetin ampules and capsules
- Chlorpheniramine maleate tablets
- Chlorpromazine tablets and ampules
- *Chlothamine tablets
- *DCT tablets
- Demerol
- *DPP in tablets
- Ephedrine
- *Furazolidone tablets
- Lactobacillus tablets
- *Lobodura tablets
- Luminal tablets
- Nikethamide ampules
- Nitrofurantoin tablets
- Penicillin, crystalline
- Penicillin, procaine
- Phenolax tablets
- *8-p-phenylbenzylatropinium bromide tablets
- Phenylbutazone tablets
- Piperazine citrate tablets
- Promethazine tablets
- Probanthine tablets
- Reserpine tablets
- Sodium bicarbonate tablets
- Sulfadiazine tablets and ampules
- Sulfaguanidine tablets
- Sulfamethazine tablets
- Sulfamethoxpyridazine tablets
- Sulfathiazole tablets
- *Syntomycin capsules
- Terramycin tablets
- Tetracycline tablets
- Valium tablets
- Vitamin B1 tablets
- Vitamin B2 tablets
- Vitamin C tablets
- Vitamin K tablets
- *Vitamin U tablets
- Yeast tablets

TOPICAL AGENTS
- Alcohol
- Boric acid ointment
- Eye drops
- Gentian violet
- Iodine tincture
- Mercurochrome
- Nose drops
- Sulfa ointment

EQUIPMENT
- Acupuncture needles
- Adhesive tape
- Bandages and gauze
- Cotton sponges and swabs
- Drinking cups
- Forceps
- Fountain pen
- Hypodermic needles
- Notebook for records
- Paper bag
- Rubber tubing
- Scissors
- Sphygmomanometer
- Syringes (2 cc. and 5 cc.)
- Thermometers (oral and rectal)

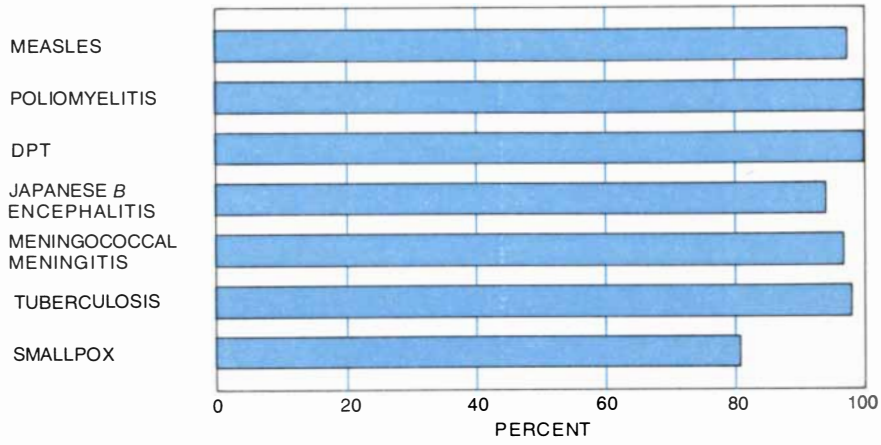
MEDICAL SUPPLIES at the disposal of barefoot doctors include 50 medications and eight topical agents. The names are the most familiar ones; some are proprietary and some generic. Of the 39 medications used in U.S. practice, 30 (*color*) are available only by prescription. The items marked with an asterisk are not generally available in the U.S.

nished with an examination couch, a cabinet for medical supplies, a table and chairs. Three Red Medical Workers were present; we talked at some length to one of the women, Yang Hsio-hua. Yang was 38 years old and had worked at the health station for two years. Married some 20 years ago, she has three children, 19, 15 and 11. She had become a Red Medical Worker by volunteering for a month's basic training at the Fengsheng Neighborhood medical facility. During that time she learned how to take a medical history and how to conduct simple physical examinations, including such routines as measuring blood pressure. She was instructed in the uses of a number of drugs, both Western and traditional, and she had learned the techniques of acupuncture and intramuscular and subcutaneous injection.

A physician from the neighborhood medical facility visited the Wu Ting health station as often as three times a week. Yang and her fellow medical workers visited the neighborhood facility when they had questions. She observed that seven to 10 patients visited the health station during morning hours and that four or five more might come in the afternoon.

At the health station much of the emphasis is on preventive medicine, in particular immunization against infectious diseases. Most immunization of local children is accomplished at the station. If necessary, one of the medical workers will call for a child at home or even administer an inoculation there. For example, one Lane Committee health-station immunization chart we saw showed that of 160-odd children eligible for immunization against measles in 1971, a total of 154 had received inoculations by the time of our visit. Other charts showed the comparative incidence of infectious diseases from 1958 on. Measles has evidently become uncommon since immunization first began in 1965 [*see illustration at right*].

Neighborhood factory workers seldom use the Lane Committee health stations because their own factory medical facility is more convenient. Their dependents, however, do use the Lane Committee stations, and the factory will reimburse half the cost of treatment. The charge is never more than 10 *fen* and is usually smaller. (One hundred *fen* make one *yüan* and one *yüan* is equal to about 40 U.S. cents.) Nonetheless, the income from the fees, together with a subsidy that each Lane Committee health station receives from its Neighborhood Committee, allows the Red Medical Workers to receive a monthly stipend of about 15 *yüan*, roughly a third of the wage that a beginning factory worker would be paid.

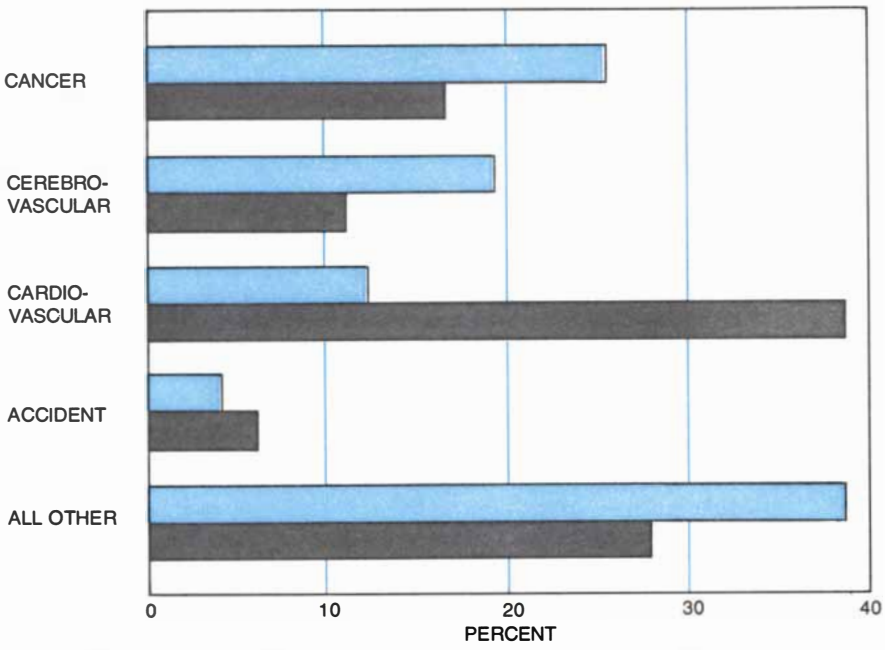


IMMUNIZATION OF CHILDREN against nine infectious diseases is routine. Bars show the percentage of eligible children immunized in 1971 at one Lane Committee health station.

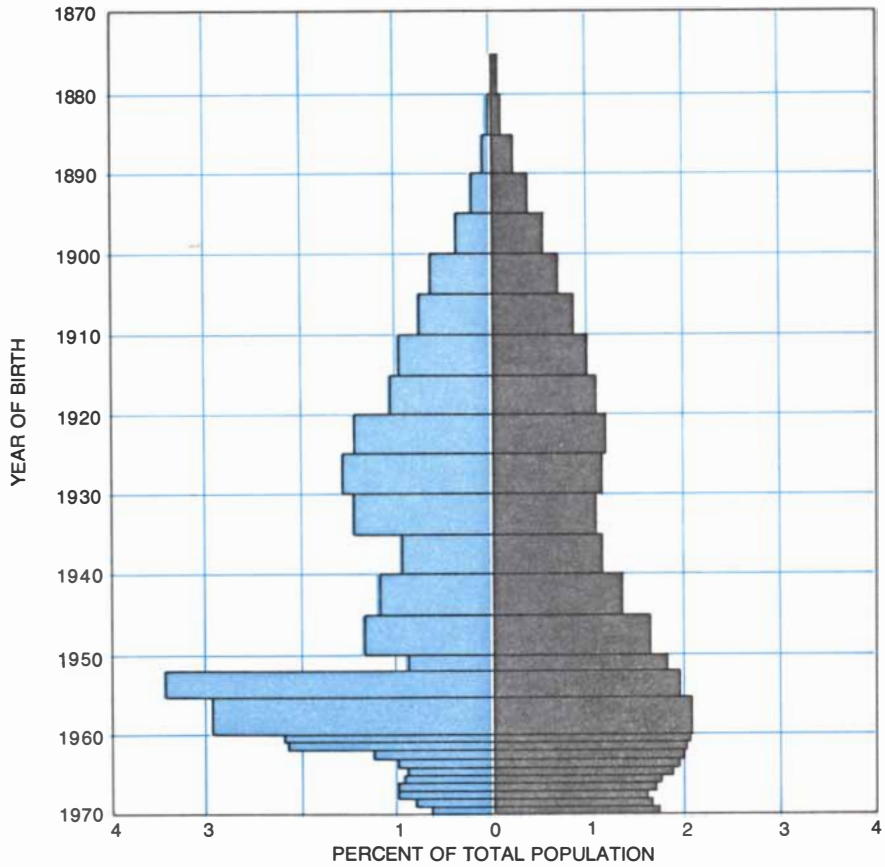
China has by no means solved its medical problems. For example, both tuberculosis and trachoma are far more prevalent than they are in richer nations. What we have seen of the delivery of medical care in both rural and urban areas, however, convinces us that public health and medical care in China are better than they are in other nations handicapped to a similar degree by technological underdevelopment. One striking instance of this is the success of the Chinese campaign for birth control.

Health workers both in the countryside and in the cities have as one of their main responsibilities a program to make contraception popular. The urban Red Medical Workers make a point of explaining to their patients how a lower birth rate will benefit not only their neighborhood and their city but all China. At the same time the medical workers stress the part that birth control plays in the "liberation" of women, a concept that is substantially more radical in tradition-bound China than in Western countries. An example of the effec-

INCIDENCE OF MEASLES in one Peking neighborhood has fallen sharply since 1965, the year routine inoculation of children against measles was begun. All immunization is free.



LEADING CAUSES OF DEATH in Shanghai during the first six months of 1972 (*color*) are contrasted with those in the U.S. as a whole during 1968. In both instances cancer, stroke and heart disease were responsible for more than 55 percent of all deaths; the parallel indicates that the Chinese are living longer and more of them are dying from the diseases of old age.



POPULATION PROFILE of the 478,000 inhabitants of the Luwan district of Shanghai in 1971 (*left*) is compared with that of the U.S. as a whole in 1970 (*right*). The "rectangular" U.S. profile is characteristic of the age distribution found in more developed countries. The "bulge" in the number of Chinese births between 1952 and 1960 may partly reflect the absence of official support for birth control during this period. The "pinch" in Chinese births in the 1963–1970 period, with an eight-year total that is scarcely half the U.S. total in the same period, probably reflects a greater use of birth-control methods by the urban Chinese.

tiveness of the urban effort is provided by the records of one Lane Committee health station in the city of Hangchow. The data represent only a small sample: 369 married couples with wives of childbearing age. Within this group 24 percent of the women and 3 percent of the men had been permanently sterilized. Another 27 percent of the women and 19 percent of the men regularly used contraceptives. Of the 98 wives who reportedly did not use contraceptives, 10 were pregnant, seven were newly married and 16 were still nursing (a period when, it is mistakenly believed by some, a woman cannot conceive). The crude annual birth rate for the population served by the Lane Committee health station is at the remarkably low level of eight per 1,000.

Birth statistics for Shanghai in 1972 are equally remarkable. The reported rate for the city proper is 6.4 per 1,000 and for the entire independent municipality 10.6 per 1,000. By way of comparison, in 1972 the lowest rate in any of the 50 states in the U.S. was 12.6 per 1,000 (in Maryland). In this connection the 1972 crude death rate for Shanghai proper is reported as being 5.6 per 1,000. Assuming that both figures are correct, this means that the city's natural growth rate is less than a tenth of 1 percent (0.8 per 1,000). This is one of the lowest natural growth rates in any urban area in the world.

Birth-control statistics from rural areas are substantially different. For example, one commune in the rural counties outside Peking, with a total population of 46,000, has compiled contraception statistics for 5,777 married couples where the wife is of childbearing age. Only 8 percent of the wives and 2 percent of the husbands have been permanently sterilized. Another 41 percent of the wives use contraceptives, the "pill" being favored over intrauterine devices by 23 percent to 18 percent. Among the husbands 9 percent use the condom, bringing the total of contraceptive users to 50 percent. During 1971 there were 1,181 births in the commune, so that the crude birth rate was slightly less than 24 per 1,000. The 1971 statistics for the 450,000 inhabitants of Shunyi County, a rural district outside Peking, are quite similar. There the married couples where the wife is of childbearing age number 49,297, and 59 percent of them practice contraception. The number of births in the county in 1971 was 9,504, which means a crude birth rate slightly above 21 per 1,000.

The birth rates in both of these rural samples seem high compared with the rates in the urban areas we have cited. Chinese health officials hope to see the national rate eventually fall to about 15 per 1,000. Nonetheless, even the rural rate is substantially below the former national rate, which is estimated to have been 45 per 1,000 or higher. (For purposes of comparison the current crude birth rate for Southeast Asia is estimated to be 43 per 1,000.)

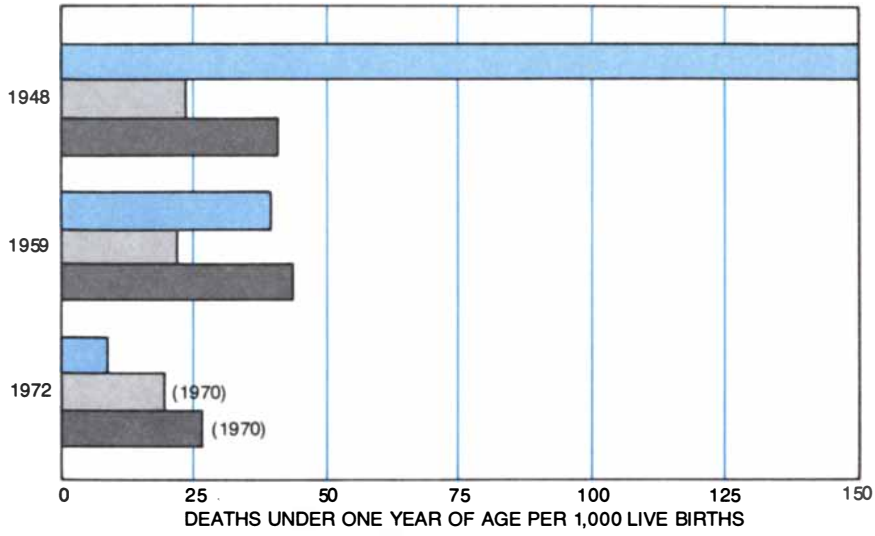
One still unresolved medical issue is how to achieve an effective union of Western and traditional medical practices. It will probably never be known exactly how many practitioners of traditional medicine there were in 1949, but they were estimated to number in the hundreds of thousands, and they provided at least some degree of health care to a large and faithful clientele, particularly in the rural areas. Nonetheless, then as now certain difficulties stood in the way of integrating the traditional and the Western-style practitioners.

Traditional Chinese medicine is much more than a collection of empirical remedies. It is based on a large body of theory, accumulated over some 3,000 years, that includes, for example, the concept of a "natural" balance between yin and yang. One declared objective of the People's Republic of China, however, is to lead the population away from the "superstitions of feudalism" and toward the practice of "scientific methods."

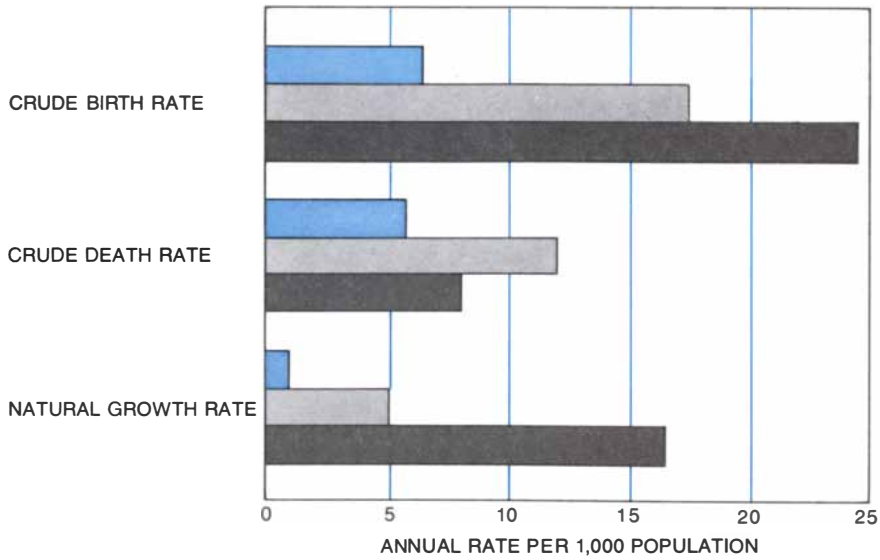
With eminent practicality Mao dismissed this theoretical conflict early in the 1950's by declaring that traditional Chinese medicine was a "great treasurehouse" and urging that traditional and Western practices be merged. Some progress in this direction was made in the 1950's and early 1960's, but it is only since the Cultural Revolution that emphasis on unifying the two streams has increased significantly. For example, students in the Western-style medical schools now receive more than casual instruction in traditional medicine, and those who study traditional medicine are also taught Western practices. At the same time the years since 1949 have seen the general adoption of certain traditional techniques: the use of herbal preparations, of gymnastic and respiratory exercises, and of two related treatments, moxibustion and acupuncture.

Traditional Chinese pharmacology emphasizes herbal remedies, usually in the form of a broth or tea that the patient drinks. In the countryside today health workers not only gather and prepare wild herbs but also cultivate a number of them. Moreover, the medicine cabinets in rural and urban health stations are stocked with herbal remedies as well as with Western ones. Some herbal remedies are even available as sterile preparations for injection. Medical investigators in China suspect that, just as one traditional herbal remedy, *ma huang*, has been found to contain ephedrine as its active principle, others may prove to contain similar specific pharmacological compounds. A substantial part of current pharmacological research is concerned with examining this possibility.

One tenet of Chinese traditional medicine is that the internal organs of the body are connected to points on the skin by way of channels called *ching lo* that run throughout the body close to the skin. Stimulation of a point along a channel is supposed to affect the internal organ attached to that channel. Although the existence of such channels has never been demonstrated anatomically, it is on this theoretical foundation that moxibustion and acupuncture rest. In moxibustion the stimulus to the skin is produced



INFANT DEATHS in the first year in urban Shanghai (*color*) are compared with the deaths of white (*light gray*) and nonwhite (*dark gray*) infants in New York City in 1948, in 1959 and in 1972. (The comparable New York figures are for 1970, the latest year for which they are available.) The data for Shanghai are from the Chinese Medical Association; the 1972 rate, less than 1 percent of all live births, is so low that it has met with some skepticism.



NATURAL GROWTH RATE, the excess of births over deaths, is shown for urban Shanghai (*color*) and for the white (*light gray*) and nonwhite (*dark gray*) population of New York City. The New York data are for 1970; Shanghai data for 1972 suggest a notably low rate.

by heat. The pulverized leaves of a herb, the mugwort, are wrapped in a paper cone, the tip of the cone is ignited and the smoldering herb is placed on or near the appropriate point.

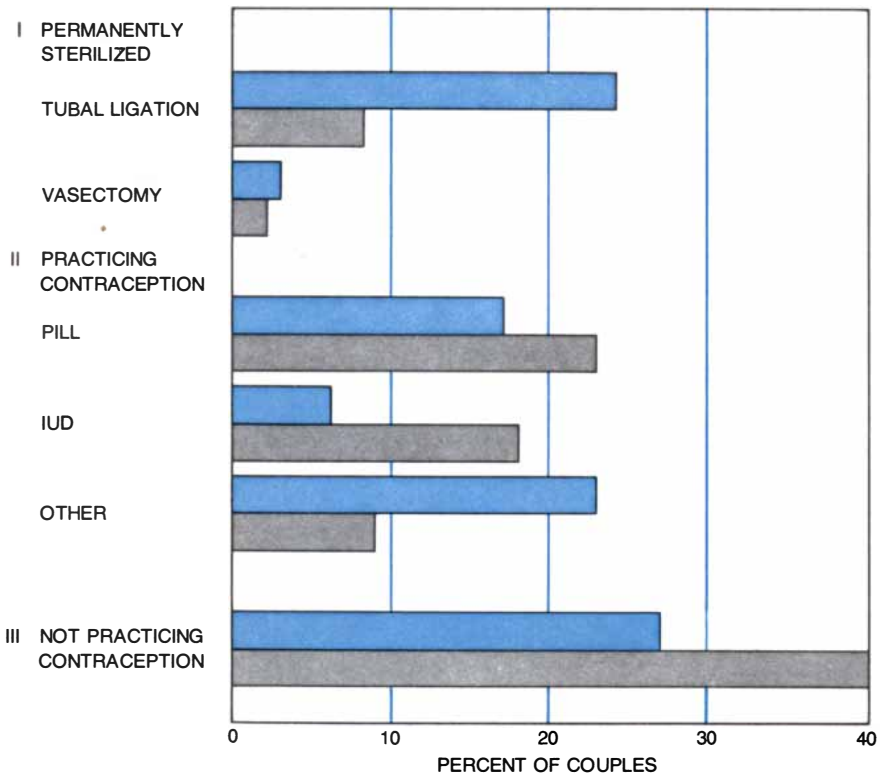
In acupuncture the stimulus is applied by inserting a needle to a predetermined depth in the patient's skin. Acupuncture can be used both for diagnosis and for treatment. It is even taught in secondary schools in much the same way that first aid is taught in some U.S. school systems. Indeed, health workers whom we met in both rural and urban areas said that as a routine treatment for headache they preferred acupuncture to aspirin.

The Chinese use of acupuncture to produce insensitivity to pain during surgery has attracted much attention in the U.S. The procedure is actually less common in China than widely circulated American accounts might lead one to believe. It is usually called acupuncture anesthesia in English (and in English-language publications of Chinese origin). "Anesthesia," of course, implies a general loss of sensation by the patient. It would be more appropriate to use the term analgesia, a loss of pain sensation. In any event the Chinese say that the effect is most successful when the surgical procedure is "above the waist," so that acupuncture for this purpose is largely confined to surgery of the head, the neck and the thorax. Because the technique is still considered experimental it is applied only with patients who specifically request it and, within this group, only with those who are considered to be "not too anxious" about surgery. Most surgical patients in China receive conventional anesthetics.

With respect to "mass movements" in Chinese health care, the first nationwide "Patriotic Health Campaigns" were launched in the 1950's. The targets were "four pests": flies, mosquitoes, rats and the "grain-thieving" sparrow. (The ecologically unjust charge against sparrows was later withdrawn and bedbugs were substituted.)

Similar mass movements continue today, expanded and redirected to include other public-health problems, such as the handling of human excreta, the purity of the water in local wells and streams, methods of food preparation and even the disposal of trash. In the countryside the departments of health in commune and brigade hospitals supervise the campaigns; in the cities district and neighborhood hospital personnel have the same responsibility. In both rural and urban areas the ultimate responsibility rests with the local health workers and the aides they have trained. For example, in Hangchow the Red Medical Workers at one Lane Committee health station set aside three days a month for "cleanup work" with the assistance of their fellow residents. The time is spent removing trash and inspecting potential pest breeding places. Full-scale sessions, however, usually coincide with major festival days and with the state celebrations on May 1 and October 1.

The mass-movement approach has apparently helped the Chinese to dispose of such "social" illnesses as drug addiction and venereal disease. For example, early in the 1950's checklists of the symptoms of syphilis were posted in every community center. A slogan ("We don't want to take syphilis into Communism") was promulgated, mass surveys were conducted and social pressure was brought to bear on suspect individuals who failed to seek medical attention. Elimination of prostitutes as carriers of venereal disease was accomplished by giving them the opportunity to engage in "socially constructive" work. As with the birth-control campaign, the Chinese effort to control "social" illnesses seems to have met with greater success than similar efforts in comparably poor nations and even in some technologically advanced ones.



CONTRACEPTIVE PRACTICES in an urban area (*color*) and in a rural area (*gray*) are compared; the urban sample is from the city of Hangchow, a provincial capital, and the rural sample is from a commune outside Peking. Sterilization is nearly three times commoner in the urban than in the rural sample and substantially fewer rural males use contraceptives. The bias seems to be reflected in the difference between urban and rural birth rates: below 10 per 1,000 in some urban areas and above 20 per 1,000 in some rural ones.

A summary of China's achievements in transforming the delivery of medical care since 1949 shows an interweaving of three main threads: decentralization, demystification and continuity with the past. Following a pattern that many students of community medicine would be happy to see more widely emulated in Western countries, the delivery of medical care in China begins at the lowest possible level in both the city and the countryside. Initial medical attention is in the hands of health aides who are part of the community they serve. From this initial point of contact a clearly organized system of referral leads level by level up to a plateau of sophisticated medical specialization. (It is worth noting that in some technical areas, such as the treatment of severe burns and the replantation of limbs, Chinese medicine may be ahead of Western medicine.) The patient with a problem that cannot be handled at one level of this decentralized structure moves on to the level above. The system is an efficient, low-cost one. Moreover, it has the advantage of building social cohesion and local self-reliance

CALISTHENICS GROUPS exercise in a riverside park in the city of Shanghai. One group (*right*) is exercising in Western style and the other is following traditional Chinese routines. Exercise is accepted by traditional Chinese medicine as something more than simple body-building. Along with certain respiratory practices it is used as part of both physiological and psychological therapy.

by emphasizing neighborliness and service for others from the lowest level up.

From the Chinese point of view demystification runs parallel to decentralization. The front-line medical workers are men and women with little in the way of formal education. They work part time and receive their instruction in health care through brief programs that emphasize the practical. They urge participation on the part of the people they look after. For example, each individual is expected, as a patient, to look out for his own health and, as a citizen, to look out for the health of the community. Under these circumstances it is no wonder that much of the mystery medicine so often holds for the layman has been effectively dispelled. Demystification has also been furthered by shortening the term of formal medical education and by assigning urban physicians to periodic tours of duty in the countryside. The two steps express with respect to medicine the determination of Mao and others to eliminate "elitism" in general.

As for continuity with the past, we have already noted the pragmatic adoption of certain traditional Chinese medical practices. What is less commonly appreciated is that the social structure involved in the decentralization process (for example the pyramidal succession that leads from courtyard to lane, from lane to neighborhood and from neighborhood to district) preserves, albeit with differences in method and purpose, much of the traditional Chinese social organization. Thus the past is interwoven with the present in many ways. Indeed, in assessing the applicability of Chinese medical methods abroad it is often difficult to judge which methods might be used successfully in other societies and which are so culture-linked as to be uniquely Chinese.

For ourselves, with a concern for the improvement of medical and social services in the U.S., the lessons from China may be more general than specific. Certainly they transcend medical technology and enter the spheres of politics and economics. It seems to us that the Chinese have managed to overcome severe problems, to improve their system of medical care and to enhance the health of their population only by making medical change an integral part of a change in Chinese society as a whole. We in the U.S. face social and medical problems that, although they are quite different from China's in many ways, are equally difficult. It remains to be seen whether or not we shall meet them in as determined and comprehensive a way as the Chinese have met theirs.

INTEGRATED OPTICS

Tiny lasers, lenses, prisms and switches made of thin films can manipulate light to serve the same purposes as the manipulation of electrons in the thin-film devices of integrated electronics

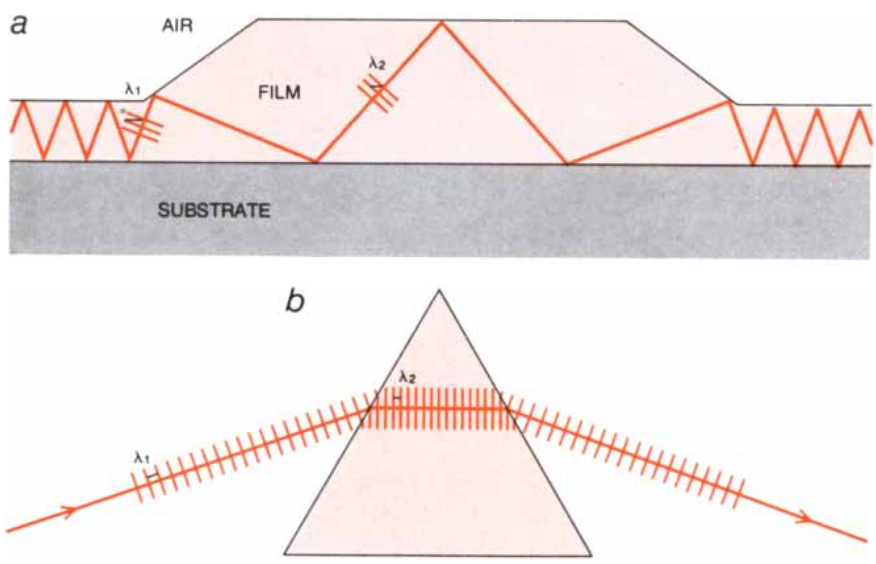
by P. K. Tien

The invention of the transistor in 1947 started a profound evolutionary trend in the fields of communication, information processing and automatic control. In time hundreds of transistors and other solid-state devices could be precisely laid down in integrated circuits only a few millimeters across. Such circuits were not only compact; they were also fast and eventually they were quite cheap. Today a new kind of integrated circuit is under active investigation: circuits based on optical devices. Integrated optical circuits can be laid down in thin films in much the same way that integrated electronic circuits are. The thin films, however, form miniature lasers, lenses, prisms, light switches and light modulators. Since the frequency of light is some 10,000 times higher than the highest frequency of an electronic device, the amount of information that can be carried by a light signal is correspondingly greater. Moreover, optical circuits are in principle considerably faster than electronic circuits.

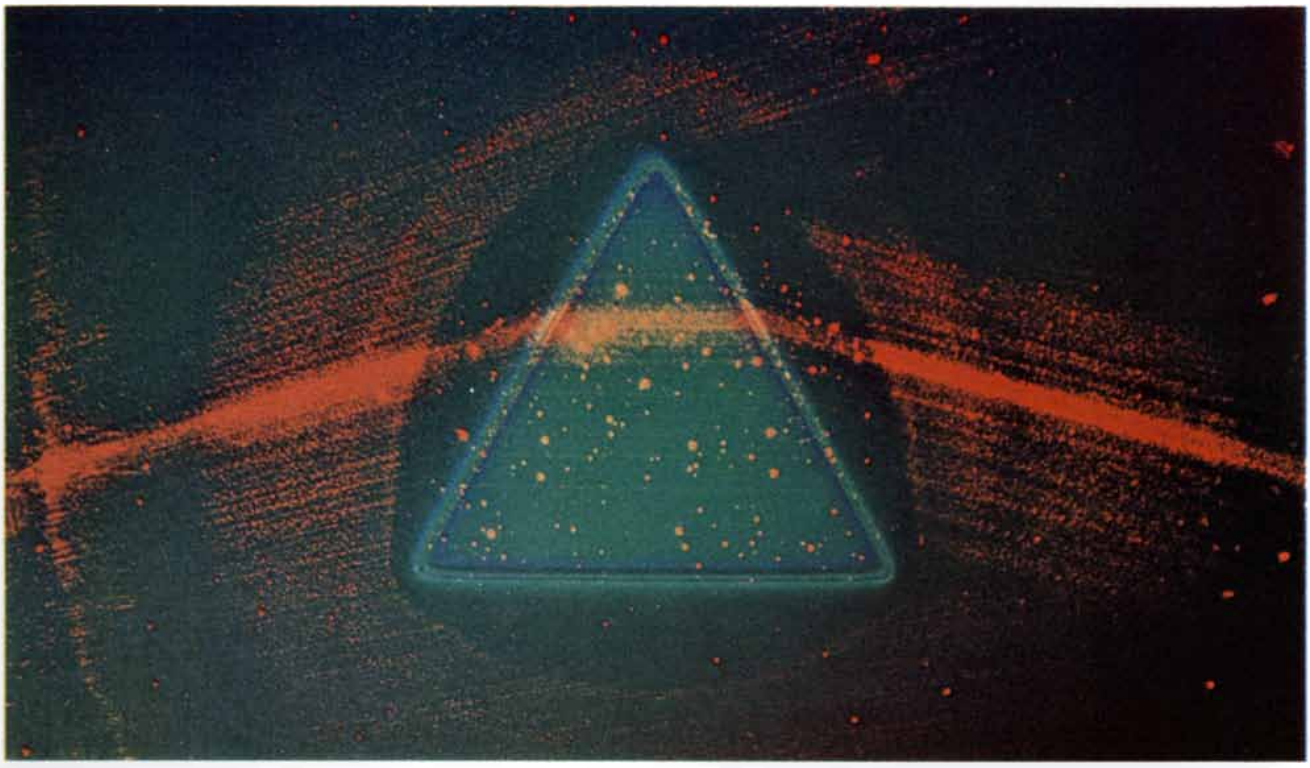
Such possibilities were apparent with the emergence of the laser in the early 1960's. The coherent light of the laser, like the coherent oscillations of an electronic circuit, could in principle be modulated or otherwise manipulated for practical purposes. To manipulate laser light in this way, however, calls for optical components such as prisms and lenses having accurately polished surfaces. Such components tend to be bulky and expensive. Furthermore, in order to provide a stable environment for experiments with lasers an optical laboratory needs a large optical bench that is typically 20 feet long. Such a bench is made of a solid block of aluminum supported by a rigid wood frame. The entire structure may be floated on eight airplane tires to isolate it from the vibrations of the building. On the bench one can put a number of gas lasers, each up to a few meters in length, which serve as sources of radiation in various regions of the electromagnetic spectrum from the ultraviolet to the far infrared. Radiation from the laser in operation passes through a series of polarizers, apertures, prisms and lenses, finally reaching an area where the experiment is being conducted.

All these facilities are quite necessary for a good optical experiment. It is clear, however, that they are much too massive and elaborate for any system that is to be used daily for any practical purpose. Thus the question is: Is it possible to compress the entire optical system, including all the necessary auxiliary equipment, into a space the size of a small button? To do so light would have to be generated by a tiny laser, modulated by a tiny modulator and directed by tiny lenses and prisms.

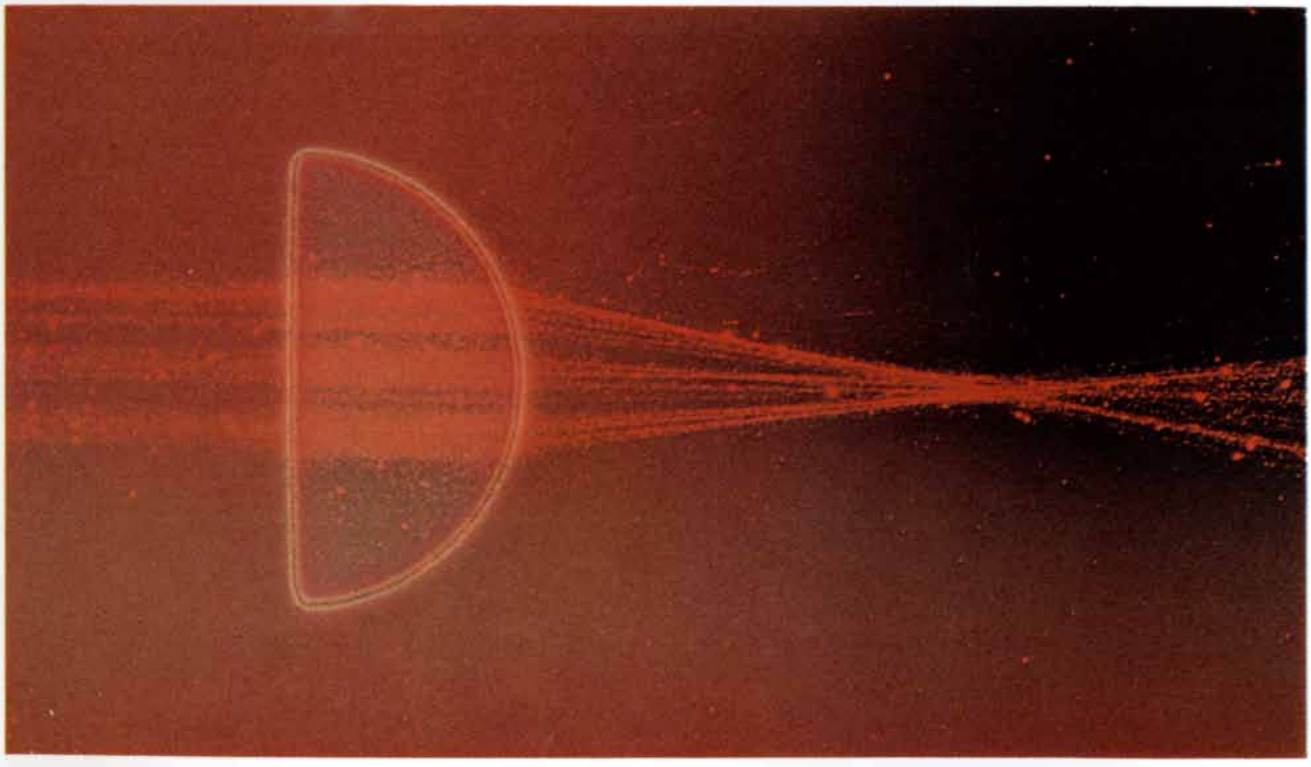
Such miniature optical devices can be made of thin films, making it possible for some of the technology already developed for integrated electronic circuits to be applied in their fabrication. Building even a simple integrated optical circuit, however, calls for techniques even more refined than the techniques employed for integrated electronic circuits. Thin



CROSS SECTION OF A THIN-FILM OPTICAL DEVICE (*a*) such as a prism or lens (*see illustrations on opposite page*) shows that it is basically a thicker section of the film. Laser light is a series of parallel wave fronts, that is, crests and troughs that propagate in a zigzag path within the film. The velocity of the light (v) traveling in the film is equal to the frequency of the light (f) times its waveguide wavelength (λ), where the waveguide wavelength is the distance parallel to the film between two crests ($v = f \times \lambda$). The waveguide wavelength in the thicker section of the film (λ_2) is shorter than the waveguide wavelength in thinner section (λ_1). Therefore light from the laser propagates more slowly in a thicker film than it does in a thinner one. When the light enters a medium where its velocity changes, it is refracted or bent as is shown in the plan view of a thin-film prism (*b*).



THIN-FILM PRISM refracts a beam of laser light just as an ordinary prism does. The prism, enlarged 30 diameters, is only about one millimeter in height and several microns (thousandths of a millimeter) thick. It was made by depositing a uniform film of zinc sulfide one micron thick on a substrate of glass and then adding a second layer of zinc sulfide through a triangular mask. The edge of the prism is tapered over a distance of several microns; beginning and end of the taper give rise to double line outlining the triangular shape. Circular bright area surrounding prism is laser light reflected by microscope lens through which photograph was taken.



THIN-FILM LENS brings three laser beams to a focus just as an ordinary lens does. The lens, also enlarged 30 diameters, was fabricated in the same way as the thin-film prism. The picture was made by photographing one laser beam three times, once at each of the three positions. Bright circle at bottom left is light scattered by a flaw in the film. Thin-film prisms and lenses are two of a number of components that will direct light through integrated optical circuits that will eventually be useful in communication, information processing and control. Both photographs on this page were made in author's laboratory at Bell Laboratories in Holmdel, N.J.

single-crystal films of various materials must be formed into the lasers and modulators. Since optical wavelengths are of the order of one micron (a thousandth of a millimeter), the techniques must be advanced enough for the fabrication of structures with a precision of even smaller dimensions. The new field of integrated optics thus requires the development of miniature optical devices that can be laid down in an integrated unit of optical circuitry. Optical circuits could then be fabricated in standard modules for service in practical systems. The prospect is excellent that reliable units of this kind can be produced economically in large quantities.

The basic concept of integrated optical circuits is not new. Because of the obvious difficulties involved in realizing such devices, however, active research in the field was not begun until 1968. Even then no more than a dozen people at a few institutions were working on the problem. (The institutions include Bell Laboratories, the University of Washington, the California Institute of Technology and the International Business Machines Corporation.) Since then the field has mushroomed. When the first professional meeting on integrated optics was held in Las Vegas, Nev., in January, 1972, it was attended by more than 150 investigators from 11 countries.

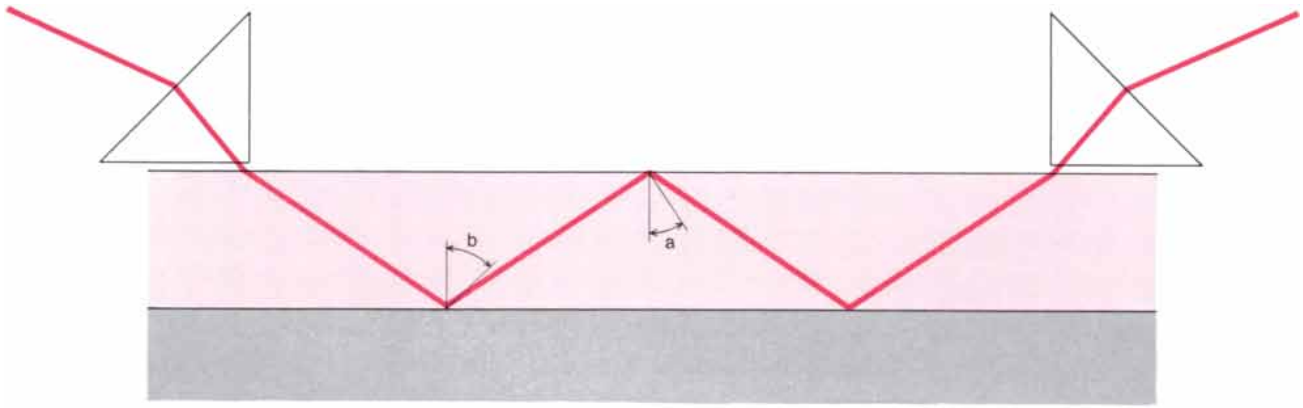
This surge of interest was the result of four favorable circumstances. First, glass fibers that can conduct light with very small losses were developed by the Corning Glass Works. Today the best optical fiber loses only two to four decibels (roughly half the input power) per kilometer, a loss comparable to that in a waveguide for millimeter-wavelength radio frequencies. Second, diode lasers made of gallium arsenide emerged, which meant that small, long-lived, low-cost, solid-state light sources might soon be available. These two developments were particularly promising for optical communication systems, which call for a means of generating light signals and transmitting them over long distances.

The third circumstance was that the art of masking and etching surfaces into complex patterns by electron beam has advanced so rapidly it is now feasible to form structures less than a micron across. The fourth circumstance is the seemingly insatiable demand of modern science and technology for miniaturized elements and compact systems.

The progress in the field of integrated optics has been rapid. When the work was started in 1968, the first efforts were directed toward answering such questions as: What materials should be used for thin films? How can such films be fabricated? How does a light wave propagate through a thin film? How can the light wave be put into the film and then taken out again at will? This last question was a particularly difficult one for a time. Because most of the films used in the early experiments had a thickness equivalent to a fraction of an optical wavelength, it was quite difficult to focus a laser beam directly on the edge of the film; the edge of the film is rough and scatters light. It was soon discovered that a laser beam could be inserted into the film with a prism. First the laser beam is focused into a thin pencil typically 50 microns (about 100 optical wavelengths) in diameter. When the beam is inserted through the prism coupler into a film that is, say, one micron thick, the beam is "compressed" into a thin ribbon with a rectangular cross section one micron thick and 50 microns wide. When the beam is extracted from the film by a second prism coupler, it returns to its original 50-micron round cross section.

The invention of the prism coupler, followed by the grating coupler and the tapered-film coupler, was a major forward step for integrated optics. In the three years from 1969 to 1972 a long list of thin-film optical devices were invented and demonstrated. They include devices that passively direct the light (such as thin-film prisms, lenses, mirrors and polarizers) and devices that actively generate or switch the light (such as thin-film lasers and light switches). Since the emergence of these devices is so recent, it was only at the beginning of last year that investigators began to study ways of combining them into complete integrated optical circuits. This work is still very much in progress. When a complete working optical circuit that is stable over long periods of time has been built, the central problem will have been solved.

How does light propagate in a thin film? A one-micron film is so thin that it must be supported by a thicker substrate such as glass. In order for the film to serve as a waveguide for light, it must have a higher index of refraction than the surrounding mediums [*see illustration below*]. Light traveling through matter moves at a lower velocity than it does in a vacuum; the ratio of the speed of light in a vacuum and the speed of light in a given substance is the index of refraction of that substance. The index of refraction for air is close to 1.0; the index for water is 1.3; the index for ordi-



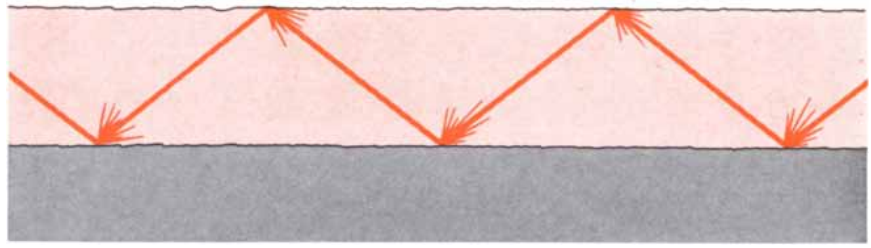
LIGHT WAVE IS COUPLED INTO THE FILM and taken out again by a prism coupler. The wave is alternately reflected by the upper and lower inside surfaces of the film because the substance of the film has a higher index of refraction than either the substrate or the air. The index of refraction of a substance is the ratio of the velocity of light in a vacuum to the velocity of light in the substance. There is a certain critical angle for the film-air interface (*angle a*) and a different critical angle for the film-substrate interface (*angle b*). When the angle at which a light wave approaches either interface exceeds the critical angle, wave is totally reflected.

nary glass is 1.5. All thin-film optical devices are simply thin-film waveguides made of substances with various refractive indexes and physical properties. An integrated optical circuit is formed when such devices are fabricated on a common substrate and interconnected.

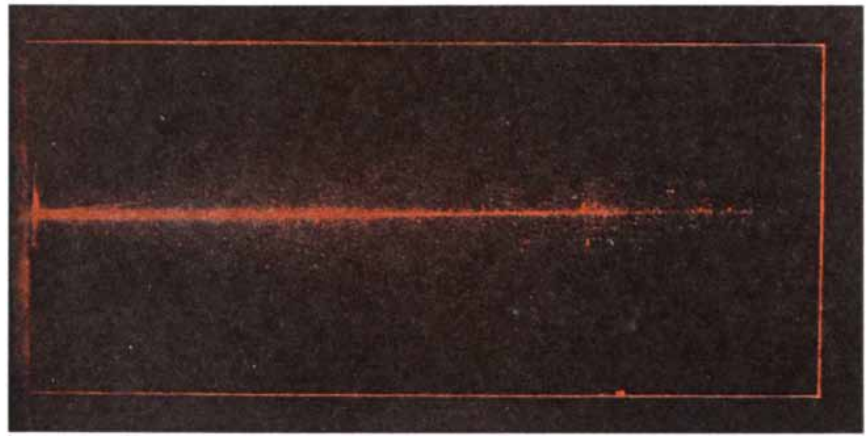
When light propagates within a thin film, it follows a zigzag path, alternately bouncing off the upper surface and the lower surface. Imagine a ray of light propagating toward the upper surface of the film. The light is totally reflected at the upper surface for two reasons. First, the index of refraction of the film is higher than that of the air. Second, the angle between the ray of light and a line perpendicular to the film-air interface is so large that it exceeds a certain critical angle for that substance, so that the upper surface of the film acts as though it were a very effective mirror. The phenomenon is known as total internal reflection. The light thus reflected propagates toward the lower surface of the film, where for the same reason it is again totally reflected. As a result the light wave follows a zigzag path.

The concept of the zigzag wave is important because it forms the basis for understanding how light behaves in thin-film devices. Moreover, the concept explains how light waves in a thin-film waveguide can carry a great deal of information. A particular zigzag wave can be identified by the angle between the two legs of the zigzag. Waves with different angles propagate in the same film independently of one another. Thus each zigzag wave can be regarded as a separate waveguide mode. The number of waveguide modes that can be carried by any single thin-film waveguide depends on the wavelength of the laser light, the thickness of the film, the refractive index of the film and the refractive index of the substrate. In a communication system each mode can carry one channel of information; a thin-film waveguide can therefore carry many channels simultaneously. In order to excite a particular mode the laser beam must enter the prism coupler from a specific direction. By carefully adjusting the direction of the incoming laser beam one can selectively excite the waveguide mode of one's choice.

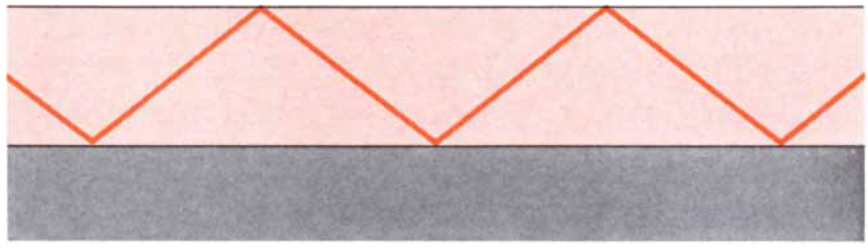
In a film one micron thick a light wave typically bounces 1,000 times in order to propagate over a distance of one centimeter. If the surfaces of the film are not smooth, some of the light will be scattered each time the wave hits the upper or lower surface. The wave will not be able to propagate very far before all the



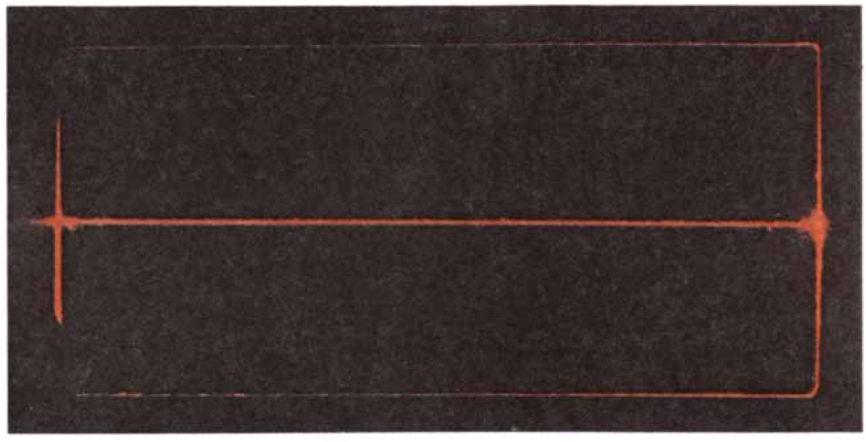
LIGHT IS SCATTERED each time the propagating light wave encounters a rough spot on the upper or lower surface of the film. Scattering accounts for major loss in a film waveguide.



POOR FILM loses a great deal of light from scattering. In this photograph the light wave traveled only six centimeters before being scattered out completely. Streak follows path of the light in an amorphous (noncrystalline) film of Ta_2O_5 deposited on a glass substrate.



LIGHT IS RETAINED only by a film that is very smooth. Ideally loss of light per reflection of zigzag wave must be less than one part in 10,000, a tenth the loss of mirrors in lasers.



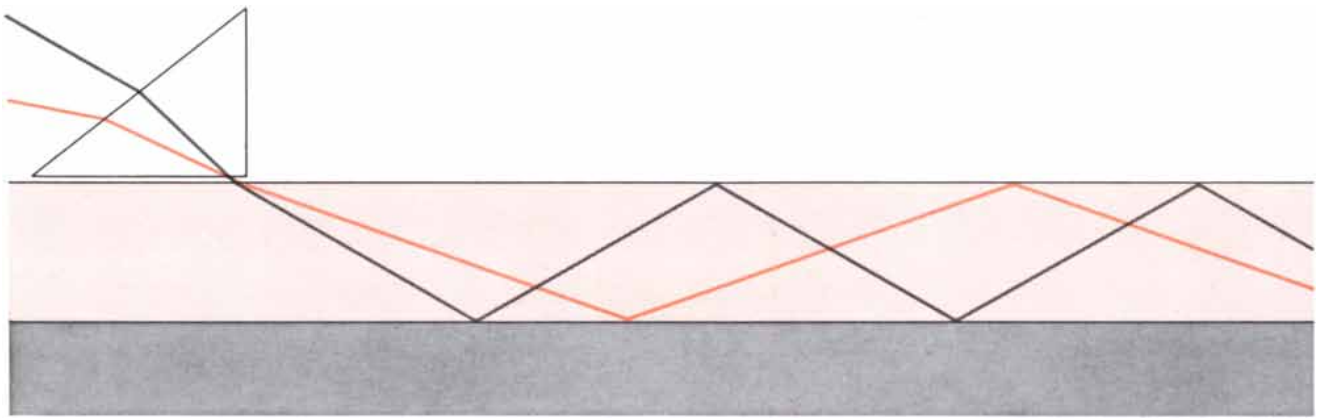
ALMOST PERFECT FILM shows the light beam as a thin bright line with almost no scattering. Film, also on a glass substrate, is made from the organic plastic vinyltrimethylsilane.

light is scattered out of the film. In order to limit the light loss of a film to less than one decibel per centimeter the loss for each reflection must be less than one part in 10,000. That loss is a tenth the light loss for the better mirrors used in lasers. How to make a smooth film free of pinholes and other imperfections is thus a major concern of integrated optics. In spite of the progress made in the past few years most of the loss in a thin-film waveguide is still due to the scattering of light.

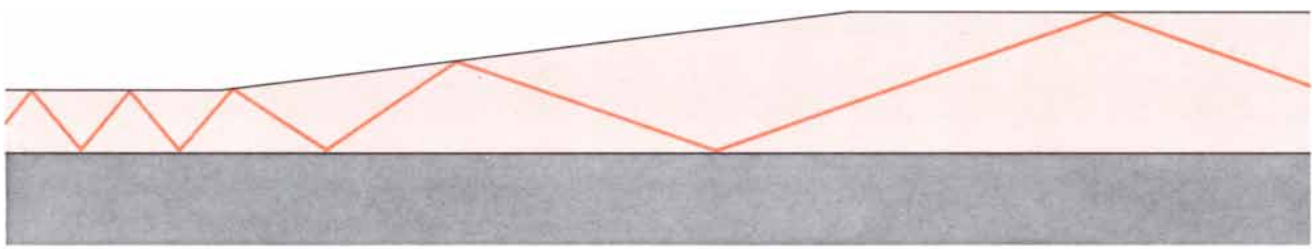
The velocity of the wave's forward direction of travel in a thin-film waveguide depends on the thickness of the film; the velocity is higher in a thinner film and lower in a thicker film. A thinner film can be joined to a thicker one by a tapered section. If the taper is gradual enough and covers a distance of many optical wavelengths, a light wave can propagate from the thinner film into the thicker through the taper without being scattered. This fact is quite important in the construction of thin-film prisms and lenses.

When light enters an ordinary glass prism from the air, it is slowed down within the prism. The result is that it is bent. A prism can be made in a thin-film waveguide by depositing a uniform film on the substrate and then adding another layer on top of the first through a mask with a triangular opening. Since the light wave propagates more slowly through the thicker prism than through the thinner film surrounding it, it is refracted as it enters or leaves the prism in precisely the same way it would be if it were going through an ordinary prism. As a result the triangular portion of the thin-film waveguide constitutes a thin-film prism. On the same basis one can construct a thin-film lens. The advantage of these prisms and lenses is that they are very small and simple to make; hence many of them can be formed simultaneously in a single step.

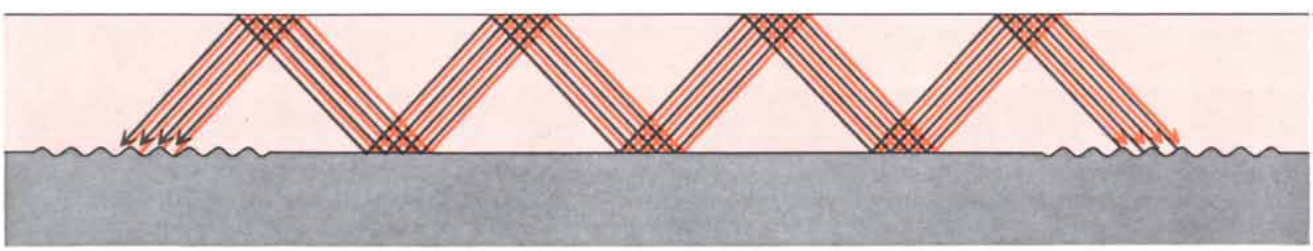
Light can be reflected not only from



MANY WAVEGUIDE MODES, that is, different zigzag waves, can simultaneously propagate within a single thin-film waveguide without interfering with one another. To excite a particular zigzag wave the incoming laser beam must enter the prism at a specific angle.



TAPERED SECTION of a thin-film waveguide allows the light to propagate from a thinner section of the film to a thicker section without being scattered in transit. The taper must be gradual and cover a distance of many optical wavelengths (50 to 100 microns).



THIN-FILM LASER can be made by placing a thin-film mirror at each end of a waveguide. Thin-film mirrors are gratings ruled on the substrate and covered with the film. A grating .1 millimeter wide with 500 grooves reflects virtually all the light. If the film is made of a laser material such as a plastic doped with a dye, light wave will be continuously amplified as it is reflected back and forth.

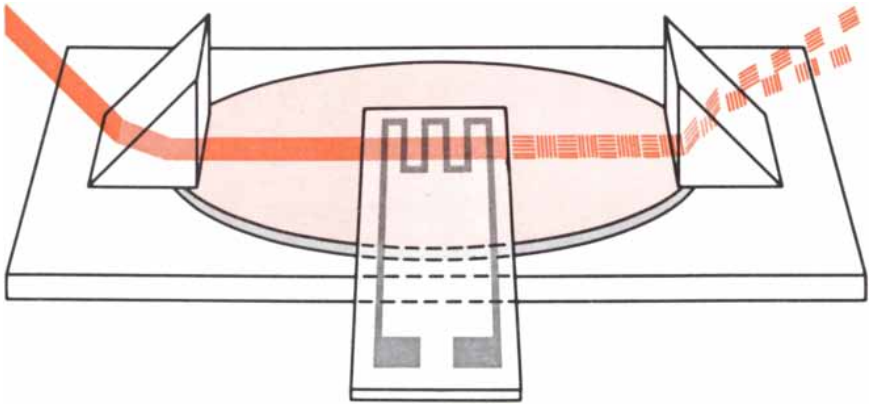
the top and bottom of the thin-film waveguide but also within it. For some time it was not known how to make a thin-film mirror, because any metallic structure put into a thin-film waveguide was found to absorb light rather than reflect it. An efficient mirror is formed, however, by a diffraction grating: a series of parallel grooves spaced precisely half a wavelength apart. When the technique of etching materials with a beam of ions recently became practical, one could fabricate miniature gratings in one's own laboratory.

The thin-film mirror is made by etching a grating onto the top surface of the substrate and then depositing a layer of film that covers both the grating and the substrate. A grating only .1 millimeter long consisting of some 500 grooves is quite big enough to reflect all the light that falls on it, so that a thin-film mirror is a very small device indeed.

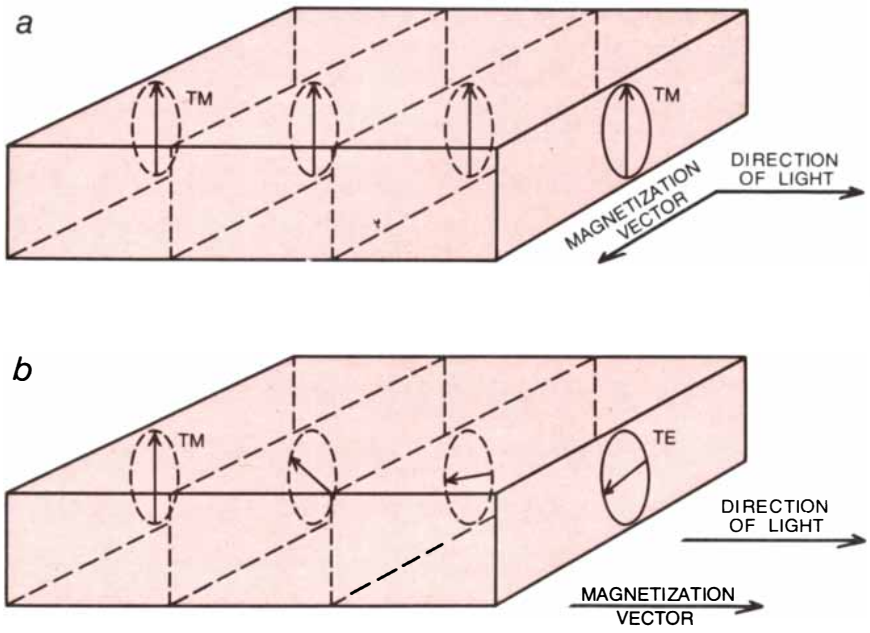
Now, consider two such thin-film mirrors, one at each end of a thin-film waveguide [*see bottom illustration on opposite page*]. A light wave propagating toward one of the mirrors will be reflected back toward the other mirror, where it will be reflected again by the first mirror. The light wave can in fact be reflected back and forth many times. The light wave is trapped in the area between the two gratings, which thus becomes a resonant optical cavity. If the film happens to be made of a laser material capable of amplifying the light, the wave will be continuously amplified as it travels back and forth. The cavity can therefore generate a sustained radiation as sufficient light energy builds up from the continuous amplification of the spontaneous emission in the laser material. In other words, the thin-film cavity is a laser.

Thin-film lasers have been made with a polyurethane film doped with the dye rhodamine 6G. The doped film emits a red-orange light when it is optically pumped by intense ultraviolet radiation. A thin-film dye laser of this kind cannot, however, serve in an integrated optical circuit; the dye bleaches too easily, and the operating lifetime of the laser is only a few hours.

The glass fibers used for transmitting laser radiation lose the least power at the infrared wavelengths from .80 to .86 micron and from 1.03 to 1.08 microns. Therefore a thin-film laser that generates infrared radiation rather than visible light would be best for integrated optical devices. A thin-film neodymium laser that generates radiation at a wavelength of 1.064 microns is one possibility. More-



MAGNETO-OPTICAL SWITCH is one thin-film modulator that can impose useful information, such as a voice signal for telephone communication, on a laser beam propagating through an integrated optical circuit. The switch consists of a magnetic iron-garnet film on a substrate of garnet (*circular disk*). In close contact with the film is an electric circuit in which the voice signals actuate a current. A beam of infrared radiation from a helium-neon laser is coupled into the switch from the left. Switch acts by polarizing the light in one direction when current in the circuit is on and polarizing it in a perpendicular direction when current is off (*see illustration below*). If output prism is made of a birefringent material, a material whose index of refraction is different for light of one polarization from what it is for the other, beams of different polarizations will emerge from prism at two angles.



CROSS SECTIONS OF TWO LIGHT WAVES illustrate in detail how the magneto-optical switch works. The electrons in the atoms of the film behave as though they were spinning on their axes, generating a small magnetic moment. When a small magnetic field is applied to the film through the circuit, the spins can be aligned in the direction of the field and the film becomes magnetized. The magnetization vector represents the direction and magnitude of the magnetic moment produced by the oriented spins. The vector can rotate in the plane of the film. The light wave entering the film can be polarized in one of two directions. If its electric field (*small arrows*) is parallel to the plane of the film, it is a TE (transverse electric) wave. If its electric field is perpendicular to the plane of the film, it is a TM (transverse magnetic) wave. In the absence of any external magnetic field (*a*), that is, with no current in the circuit, the magnetization vector is perpendicular to the direction in which the light wave propagates. If the light wave coupled into the film is initially a TM wave, it remains a TM wave as it travels through the film. When electric circuit is turned on, however, current generates a small magnetic field that causes magnetization vector to rotate so that it is parallel to the direction in which light wave propagates (*b*). A TM wave coupled into film would be gradually converted into a TE wave as it propagates. If it entered a polarizer instead of a prism, TM wave would be absorbed while TE wave would pass through.

over, a neodymium laser can be optically pumped by a light-emitting solid-state diode, so that a complete solid-state package could be fabricated.

Another possibility is an injection laser made up of a compound of aluminum, gallium and arsenic with the formula $Al_xGa_{1-x}As$; this laser radiates at a wavelength of .85 micron. Injection lasers have many advantages because they convert electrical energy directly into radiation, avoiding the need for the separate light source (such as a light-emitting diode) required by optically pumped lasers. One disadvantage, however, is that gallium arsenide (GaAs) has the very high refractive index of 3.6. This refractive index is not compatible with that of materials used to construct other thin-film devices, whose refractive indexes range from 1.5 to 2.0.

A second disadvantage is that a grating structure is difficult to fabricate in an injection laser. The present lasers have an optical cavity that uses two cleaved end surfaces as the reflectors. Incorporating such cleaved surfaces into an integrated optical circuit appears to be a difficult problem. To avoid the difficulty we are currently studying a thin-film corner reflector. Such a reflector is basically a thin-film right-angle prism, except that here a light wave enters the prism through its hypotenuse, is reflected internally from the other two sides and then passes out through the hypotenuse. It remains to be seen whether or not these reflectors can replace the cleaved-end-surface ones.

Since all thin-film lasers are still under development, the construction of a complete integrated optical circuit incorporating a laser may be a few years away. Meanwhile the realization of a modulator, another vital component in such a circuit, has progressed rapidly. In order for laser light to carry information over, say, a telephone line, it must be modulated by the voice signal through a thin-film switch. Only then can the light wave carrying the useful information be coupled into a glass fiber and transmitted many miles away.

Among the thin-film modulators available at present I shall discuss only one: the magneto-optical switch. This particular switch can turn the laser beam off and on at a rate of 3×10^8 (300 million) times per second. That rate is three times faster than the most advanced transistorized communication coaxial cable in use today. The switch consists of a film of a magnetic iron-garnet film on a substrate of garnet. Coupled into the film is the infrared output of a helium-neon laser, and in close contact with the film is a looped electric circuit.

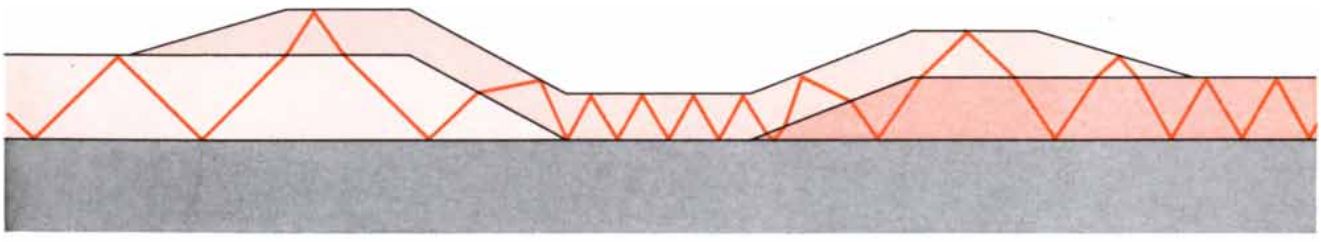
Within the magnetic film the electrons in the atoms behave as though they were spinning on their axes, generating a small magnetic moment. Usually electrons have a strong tendency to pair their opposite spins and cancel their magnetic moment. In this magnetic film, however, one has a large number of unpaired electron spins. If a small magnetic field is applied to the film, the unpaired electron spins are aligned in the direction of the field and the film is magnetized. A quantity called the magnetization vector can be used to represent the direction and magnitude of the magnetic moment produced by the oriented electron spins.

The magnetization vector can rotate rather freely in the plane of the film. In the absence of any external magnetic field the magnetization vector is in the plane of the film but is perpendicular to the direction in which the light wave is propagating [*see bottom illustration on preceding page*]. When a current is applied to the electric circuit in this kind of device, it gives rise to a magnetic field that rotates the magnetization vector into a direction parallel to the direction in which the light wave is propagating. Therefore, depending on whether the current in the circuit is on or off, the magnetization vector of the film alternates between being parallel to the light beam and being perpendicular to it.

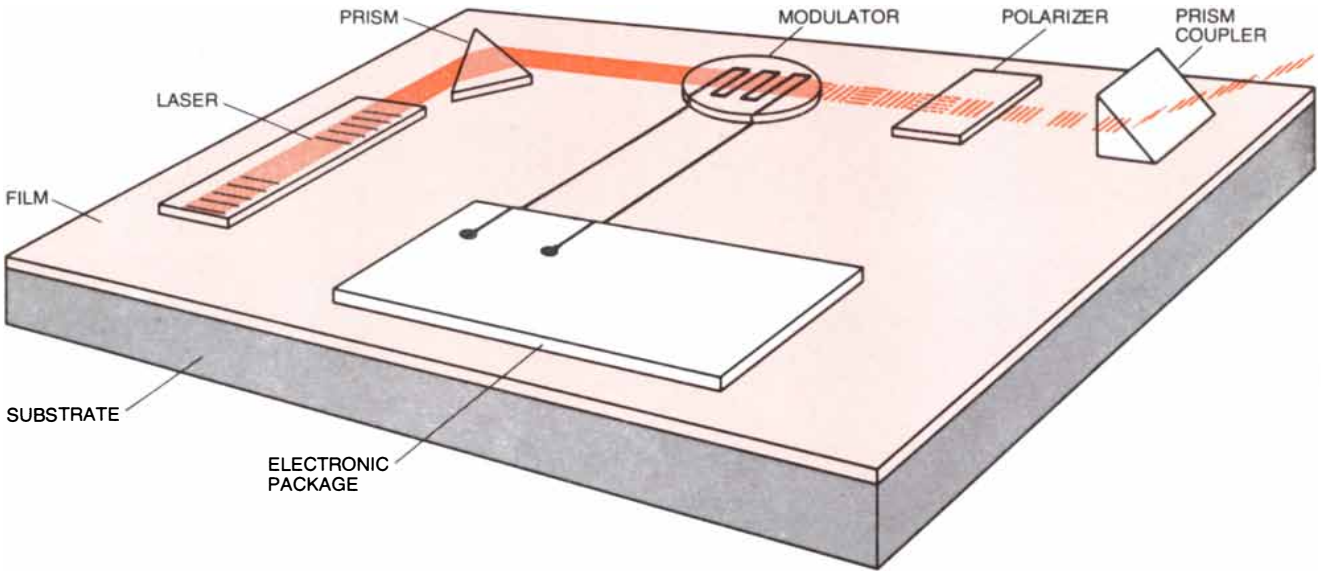
The effect of the different directions of the magnetization vector is that the light wave is polarized first in one direction and then in a perpendicular direction. In a real integrated optical circuit the light that had been modulated by an optical switch would proceed on to a thin-film polarizer (simply a strip of aluminum film deposited on top of a thin-film waveguide). Depending on the direction in which the light is polarized, the beam will either be absorbed or be transmitted by the polarizer. It can thus be turned on and off according to the code imposed on it by the electric circuit.

Active devices such as lasers and modulators must be formed out of films that have properties much different from those required by passive devices such as prisms or lenses. For this reason there are two different kinds of thin film, and they are made in two different ways. The first kind, which is used for the active devices, is a single-crystal film: a film consisting of a regular lattice of atoms. Single-crystal films are made by taking a substrate that has a certain lattice structure and dipping it into a vessel containing a solution of the film material. The film grows epitaxially, that is, it builds up atomic layer by atomic layer in a lattice structure that matches the structure of the substrate. The material of the film can be either different from that of the substrate or the same. If it is different but has the same lattice structure, the film is said to be heteroepitaxial.

The second kind of film, which is used for the passive devices, is amorphous: it has no crystal-lattice structure. It can be deposited on the substrate by one of two methods. In the evaporation method the film material is heated until it evaporates. It then condenses on a cooler substrate to form a thin film. Alternatively



TWO DIFFERENT FILMS CAN BE JOINED in an actual circuit by a third film that overlaps them both. All three films must have tapered edges in order for the light to travel from one to the other without scattering. Refractive indexes of the two films can be different and can be joined by a third film of intermediate index. Moreover, connecting film can join devices of different thicknesses.



COMPLETE OPTICAL CIRCUIT made entirely of thin-film devices may soon be fabricated on a substrate having an area of only a few square centimeters. In this circuit the light is generated by a tiny laser and directed around a corner by a prism, where it enters a magneto-optical switch. Signals from the small electronic package modulate the light, which then enters a polarizer. The light wave, now modulated by the signals, can be coupled into a glass fiber and transmitted many miles away for communication purposes.

the film can be sputtered onto the substrate in a vacuum chamber. Ions and electrons are accelerated toward a target of the film material with such force that they knock atoms out of it. The atoms settle rather like falling snowflakes and build up a thin layer of the film material on the substrate.

All the thin-film devices I have been discussing are basically thin-film waveguides built up out of different materials. It would be convenient if the devices could be manufactured first as separate units on a common substrate and then interconnected into a circuit. A light-conducting connection between two films of different thicknesses and refractive indexes is in itself an interesting problem. One simple method has recently been demonstrated experimentally. First the edges of the two thin-film devices that need to be connected are tapered. Next a third film is deposited between the devices, overlying part of each device and the substrate between them. The film is thus in direct contact with both the devices and the substrate. The ends of the connecting film are also tapered, providing a smooth transition for the light traveling from one device to the next [*see illustration on opposite page*].

The continued evolution of integrated optics depends heavily on the availability of materials suitable for the construction of thin-film devices. Single-crystal films are needed for modulators and lasers, and the higher the quality of the crystal is, the more efficiently the devices can function. Since the refractive index of the film must be larger than that of the substrate, it is necessary to use a material for the film that is different from the substrate material. To grow such a heteroepitaxial film, however, requires that the crystal-lattice structure of the film be closely matched with that of the substrate. If it is not, the surface of the resulting film will be rough and will need to be polished, and polishing a thin film is quite tedious.

Two unique systems of materials have been studied intensively. One is the aluminum-gallium-arsenic system. The lattice structure of aluminum arsenide exactly matches that of gallium arsenide at a temperature of 900 degrees Celsius. At that temperature a perfect film of $Al_xGa_{1-x}As$ can be grown on a gallium arsenide substrate. The other system is represented by the garnets, whose lattices can be modified by adding different ions of the rare-earth elements. One can therefore choose one garnet for the film and another for the substrate, so that the two lattices match exactly. So far the garnet films are the best single-crystal films available.

For some time to come integrated optical circuits will probably consist of devices made of different systems of materials. The trend for the future, however, is toward monolithic integrated optics: circuits built of materials in one system. It is already possible to develop monolithic integrated optics based on either the Al-Ga-As system or the garnet system. One difficulty inherent in the present Al-Ga-As system, however, is that gallium arsenide waveguides absorb the light from $Al_xGa_{1-x}As$ lasers instead of transmitting it. The system could be enlarged to include the elements indium (In) and phosphorus (P). Lasers, modulators and waveguides made out of the materials in such a system would be compatible with one another. One system being considered now utilizes $In_yGa_{1-y}As$ injection lasers for generating radiation at 1.05 microns and $Al_xGa_{1-x}As$ films for waveguides and modulators. Alternatively garnets can be used for neodymium lasers, magneto-optical switches and waveguides.

Although there is much work to be done before practical integrated optical circuits can be realized, the progress made over the past few years has laid a firm foundation for future development. Integrated optics calls for an integrated effort by physicists, materials scientists and electronic engineers. One can expect the field to continue to expand. Miniature integrated optical circuits may well be the next major development in the evolution of faster and cheaper devices to satisfy the ever increasing demands of communication, information processing and control.

THE EMBRYO AS A TRANSPLANT

Pregnancy can be regarded as a graft-host relationship, like that established by an organ transplant. It is an unusual one in that the embryo seems to be indifferent to the maternal immune response

by Alan E. Beer and R. E. Billingham

A graft of living tissue obtained from someone unrelated to the recipient is ordinarily destroyed by the host within a week or two. An important instance of this immune response is the rejection of a transplanted organ, such as a kidney, but grafts of less complex structure (such as skin) and even suspensions of dispersed cells (such as leukocytes) are also vulnerable.

One kind of graft tissue, however, seems to be exempt from the consequences of the host's immunological reactivity: the tissue of the fetus. At the placenta the tissue of one individual is in intimate and extensive contact with, and indeed is united with, that of another, a condition that ordinarily would provoke a strong reaction, yet the graft of the fetus to the mother is a perennial success. The reproduction of all mammals depends on it.

This paradox has often been neglected, primarily because mammalian reproduction has not traditionally been recognized as a grafting process. That it is such a process can be readily demonstrated, although the mechanism of this peculiar exemption from immunological attack has not yet been fully explained.

The function of the immune system is to recognize and inactivate pathogenic microorganisms and their products; there is evidence that the system can also act against cells that have become malignant. In order to perform this function the immune system must discriminate between "self" and "nonself"; it does so by detecting the presence of antigens, chemical groups on the surfaces of cells and other biological materials that determine their identity [see "The Immune System," by Niels Kaj Jerne, SCIENTIFIC AMERICAN, July, 1973, and "Markers of Biological Individuality," by Ralph A. Reisfeld and Barry D. Kahan, SCIENTIFIC AMERICAN, June, 1972].

Because an individual's complement of antigens is genetically determined and virtually unique, the immune system can also recognize as alien the cells of another member of the same species. If the individuals are not closely related, their cells are likely to differ considerably with regard to the particular antigens that determine the response to transplantation. Tissue transferred between two such individuals is called an allograft or a homograft. (An interspecies graft is a xenograft or heterograft.)

The recipient's response to an allograft has two components. Transplantation antigens and other foreign material are carried through the lymphatic system to the lymph nodes, where they stimulate the production and release of "sensitized" lymphocytes, or effector cells. These cells enter the bloodstream and, on reaching the graft tissue, infiltrate and begin to destroy the alien cells. In addition to killing graft cells directly the sensitized lymphocytes secrete a variety of chemical agents called lymphokines, some of which act directly on the foreign cells and some of which attract other leukocytes, including macrophages, which digest damaged cells and cell fragments. The mixed population of leukocytes that assembles in the allografts is in some way induced to cooperate in the attack on the foreign tissue [*see illustration on page 38*].

The "humoral" antibodies, which compose the second component of the immune system, are protein molecules whose structure permits great variation at specific sites; in concert with the complex of enzymes known collectively as complement they are capable of destroying certain kinds of allograft cells, in particular those of the lymphoid and blood-forming tissues. In the rejection of almost all grafts of solid tissue, however, it is not the humoral antibodies but the cell-mediated component of the immune response that is responsible for tissue destruction. In fact, in most kinds of grafts antibodies tend to inhibit both the development and the expression of cell-mediated immunity.

The host's immune reaction to an allograft can be palliated by either of two procedures. In experiments with animals tolerance of specific transplantation antigens can sometimes be induced by exposing the animal to them early in life. In human patients the repudiation of transplanted organs is averted by the administration of immunosuppressive drugs. Both of these methods involve tampering with the immune system and, in the latter case, disabling it. The "natural" graft represented by the fetus, on the other hand, prospers with the maternal immune system intact.

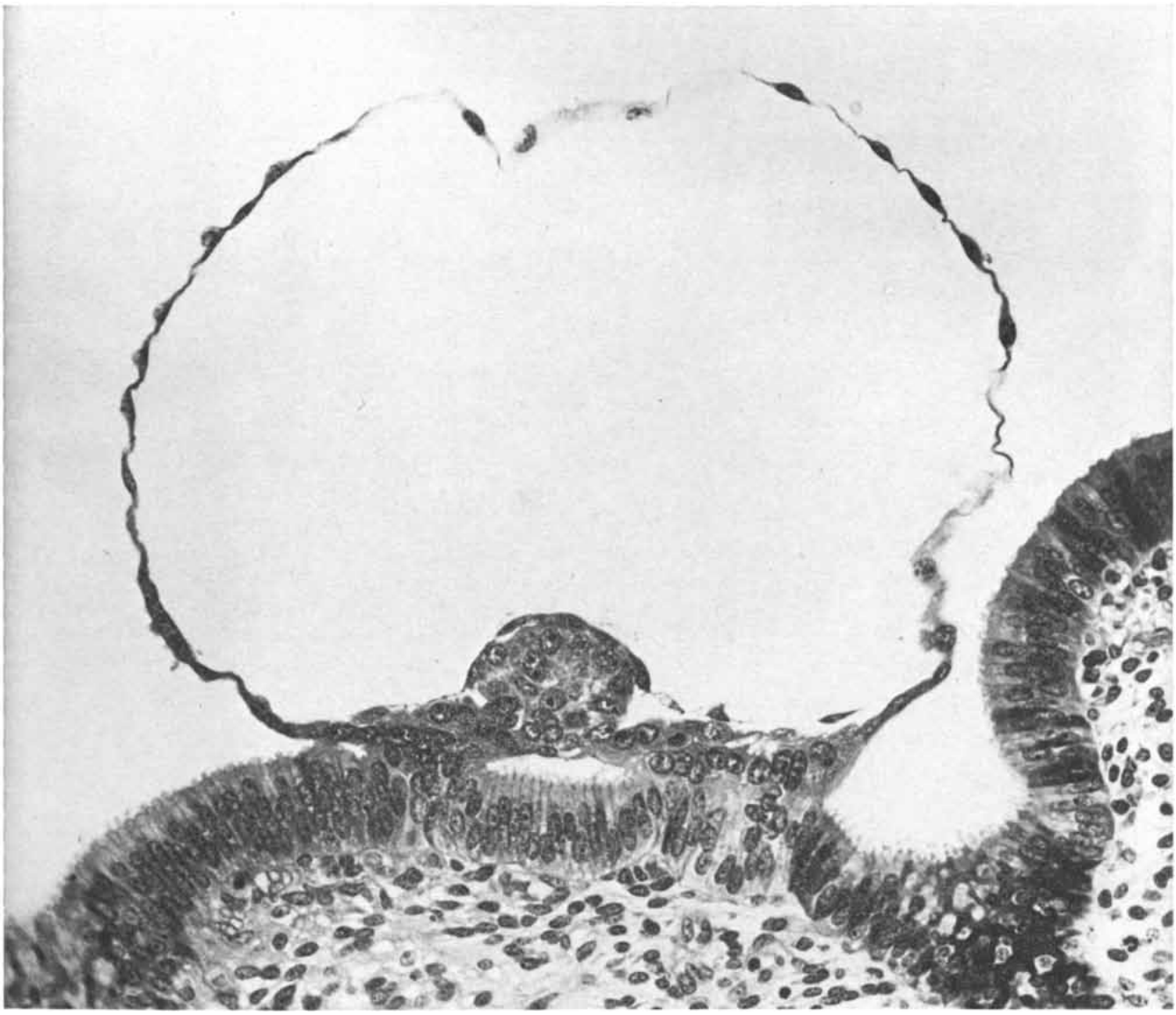
The fetus can be regarded as an allograft at all stages of its existence. Conception normally begins with the "inoculation" of the female host with spermatozoa from a male donor not genetically related to her. These specialized cells have been shown to bear transplantation antigens and therefore have the potential of eliciting an immune response. The zygote formed by the union of a spermatozoon and an ovum has equal genetic endowments from each parent, including the histocompatibility genes that determine the offspring's transplantation antigens. Of these genes an indeterminate number are not identical with the mother's. The zygote subsequently divides many times to form the hollow mass of cells called a blastocyst. The blastocyst eventually becomes embedded in the surface layer, or endometrium, of the uterus, that is, it is engrafted. Even at a very early stage of de-

velopment, when the zygote has divided into only two cells, transplantation antigens can be detected.

The healing in of the zygote on the endometrium initiates the intimate parasitic or parabiotic graft-host relationship called gestation. The embryo is confined within a quarantining, fluid-filled sac, the amnion, which in man consists of embryonic tissue closely applied to but not engrafted to the uterine wall. At the placenta, however, a layer of embryonic cells, the trophoblast, is united with maternal tissue. The trophoblast develops as soon as the blastocyst is implanted and, in a cancerlike manner, invades the neighboring endometrium, including its blood vessels. (There is, however, no direct connection between the maternal and the fetal circulatory systems.)

As a final indicator of the fetus's status as an allograft, birth might be considered the repudiation of the graft by the host. In this case, however, the relation of reproductive to grafting biology is superficial: parturition is not initiated or mediated by the immune system, and it is not therefore truly equivalent to the rejection of a graft. In fact, the conspicuous disparity between the ultimate fate of the fetus and that of a more conventional allograft leads one again to ask what special dispensations exempt the fetus from destruction.

A first assay at answering this question might be to postulate that the fetus is protected because half of its chromosomes, with their associated histocompatibility genes, are inherited from the mother. This notion must be discarded. If it were correct, mothers would be expected to tolerate the tissues of their offspring; actually, in experimental animals skin grafts surgically removed from the fetus, or from the progeny at any stage after birth, are routinely rejected by the mother. Furthermore, an embryo produced by the mating of two unrelated individuals will thrive when it is transferred to the uterus of a third and also unrelated surrogate mother. The fetus is protected even though the "mother" in whose uterus it grows has



EMBRYO ADHERING TO THE UTERUS represents a graft of complex tissue that is genetically alien to the mother. The embryo, that of a rhesus monkey, is the hollow shell extending to the top; it is enlarged about 400 times in this photomicrograph. It is seen in cross section at the age of nine days, shortly after the process of implantation began. The layer of vertically aligned cells on which the embryo rests is the endometrial epithelium, the lining of the uterine cavity; below the endometrium is the decidual tissue, which, as the embryo matures, proliferates in response. In the two areas where the embryo and the uterus are joined the trophoblast has begun to invade the endometrium. The trophoblast is of embryonic origin and eventually contributes to the formation of the placenta.

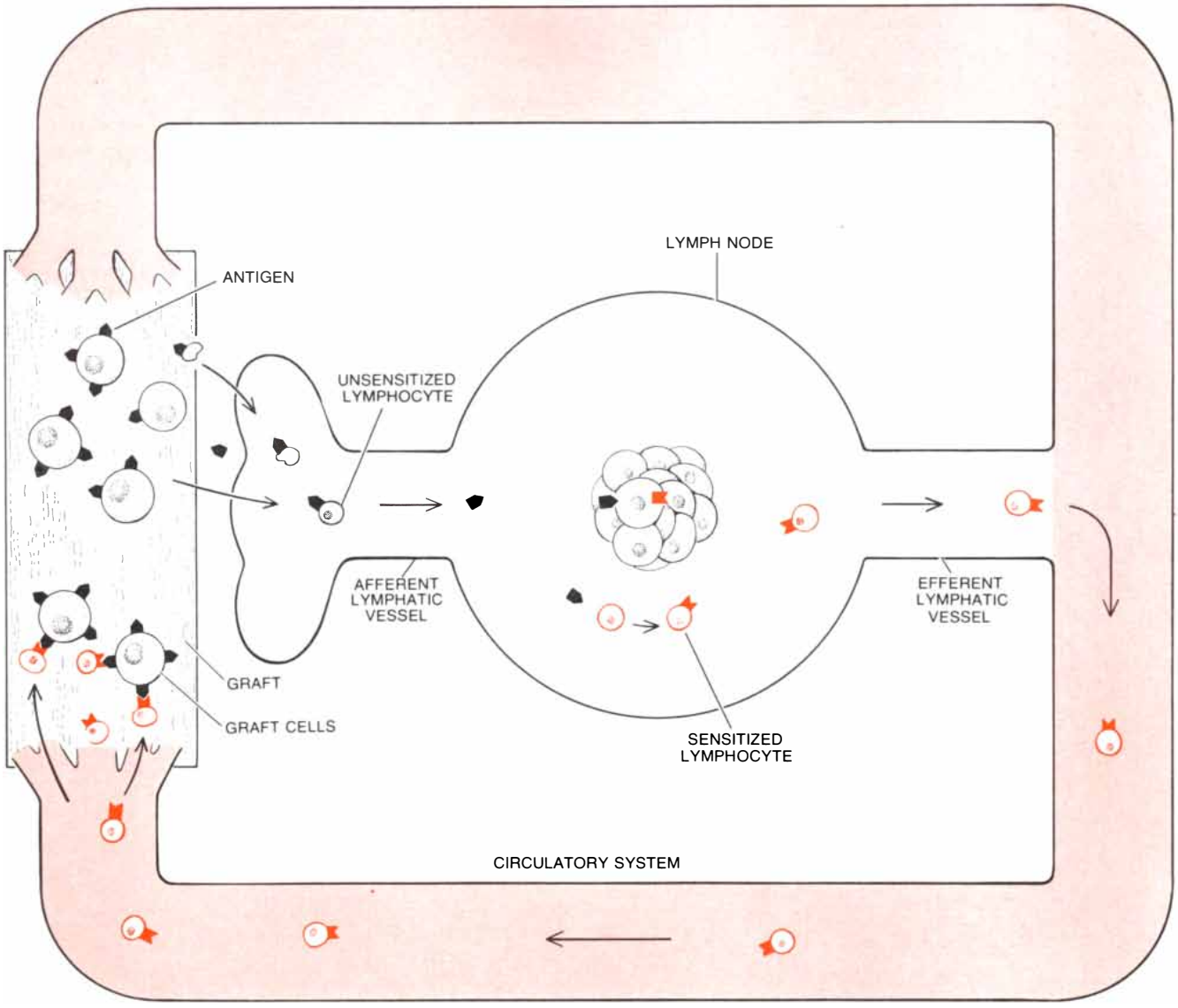
made no genetic contribution to it and has few histocompatibility genes in common with it.

If the mechanism that protects the fetus does not reside in the partial correspondence between fetal and maternal histocompatibility genes, it is reasonable to ask if the site of the graft, the uterus, might be an immunologically privileged one. There are a few such sites in the body, notably the brain and the anterior chamber of the eye, where allografts are sustained for anomalously long periods. The most extensively studied of these sites is the cheek pouch of the Syrian hamster [see "Skin Transplants and the Hamster," by Rupert E. Billingham and Willys K. Silvers; SCIENTIFIC AMERICAN, January, 1963].

These sites have in common an absence or a paucity of draining lymphatic vessels; tissue grafted to them is not destroyed because antigenic material is not readily transported to the lymph nodes, and sensitized lymphocytes therefore are not produced. Clyde F. Barker of the University of Pennsylvania School of Medicine has shown that in rodents privileged sites can be created artificially. Barker raised a flap of skin from the trunk of guinea pigs, maintaining communication to the body of the animal only through a narrow pedicle that contained an artery and a vein but no lymphatic vessels. Skin allografts transplanted to the flaps were tolerated for many weeks [*see bottom illustration on opposite page*].

Experiments with animals that have previously been immunized against the antigens of an allograft donor tend to confirm this explanation of privileged sites. The host can be immunized by inoculation with leukocytes or spleen cells from the donor, or by grafting donor skin to a conventional site. Either procedure results in the production of lymphocytes sensitive to the donor's transplantation antigens. A subsequent graft to a privileged site will be destroyed in the conventional manner and the viability of a hitherto sustained graft will be terminated since lymphatic connection to the graft tissue is required only to stimulate the immune response, not to execute it.

The uterus is supplied with lymph



IMMUNE RESPONSE to a graft of alien tissue is centered in the lymphoid tissue. Antigen characteristic of the graft, in the form of cells and cell fragments, is transported through an afferent lymphatic vessel to a lymph node, where it stimulates the proliferation of lymphocytes sensitized to the specific antigen. These cells leave the lymph node through the efferent lymphatic vessel and are eventually released into the bloodstream and carried to the site of the graft. There they infiltrate the graft tissue and initiate the processes that destroy it, both by attacking cells directly and by enlisting the cooperation of other components of the immune response. In this diagram the immune system has been greatly simplified; in particular the role of the humoral antibodies is not portrayed.

vessels and nodes. It therefore appears, at least at first glance, that it cannot be privileged in the same way that the cheek pouch of the hamster is. We nevertheless decided to investigate the possibility that it may provide a particularly hospitable location for allografts. In our laboratory at the University of Texas Health Science Center at Dallas we introduced small skin grafts and suspensions of epidermal cells into the uterine lumen, or cavity, of laboratory rats. The grafts served as false, or model, embryos [*see illustration on next page*].

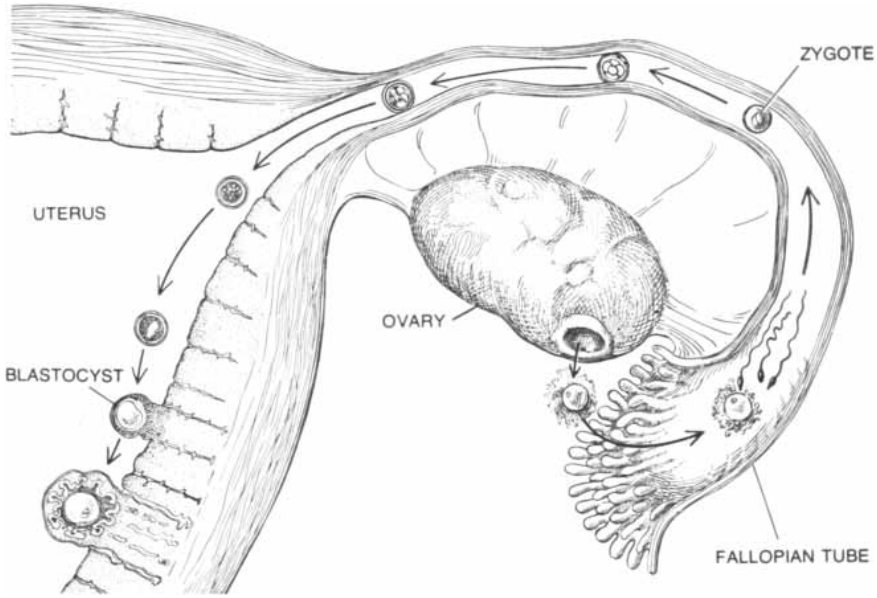
When the graft tissue was obtained from genetically compatible donors, it readily became established on the endometrium of the uterus. It was not necessary to implant it in the uterine wall, nor was the maternal tissue specially prepared for the graft. One endocrinological event that normally accompanies pregnancy was a prerequisite to the success of the transplantation: a surge in the serum level of estrogen, a hormone secreted by the ovaries. In our experiments estrogen was injected into the rats, or its production was stimulated by mating the rats before attempting the graft. If these preconditions were neglected, the grafts failed to become established and were soon extruded through the vagina.

These grafts, of course, were from compatible donors. Allografts introduced into the uterine lumen also became implanted if a suitable estrogen level was maintained. After implantation, however, they were soon rejected. They survived no longer than similar tissue grafted to the trunk of the host animals.

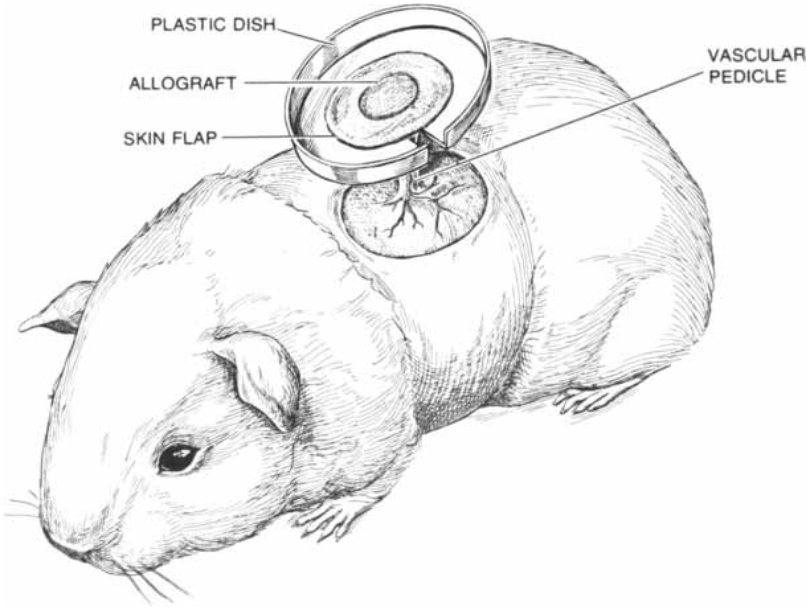
Even when the estrogen level was elevated, our experiments did not entirely simulate normal pregnancy, and our grafts of skin cells were not perfect model embryos. In particular, the grafts failed to stimulate the "decidual response," a proliferation of the connective tissue of the uterus in the region of the implanted blastocyst. It is this decidual tissue that the late David Kirby of the University of Oxford suggested might be partly responsible for the immunological protection of the fetus.

We tested this hypothesis in normal pregnant rats at the stage before the blastocyst is implanted and in rats made pseudopregnant by the injection of progesterone, another ovarian hormone, as well as estrogen. Under these circumstances skin allografts placed in the uterus did provoke a decidual response, and their survival was significantly prolonged. The possibility that this attenuation of the host's response was a result of the systemic effects of elevated hormone levels was tested by making allografts to conventional sites on pregnant and pseudopregnant hosts; the grafts were promptly rejected.

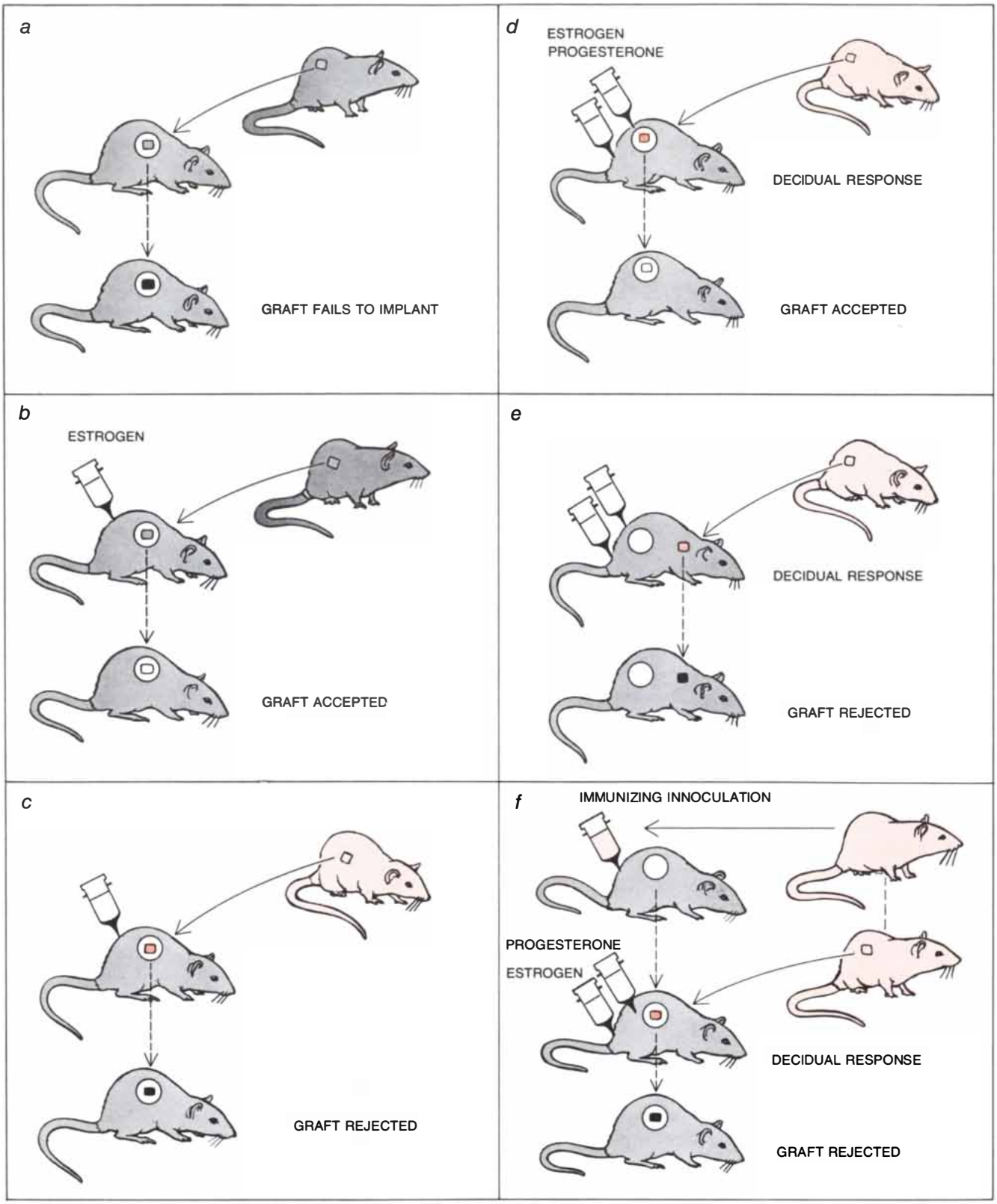
These promising results were contravened by the discovery that decidual tissue gives no protection to allografts when the host has previously been immunized against the antigens of the graft tissue. This finding is contrary to the observation that the fetus is indifferent to attempts to immunize the mother against it. Experimental efforts to prevent the implantation of the blastocyst or to de-



GRAFTING OF EMBRYO TO UTERUS takes place about a week after fertilization. The zygote is formed by the union of the ovum and spermatozoon in the fallopian tube; while still a free-floating organism it divides repeatedly to form the hollow sphere of cells called a blastocyst. At a very early stage, when the embryo consists of only two cells, transplantation antigens can be detected on its surface. Its implantation on the endometrium initiates the trophoblastic invasion and the associated decidual response of the uterus.



IMMUNOLOGICALLY PRIVILEGED SITE was created artificially in guinea pigs in an experiment performed by Clyde F. Barker of the University of Pennsylvania School of Medicine. A flap of skin was raised from the animal's trunk and prevented from healing by inserting a slotted plastic dish under it. The flap was supplied with blood through a narrow pedicle containing an artery and a vein but no lymphatic vessel. Skin grafted to the flap survived for many weeks, much longer than skin grafted to conventional sites, suggesting that communication with lymph nodes is required if the immune response is to be evoked.



HYPOTHESIS that the uterus is a privileged site was tested by grafting skin cells as model embryos to the uterine endometrium of rats. Grafts from rats of the same inbred strain (*gray*) failed to implant (*a*) unless one condition of normal pregnancy was simulated: an elevated level of estrogen. When estrogen was injected, these genetically compatible grafts implanted and survived indefinitely (*b*). Estrogen had no effect, however, on grafts obtained from donors of another strain (*color*), which healed in but were soon rejected (*c*). The survival of these allografts was significantly prolonged only when another condition of pregnancy was elicited: the decidual response (*d*). This response was evoked by injecting the recipients with both estrogen and progesterone. That these hormones did not induce a systemic weakening of the immune response was demonstrated by grafting skin from the same alien strain to the trunks of similarly prepared recipients (*e*); these grafts were rejected. Although these results suggested that the uterus might be a privileged graft site, a final experiment strongly indicated the opposite. When the recipient was immunized against the donor strain before the graft was made (*f*), the model embryo was rejected even when the decidual response had been elicited.

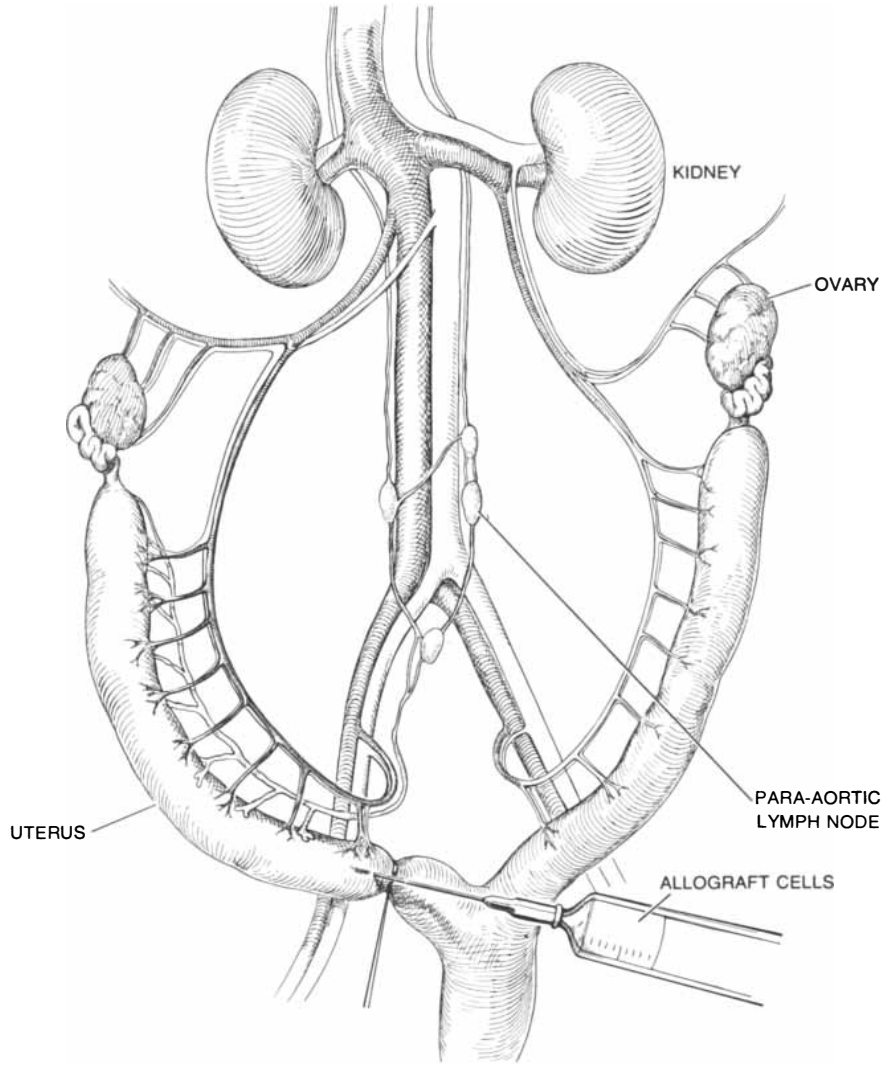
stroy the fetus *in utero* by making the mother sensitive to the father's transplantation antigens have repeatedly failed. (We later succeeded in producing damage in fetuses through the action of maternal lymphocytes, but only through a procedure that was not followed in these experiments and that does not alter the interpretation of them.)

Because the decidual tissue affords no protection to a skin allograft in a hyperimmunized host it appears unlikely that this tissue makes a significant contribution to the protection of the fetus. The invulnerability of the fetus to prior immunization of the mother is in fact the most perplexing of its attributes as an allograft.

Allografts at most sites in the body provoke hypertrophy, or enlargement, of the regional lymph nodes, an indication that immunological processes have been stimulated. Allografts introduced into the uterine lumen cause an enlargement of the para-aortic node, which is apparently the main site of reactivity against intrauterine grafts. Observations of rats, mice, hamsters and women have shown that pregnancy with genetically incompatible fetuses also causes striking hypertrophy of these lymph nodes. The obvious import of this discovery is that the mother is stimulated immunologically by antigenic material of fetal origin.

This conclusion can be further tested by experiments intended to induce hypersensitivity. Ordinarily exposure to and rejection of one allograft greatly hastens the rejection of any later grafts of genetically similar tissue. This is the "second set" phenomenon, the classical expression of transplantation immunity. Ordinary allografts in the uterus provoke this response but the fetus does not, even though the regional lymph nodes are observed to be stimulated. Pregnancy with a genetically incompatible fetus does not hasten the rejection of subsequent grafts of paternal tissue, and indeed it may delay it. (The fetus does stimulate the production by the mother of humoral antibodies corresponding to the major paternal antigens. These antibodies do not play a significant role in the rejection of grafts, but they are of value in a clinical procedure; as found in the blood serum of multiparous women, the humoral antibodies are employed in tissue typing for the selection of donors for organ transplantation.)

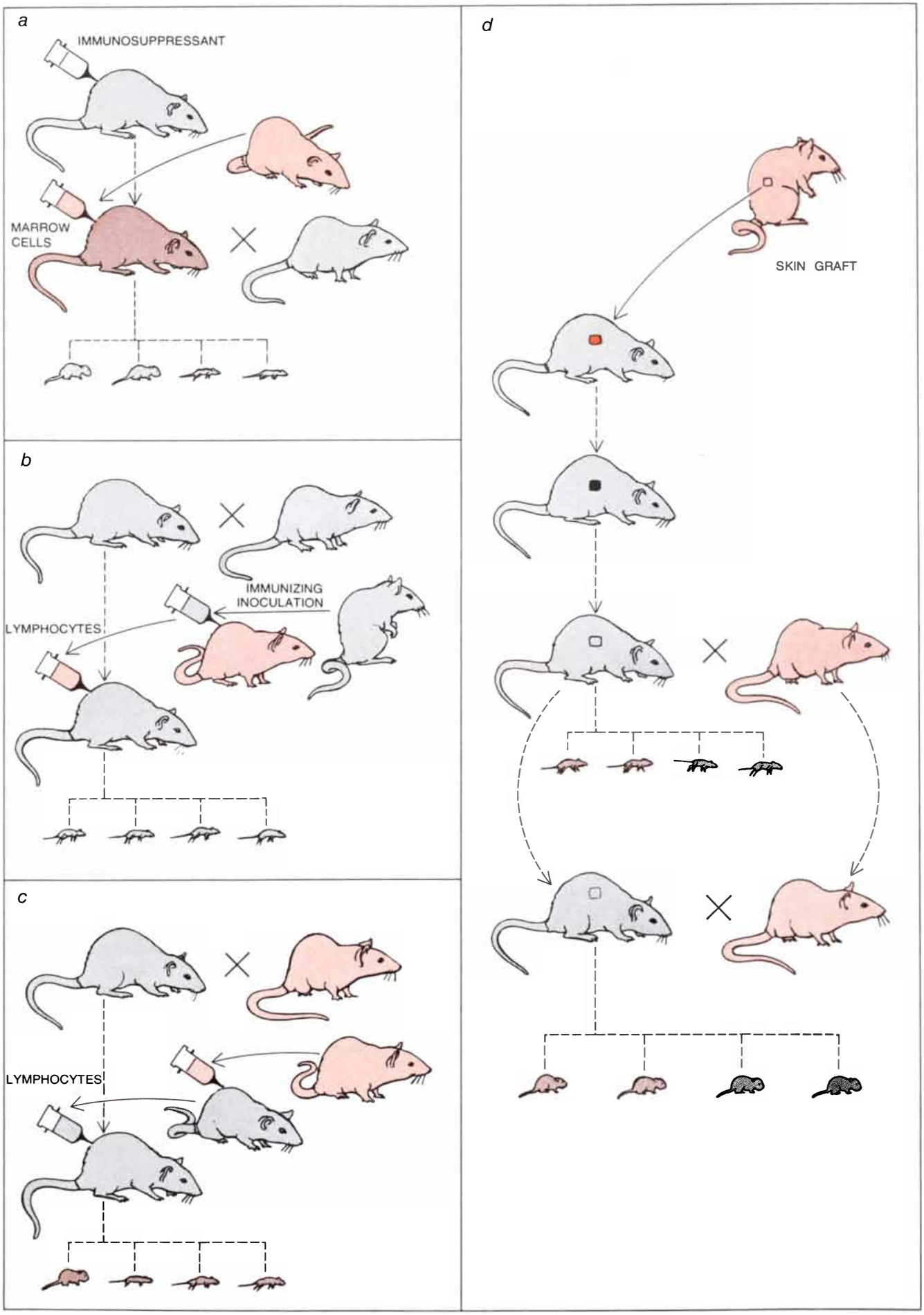
The uterus can readily be made hypersensitive to foreign tissue antigens. Following an initial exposure, rechallenge with identical cells a few weeks later provokes a profound, transient swelling and inflammation in 24 to 48 hours. This response is not observed when the initial inoculation is made at some site other than the uterus and is probably caused by the persistence of "memory" cells sensitive to the antigen that remain in the uterine wall at the site of the first exposure. This effect is observed whether the challenge inoculum consists of skin cells, leukocytes or other types of cells, including spermatozoa; we decided to determine whether or not the blastocyst or very young embryo would incite and succumb to this local hyperreactivity.



EFFECTS OF LOCAL IMMUNIZATION were tested in the bicornate, or two-horned, uterus of the rat. An allograft of dispersed cells was injected into one horn; subsequent grafts to that horn from the same donor were quickly rejected. The immunizing graft did not impair the reproductive function of the horn, however; mating with males of the donor strain produced more embryos in the immunized horn than in the unimmunized one.

The uterus of the rat is bicornate, that is, it has two "horns" that meet near the vagina. In our experiments one of the horns was immunized against the antigens of an unrelated strain of rats by an inoculation of dispersed cells [*see illustration above*]. The rats were later mated with males of the donor-cell strain. Before the pregnancy was completed the rat uteri were examined in order to compare the number of fetuses that developed in the immunized horn, the number in the nonimmunized horn and the number in each of the uterine horns of control animals that had not been immunized. We found that local immunization did not prejudice the maternal tissue against the presumably susceptible fetus; on the contrary, immunization increased the reproductive efficiency of the sensitized horns.

The weight of evidence of these various experiments indicates that the uterus is not an immunologically privileged site. Antigenic material released by intrauterine allografts does reach the regional lymph nodes, and it provokes a response there that under certain circumstances



can cause the rejection of the graft. Clearly the uterus is an organ in which transplantation antigens are readily detected and in which transplantation immunity is fully expressed.

It is equally clear, however, that the immune response of the maternal host does no harm to the fetus. Although the lymph nodes are stimulated, and even though the mother has been immunized against the paternal antigens, the fetus thrives.

Under certain circumstances the characteristic invulnerability of fetal tissue can be breached. A familiar example, in which the fetus is dramatically attacked by the maternal immune system, is hemolytic disease of the newborn, or erythroblastosis fetalis. It occurs when a mother who is Rh-negative and therefore lacks one of the important antigens of the Rh system on her erythrocytes bears children who have inherited the antigen from their father. Because fetal erythrocytes migrate across the placenta and enter the maternal circulation the mother develops antibodies to the Rh antigen, although usually not until after her first pregnancy; the antibodies readily cross the placenta, destroy the Rh-positive fetal red blood cells and produce a potentially fatal illness in the fetus. Recognition of the cause of this disease in 1939 by Philip Levine of the Newark Beth Israel Hospital and R. E. Stetson of the New York University School of Medicine established for the first time that there are immunological hazards of gestation.

The maternal agents that cross the placenta in erythroblastosis fetalis are, of course, humoral antibodies; it is now well documented, however, that cells too can pass this barrier. In man, fetal erythrocytes, leukocytes and other kinds of cells are continually entering the maternal bloodstream, while maternal lymphocytes and erythrocytes cross the placenta in the opposite direction, albeit in small numbers.

There are two possible consequences of the infiltration of the fetus by maternal leukocytes. First, the fetus may to some degree become tolerant of the maternal transplantation antigens and as an adult manifest impaired capacity to reject grafts of maternal tissue. This effect was discovered by P. B. Medawar and his colleagues at University College London in 1953. Fetal mice were inoculated with cells from adult mice of an unrelated strain. When the inoculated mice were tested at maturity, they showed partial or total inability to destroy grafts from the donor strain.

The second possible consequence is graft-v.-host disease, a wasting syndrome that results in the birth of runts, infants that initially appear to be healthy but whose development is soon grossly retarded. It is often fatal to the offspring. This graft-v.-host syndrome is a "laboratory disease" procurable at will by a variety of methods, most simply by inoculating newborn mice or rats with a few million lymphocytes from adult donors of another strain. The disease is caused by the reactivity of the mature, immunologically competent donor lymphocytes against the immunologically immature or incompetent perinatal recipients.

These two possible consequences of the interchange of maternal and fetal cells are not mutually exclusive. Previous attempts to demonstrate the first, tolerance of maternal tissues, without directly inoculating the fetus have failed. In a specially designed experiment we made a new attempt. Virgin female rats of the inbred Fischer strain were treated with an immunosuppressive drug, cyclophosphamide, that completely destroyed their lymphohematopoietic system. (This system comprises the lymphocytes and the tissues they are associated with: cells of the blood, bone marrow, spleen, lymph nodes and thymus.) Twenty-four hours later the rats received a transfusion of bone marrow cells from adult rats of the unrelated Lewis strain. George W. Santos and Albert H. Owens, Jr., of the Johns Hopkins University School of Medicine had previously shown that this procedure results in the complete replacement of the host's lymphohematopoietic system with cells from the donor. The Fischer rats became chimeric, that is, they came to consist of cells of two genetic types and were completely tolerant of grafts of Lewis skin [*see illustration on opposite page*].

When the chimeric Fischer rats had recovered from the tissue substitution, they were mated with males of their own strain. The resulting Fischer fetuses therefore developed in a uterine milieu marked by Fischer antigens, but the only "maternal" leukocytes to which they were exposed were of the alien Lewis type. We reasoned that if enough of these cells crossed the placenta and entered the fetal circulatory system, tolerance of Lewis grafts would be induced. The outcome could be tested simply by grafting Lewis skin to the infants when they were 21 days old; if the graft survived longer than similar tissue grafted to rats born of untreated mothers, transplacental traffic in lymphocytes would be strongly implicated.

Healthy litters were born, and almost all the infants we were able to test were tolerant of the Lewis grafts. A far more surprising development, however, was that within five weeks half of the infants developed runt disease and died. Eventually most of the chimeric mothers also developed the graft-v.-host syndrome. The lymphocytes responsible for producing runt disease in the offspring had to have been transmitted across the placenta. Thus in addition to procuring maternally induced tolerance we had unwittingly discovered a way to produce maternally induced graft-v.-host disease.

The occurrence of the disease in this experiment was unexpected. Ordinarily it develops only when the donor and the host possess different genes specifying the major histocompatibility locus of the species or when the donor has been previously immunized against the tissue antigens of the recipient. The Lewis and Fischer strains, although they differ at four of five histocompatibility sites and reject each other's skin within 10 days, are alike at the major site. For this reason

GRAFT-V.-HOST DISEASE, which leads to the birth of runts, is caused by the "infection" of the offspring with alien lymphocytes. It is commonly produced by inoculating newborn rats with lymphocytes from adult donors; in the series of experiments diagrammed here the authors demonstrated that it could be induced by the mother. Female rats of the Fischer strain (*gray*) were treated with an immunosuppressive drug, which destroyed their lymphoid tissues and certain other cells. A day later they received an injection of bone-marrow cells from rats of an unrelated strain (*color*) and as a result became chimeric (composed of genetically diverse tissue). When the chimeric females were mated with Fischer males (*a*), half of the offspring developed runt disease. The authors then mated pairs of normal Fischer rats (*b*) and inoculated the pregnant females with lymphocytes obtained from rats of an unrelated strain that had been immunized against Fischer antigens; almost all the progeny contracted the disease. In a third experiment Fischer females were mated with unrelated males and subsequently inoculated with lymphocytes obtained from rats that were of their own Fischer strain but that had been previously immunized against the antigens of the paternal rats (*c*). Again most of the offspring were runts. In the final procedure the mother was immunized directly against her offspring. Fischer females received grafts of skin from an alien strain. After the grafts had been rejected the females were mated with males of the donor strain and produced afflicted litters (*d*). Timing of the graft was crucial; second litters of the same parents were unaffected.

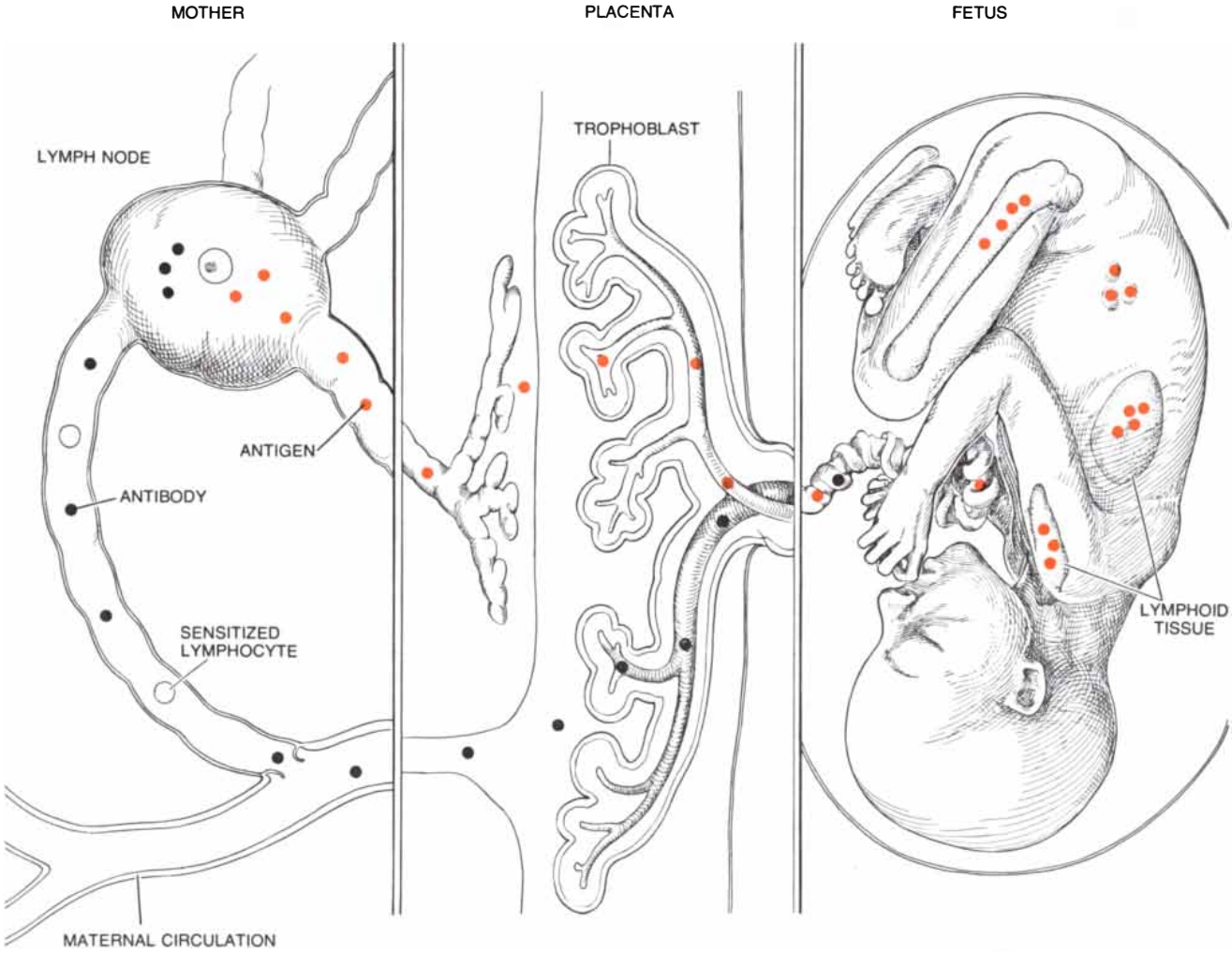
we suspected that the Lewis lymphocytes transmitted to the fetuses had become immunized to Fischer tissues, including the fetal tissue. Confirmation of this interpretation of our observations would mean that the fetal exemption from the consequences of the immune response is not absolute.

We sought to test this hypothesis by a less complex experimental procedure. Instead of creating chimeric animals we mated normal Fischer females with Fischer males and on the 15th day of gestation inoculated the females with 200 million lymphocytes obtained from Lewis rats that had previously been immunized against the Fischer-strain antigens. Ninety-one percent of the offspring were runts, but none of the mothers were harmed. Similar results were obtained with other donor and host strains.

One possible artifact of the experimental method remained to be eliminated: In all cases the transplanted lymphocytes were genetically incompatible with both the fetus and the maternal host. In order to determine if this dual incompatibility was necessary to produce the graft-v.-host reaction, we undertook another modification of the procedure. Fischer females were mated with males of the unrelated DA strain, so that the fetuses were hybrids (and therefore allografts) with transplantation antigens not possessed in common with their mothers. On about the 15th day of gestation the mothers were inoculated with 200 million lymphocytes from donors of their own Fischer strain that had been immunized against DA tissue antigens. In this case the injected cells were not incompatible with the maternal tissues but only with the fetal tissues; nevertheless, the incidence of runt disease among the progeny was high. When the Fischer females in this experiment were mated again with DA males, however, the second litter of hybrid offspring remained healthy.

The results of this experiment were difficult to reconcile with the failure of numerous previous attempts to immunize mothers against their hybrid offspring. A final experiment was devised to ascertain whether or not the effects we had observed could be obtained by direct, active immunization of the mother.

Virgin Fischer females were given small skin grafts from DA donors. One to two weeks later, when the grafts had been rejected, the Fischer rats were mated with DA males. Almost all the progeny succumbed to runt disease. This unequivocal result, which indicates that



IMMUNOLOGICAL RELATIONSHIP of mother and fetus involves the exchange of antigenic material in both directions. Cells and other material of fetal origin enter the maternal bloodstream and lymphatic system and are conveyed to the lymph node. Similarly, sensitized lymphocytes produced in the nodes, as well as other maternal cells, are transported by the maternal bloodstream to the placenta and enter the fetus. Humoral antibodies are also generated in the lymph node; their role appears to be inhibitory. Fetal lymphoid tissues, which develop relatively early in gestation, may be able to destroy limited numbers of maternal lymphocytes.

under certain circumstances the fetus is susceptible to attack by the maternal immune system, albeit rather indirectly, raises two questions: Why, in this experiment, was the immune response effective against the fetus, and why is it not effective more commonly?

Two clues are immediately apparent. The interval between immunization and mating was important to the outcome of the experiment; grafting three weeks before conception was harmless to the infants. In addition, second litters conceived by parents whose first litters had become runts were unaffected by the disease. Similarly, females that had not been immunized before giving birth to their first litter of hybrid offspring did not induce the disease in their second litters when they were subsequently immunized and mated again.

Both of these observations suggest the action of humoral antibodies in inhibiting graft-v.-host disease. It is well established that some antibodies can block the lymphocyte-mediated immune response, although the mechanism of this interference is not yet certain. The antibodies may "coat" the antigenic sites on the allograft cells and thus conceal them from the lymphocytes, or they may, by forming antigen-antibody complexes, prevent the stimulation of the lymph nodes and thereby diminish the production of sensitized lymphocytes. In the case of the fetus it is also possible that they may intervene at the placenta by combining with and thereby masking some antigenic element there that may otherwise facilitate the passage of lymphocytes from the maternal to the fetal bloodstream.

Whatever their mode of action, humoral antibodies are implicated as the inhibitors of runt disease by the required timing and sequence of the disease-causing procedure. When the antibodies are produced as a result of tissue grafting, they reach their highest level in the blood serum after the graft has been rejected; this suggests that the issue of delayed matings are protected by antibodies formed during the interim between graft rejection and mating. When antibodies are produced as a reaction to pregnancy, they can usually be detected only after the first pregnancy is completed. This, similarly, explains why second litters are not afflicted.

The failure of previous attempts to prejudice the survival of fetuses through immune reactions has two probable explanations. First, the experimenters may have hyperimmunized the mother. Many experiments called for the repeated grafting of skin, often with "booster" injections of suspended cells; this regimen is ideal for stimulating the production of blocking antibodies. Second, adverse effects of transplantation immunity on the fetus have generally been anticipated at the placenta, where the tissues of the two individuals are united. The expected consequence of the destruction of the fetal component of the placenta is spontaneous abortion; when this did not occur and litters were born normally, the investigators may have judged the experiment prematurely. Observation of the progeny for a few weeks *post partum* might have uncovered the unexpected phenomenon of maternally induced runt disease.

Experiments similar in design to those described here have shown that graft-v.-host disease with high mortality can be produced by maternal immunization in mice, guinea pigs, hamsters and rabbits as well as rats. It may also occur naturally in man. By the examination of infants who fail to thrive a few cases of runt disease in humans have been detected, some even documented by evidence of lymphocyte chimerism. It is possible that some of the fetuses that die in utero of undetermined causes are also victims of the infiltration of their mother's lymphocytes.

Normally two factors favor the fetus: the maternal blocking antibodies and the immunological competence of the fetus itself, which develops long before birth. Arthur M. Silverstein of the Johns Hopkins School of Medicine has found that the sheep fetus, which has a gestation period of 147 days, can reject skin allografts 70 days before birth. The maturation of immunocompetence in an environment that is usually regarded as sterile may represent a protective mechanism capable of disarming immigrant maternal lymphocytes, at least in relatively small numbers. In fact, maternal lymphocytes may be the first pathogens to which a fetus is normally exposed. Possibly supporting this theory is the observation that in most cases of established or suspected graft-v.-host disease in man there is evidence of congenital deficiency of the immune system.

There may be another manifestation of graft-v.-host disease in man. In 1959 Henry S. Kaplan of the Stanford University School of Medicine and David W. Smithers of the University of London noted a similarity between the abnormalities of animals suffering from experimentally induced runt disease and those of human patients with lymphomas, or malignancies of the lymphoid tissues. Recently Robert S. Schwartz and his associates at the Tufts University School of Medicine showed that these tumors may arise from the unmasking of latent oncogenic viruses by graft-v.-host reactions. The relatively high incidence of lymphomas in children could thus be explained as a result of subclinical runt disease produced by maternal lymphocytes that infiltrate genetically susceptible fetuses. The fact that lymphomas are more common in boys than in girls also conforms to our laboratory observation that maternally induced runt disease is more likely to affect males than females.

The discovery that mothers can be made to express a harmful immune response to their unborn offspring deepens the paradox of the fetus's status as an allograft. It is now beyond dispute that the mother can produce lymphocytes sensitized to fetal antigens, that these lymphocytes can cross the placenta and that once they have entered the fetal bloodstream they can cause extensive tissue damage, often resulting in death. Nevertheless, except when conditions are experimentally modified, the fetus almost never bears the wounds of this internecine battle.

It has been shown that the source of this privileged status of the fetus cannot be the uterus; it now also appears that it cannot reside in any special properties of the maternal immune system that would cause it to tolerate fetal tissue. Rather, the maternal immune response is apparently blocked or defeated by some mechanism associated with the fetus.

The most likely site of this mechanism is the trophoblast, the layer of fetal cells in immediate contact with maternal tissue at the placenta. Three items of evidence sustain this premise: experimental grafts of trophoblast tissue in mice were made by Richard L. Simmons and Paul S. Russell, who were then working at the Columbia University College of Physicians and Surgeons. Even when the trophoblast cells were grafted to anatomically unnatural sites, such as under the renal capsule, they failed to elicit the host's immune response; when the immune response was artificially stimulated, the trophoblast was resistant to it.

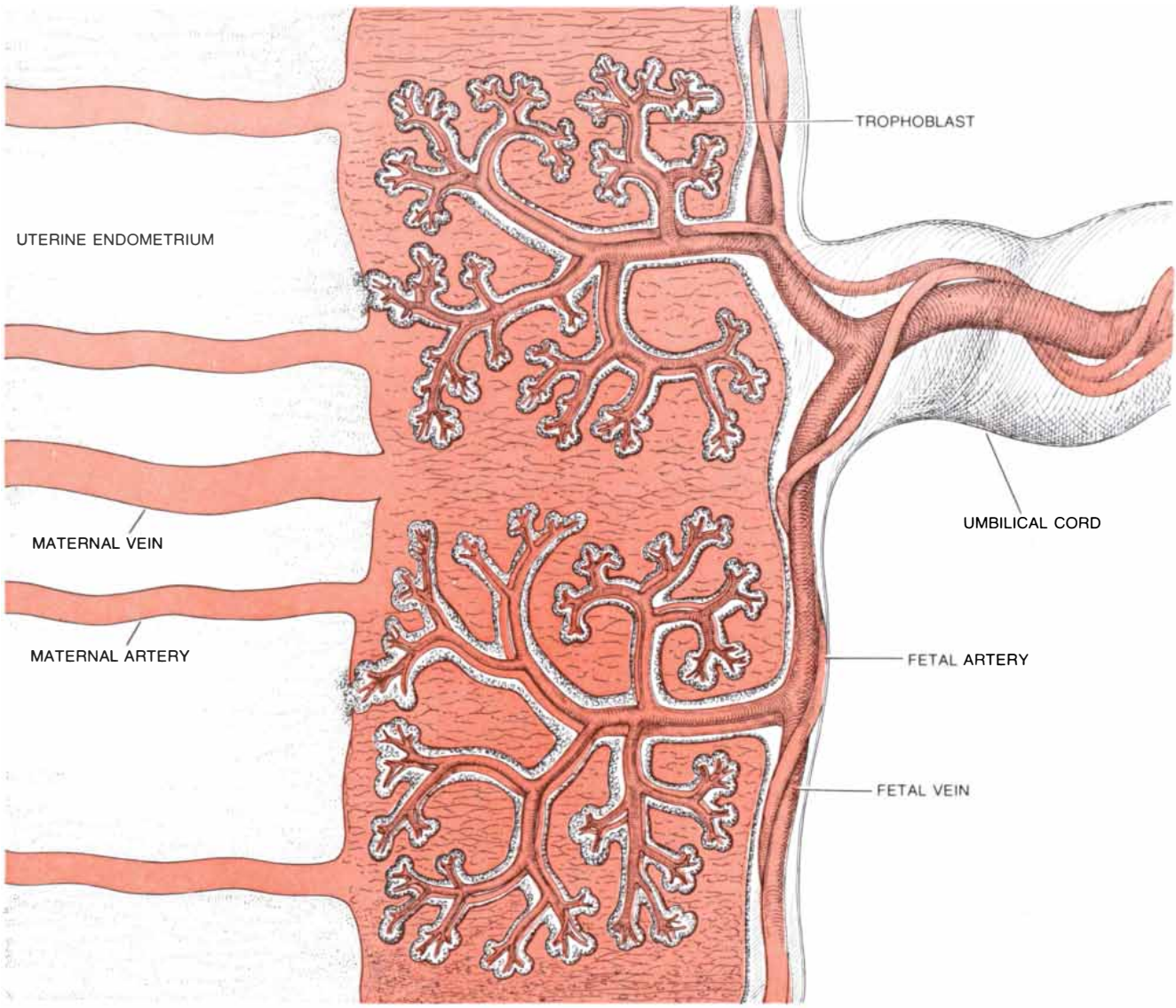
The same indifference to location and to maternal immunological reaction is demonstrated when a blastocyst is implanted at a site other than the uterine wall, such as in the fallopian tube or on the peritoneal lining of the abdominal cavity. This occurs occasionally by accident in both man and animals and can be artificially provoked in experimental

animals. In spite of the trophoblast's unsuitable environment it invades the substrate tissue and the fetus develops to an advanced stage, sometimes to term. No adverse effects of maternal immunity can be detected in the fetus.

In women, malignant and usually fatal tumors called choriocarcinomas occasionally develop from trophoblast cells. Because these tumors are of fetal origin their ability to metastasize as allografts through the body of the mother is extraordinary. Many kinds of cancer cells are antigenically different from other body cells, although the tumor-specific antigens are usually difficult to detect; choriocarcinoma cells, because they have an alien genetic component as well, should be readily distinguished and destroyed by the mother's immune system. Ingenious attempts have been made to enhance this presumed immunity to the paternally inherited antigens of the malignant cells, but the results have been equivocal.

Trophoblast cells and their malignant derivatives thus seem to be uniquely refractory to transplantation immunity. Their invulnerability to the immune response awaits a satisfactory explanation. Studies of dissociated trophoblast cells indicate that transplantation antigens are present on their membranes, but in vivo they nevertheless do not seem to incite an immune response nor are they susceptible to attack by sensitized lymphocytes. Their protection may reside partly in the glycocalyx, a cell coating that may mask transplantation antigens and may also repel lymphocytes.

A final, intriguing observation also pertains to the trophoblast and to the other placental tissues. There is evidence that the degree of antigenic disparity between mother and fetus, and the mother's status as being immunized or not immunized against the fetal antigens, may influence the extent to which the trophoblast invades the uterine tissue, the magnitude of the decidual response to trophoblastic invasion and the size and weight of the placenta. The greater the degree of genetic incompatibility, the more extensive the trophoblastic invasion and the larger the placenta.



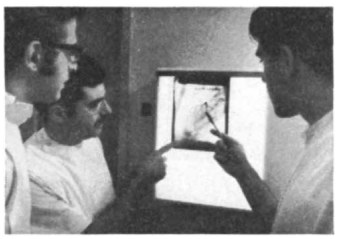
PROBABLE SITE OF FETAL EXEMPTION from the immune response is the trophoblast, shown here in a cross section of the mature placenta. Although the trophoblast is in contact with maternal blood and other maternal tissue, it apparently does not evoke an effective immunological reaction. When the immune response is artificially stimulated, the trophoblast seems to be invulnerable to it. It thus appears that the trophoblast may mask the transplantation antigens of the fetus and may also repel maternal lymphocytes.

If you were working for Kodak and this were your big accomplishment, how proud would you feel?

Rapid Processing Copy Film. A triviality? No, not when you understand its purpose.

Radiologists are physicians who more often see patients through a sheet of x-ray film than face-to-face. Meshing eye and intellect to make life-and-death judgments from the subtle shadings on that sheet of film takes a long time to learn—and a great many patient films begged and borrowed from experienced colleagues. The most interesting films that have the most to teach are, of course, the most in demand.

Obviously, interesting films can be copied and studied as slides on a projector. Why this hasn't been done much, and why so much of the learning must be done from the original film are less obvious.



Not that it's impossible to make slide copies of adequate quality from medical radiographs. It's just that copying has posed its problems.

For training the eye, the shadows should be as on the original radiograph—light against dark, not a "negative."

With this requiring two exposures and two developments, and with limitations on maximum density for projection conditions, it is not easy to hit everything just right when that critical tonal subtlety occurs at a different density from one copying job to the next. Direct-positive films can eliminate one exposure but may have complicated processing requirements. A department of radiology is not a photographic laboratory and is fussy about letting go of its patient films.

A direct-positive film that could be run through the x-ray processor right along with the daily flow of radiographs would be nice. Or, if that's not convenient, and since it should be a 35mm film to use in a single-lens reflex camera on a tripod in front of the illuminator bearing the noteworthy original radiograph, it should be suitable also for manual processing in a tank with ordinary chemicals.

All of which happens to be true of the new film. Magic has been performed on its silver halide such that, except where hit by enough light, the whole picture area goes to its maximum density upon development. That density is not too high for an ordinary slide projector. (The magic does not relieve the need to find the right exposure for copying the particular radiograph, nor do we recommend slides for original diagnosis.)

Dealers in x-ray supplies and equipment will know the product as KODAK *RP/C Film SO-185.*

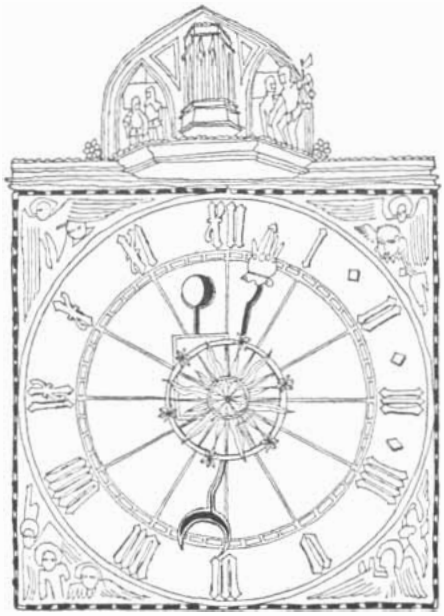
This quick route to 2″ x 2″ black-and-white slides is not just for radiologists.

KODAK RP/C Film SO-185 works just as well, of course, in copying charts or any black-and-white pictures for lecture use. Electron micrographers in particular may take to it. They often find that the generous exposures needed for best detail produce densities too heavy for good slide projection. This film, with its built-in limit on D_{max}, might turn out to be just the ticket for them. If stumped on a technical black-and-white copying problem, write Kodak, Dept. 756, Rochester, N.Y. 14650.

Kodak

An equal-opportunity employer f/m

SCIENCE AND THE CITIZEN



Triad

Ever since the development of long-range ballistic missiles in the 1950's there has been a tendency among U.S. military planners to regard this country's strategic nuclear forces as consisting essentially of three parts: intercontinental bombers, fixed land-based missiles and submarine-based missiles. During the 1960's this "triad" concept evolved to the point where it was felt that each component of the strategic forces should by itself be capable of deterring a surprise Russian nuclear attack (by being large enough or invulnerable enough to survive such a "first strike" attack and still retaliate by inflicting a certain level of "assured destruction" on targets in the U.S.S.R.).

The whole idea of the strategic triad has recently come under fire from various quarters. In the words of one critic, J. P. Ruina of the Massachusetts Institute of Technology, the "rule of three" has taken on "a sacred character quite outside the realm of the analytic decision-making process." Ruina argues that the changing technological and political situation raises the question of "how eternal this triangle need be. The current interpretation of the triad concept seems [to go] beyond the prudent pursuit of some variability in the strategic force . . . ; it appears almost as an inadvertent formula for a triple-sized arms race."

In particular, a growing number of arms-control advocates, including analysts in the Federation of American Scientists and at the Brookings Institution, have recommended that the U.S. and the U.S.S.R. agree to the phased elimination of their land-based strategic missiles, on the grounds that as such forces become increasingly vulnerable to attack by the other side their continued existence could itself become a destabilizing factor in the arms race. Dismantling the land-based missile forces on both sides, a policy statement of the federation points out, would still leave the U.S. its intercontinental bombers and submarine-based missiles, either of which could "provide a reliable deterrent" to any hypothetical Russian attack.

So far this point of view has the support of at least one prominent Administration official, Fred C. Iklé, the director of the Arms Control and Disarmament Agency. Recently Secretary of Defense James R. Schlesinger, in his annual "defense-posture statement," indicated that the U.S. might be willing to agree to a reduction of its land-based missile force as part of an overall arms-limitation agreement with the U.S.S.R. Nonetheless, in an apparent contradiction of this approach, the Administration appears determined to push ahead with the development and procurement of a large array of new strategic weapons, many of them elaborations of existing land-based missile systems.

Rime of the Modern Mariner

The instrumental operations of the *Mariner 10* spacecraft within 3,600 miles of the surface of Venus on February 5 went flawlessly and have made a substantial contribution to knowledge of the planet's nature and evolution. The flyby established an upper limit for the planet's magnetic field even lower than the very low one suggested by the earlier missions of *Mariner 2* and *Mariner 5*. It generated data for measurement of the planet's mass and determinations of its shape, improving on previous values by a factor of five. It provided analyses of the chemical composition of cloud layers. Above all, it resulted in exquisite high-resolution television pictures yielding detailed information about the structure and dynamics of the planet's atmosphere [*see illustration on opposite page*].

The television experiment, under the direction of Bruce C. Murray of the California Institute of Technology, generated images of the planet in the visible-light and ultraviolet regions of the spectrum. The visible-light pictures are completely featureless. The ultraviolet pictures, however, reveal a system of bands approximately parallel to the equator that are strikingly reminiscent of the cloud belts of Jupiter. Why these bands should be visible only in the ultraviolet is not certain, but according to Verner E. Suomi of the University of Wisconsin they may indicate a spiraling system of 200-mile-per-hour winds that carry heat energy from the equatorial region to the poles. At the poles the winds descend to lower levels and return to the equator. Since the pressure at the lower levels is 90 times greater than that at sea level on the earth, only a moderate flow of dense gas toward the equator would be needed to compensate for the high winds of the upper atmosphere.

The clouds of Venus move in the same direction as the planet's spin but much faster: whereas the solid body of the planet rotates once every 243 days, the clouds at the top of the atmosphere travel around the planet once every four days. The television pictures showed three distinct layers of haze above the opaque clouds. At least one deeper layer of clouds was suggested by the monitoring of radio signals from the spacecraft at two different frequencies. An accumulation of atomic oxygen in the upper atmosphere demonstrates that there is little vertical mixing between the layers.

A detailed profile of the temperature of the cloud tops was provided by an infrared radiometer operating at wavelengths between 35 and 55 microns. (This band of wavelengths is absorbed by the earth's atmosphere and thus cannot be used for ground-based observations.) Even though the solid body of Venus rotates very slowly, there is so much heat stored in the atmosphere that the night side of the planet is as hot as the day side. Moreover, there is little difference between cloud temperature at the equator and at the poles.

An ultraviolet spectrometer detected the glow of hydrogen in the upper atmosphere but no deuterium, the heavy isotope of hydrogen. If the hydrogen comes from the breakdown of water on the planet, one would expect to find an upper atmosphere enriched in deuterium, since the heavier deuterium could not escape into space as readily as the lighter

ordinary hydrogen. If, on the other hand, Venus receives its hydrogen from the solar "wind" impinging on its upper atmosphere, no deuterium should be evident, since the sun "cooks" deuterium into helium in its core. The observation seems to support the hypothesis that Venus formed out of dust and gas so near the sun that most of its water was driven off at early stages.

The solar wind can impinge directly on Venus' atmosphere because unlike the earth the planet has virtually no magnetic field to ward it off. A magnetometer aboard *Mariner 10* indicated that if Venus does have a magnetic field, it can be no stronger than a twentieth of 1 percent of the earth's. This observation lends support to the hypothesis that the earth's magnetic field is a by-product of its relatively fast spin.

Venus also appears to be extraordinarily round, with a hundredth the flattening at its poles that the earth has. If it ever did rotate more rapidly, either it did so when it was a plastic mass or it is less rigid than the earth and was capable of returning to a more spherical shape when it slowed down.

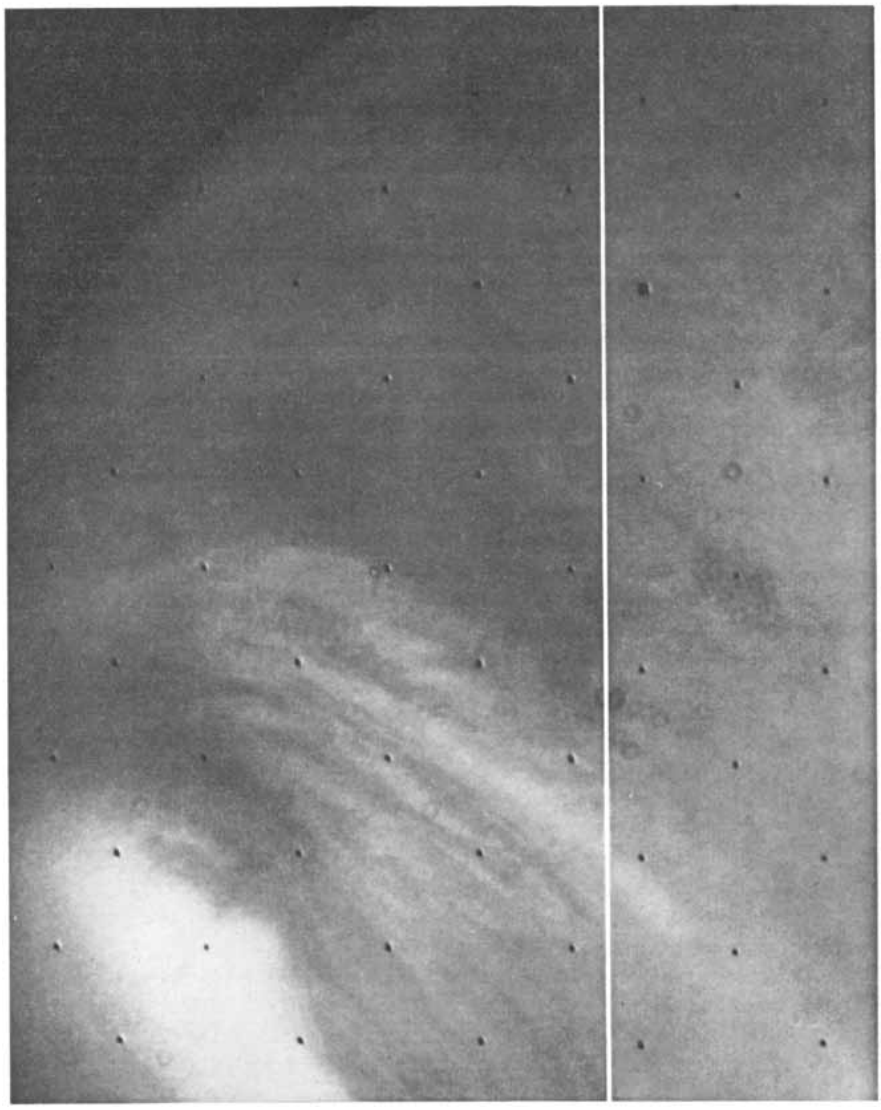
Herbicidal Warfare

Between 1962 and 1971 U.S. forces sprayed 20 million gallons of herbicidal chemicals on some four million acres of South Vietnam, or about a tenth of the country's land surface. The objectives were primarily to defoliate trees and thus uncover enemy troop concentrations and to destroy crops suspected of being destined for enemy stomachs. The environmental war has now been assessed by a committee of the National Academy of Sciences. Among the committee's conclusions are that serious long-term damage was done to Vietnamese forests, that the fertility of cropland was probably not adversely affected and that it is too soon to be sure whether or not humans were made sick or genetically damaged by the herbicides, or will be in the future.

The study was commissioned in 1970 by the Department of Defense, responding to a Congressional mandate. The National Academy formed a committee of 17 scientists from the U.S., South Vietnam, Great Britain, Sweden and Canada, none of whom had been known as activist opponents of the herbicide program. The chairman was Anton Lang, director of the Atomic Energy Commission's Plant Research Laboratory at Michigan State University. The members resolved to emphasize quantitative data rather than anecdotal reports and speculation; the result is a cautious report that makes few broad pronouncements but that documents a story of widespread devastation and frames significant questions for future research.

The herbicides sprayed over Vietnam were chemicals that are routinely applied for weed control: 2,4-D and 2,4,-5-T, picloram and cacodylic acid. The concentrations, however, were generally from two to 10 times as high as for civilian purposes. Military records indicate that 88 percent of all herbicide missions were for defoliation, 9 percent for crop destruction and 3 percent for other purposes, such as clearing areas around fire bases. About three-fourths of the total amount sprayed was delivered over inland forests, about 8 percent on mangrove forests along waterways and about 7 percent over cultivated areas (not counting temporarily cleared cropland in the highlands, which was classed as forest).

The ecological province most seriously affected by herbicides was the coastal and riverine mangrove forest, some 36 percent of which was subjected to spraying. One spray mission was usually enough to kill all mangrove trees; in extensive areas there has been no reconstitution of the vegetation because there are no sources of seed. The committee estimates that "it may take well over 100 years for the mangrove area to be reforested" unless a massive reforestation program is undertaken, which might reduce the time to about 20 years "if sufficient money and seed resources were available." Since the mangrove forest



SOUTHERN HEMISPHERE OF VENUS was photographed at ultraviolet wavelengths with a television camera aboard *Mariner 10*. The picture, which is actually a mosaic of two (*separated by the white line*), was made at 17:15 universal time on February 6 from a distance of 450,000 miles one day after the spacecraft's closest approach to the planet. Spiral markings, visible only at ultraviolet wavelengths, seem to indicate a smooth streamlined flow of air that originates at the equator and spirals toward the pole. Dots are fiduciary marks.

serves as a spawning and feeding ground for many fish and shellfish, its destruction has important economic consequences. In addition mosquitoes—and with them malaria—have proliferated in the devastated mangrove areas, and so have rats.

As for the inland hardwood forest, some 10 percent of which was subjected to one spraying or more, the assessment is more equivocal. The committee took great pains to measure the loss of "merchantable timber" as the only readily quantifiable criterion of damage. The figure arrived at was substantially lower than earlier estimates and apparently occasioned dissension within the committee; the ultimate majority conclusion was challenged by two members as being much too low. The report makes it clear, however, that the damage to inland forests has been "widespread and serious." The long-term question, which the report leaves unanswered, is to what extent the destruction of small or otherwise "nonmerchantable" trees has so disturbed the forest succession that bamboo and scrub growth will replace valuable trees.

The committee found no hard evidence of human illness or of congenital malformations attributable to the herbicides, although 2,4,5-T in particular, and its contaminant, dioxin, have been seriously implicated in animal studies. On the other hand, the report cites repeated stories, primarily from residents of the highland forest regions, of sickness and death following spraying of their hamlets and fields; these stories are so consistent, the committee states, that further investigation is required. As for congenital effects, the report also urges further study of hospital records to establish any correlation between spraying and birth defects. In addition it calls for careful monitoring of dioxin residues in the food chain to make sure they are not becoming concentrated to dangerous levels.

The Gears of Ancient Greece

In 1901 the remains of an intriguing metal object were found in an ancient shipwreck near the small island of Antikythera between Greece and Crete. In the corroded and encrusted fragments, recovered on the first underwater archaeological expedition, gears were discovered, apparently arranged in a clockwork mechanism. The remains of the device, which came to be known as the Antikythera mechanism, were placed in the Greek National Archaeological Museum in Athens.

By the 1950's it was apparent that the mechanism was an astronomical calendar or computer and that it had been built in the first century B.C. That a geared machine of any kind was constructed in that era was considered surprising; more recent work has shown that it was in fact a sophisticated machine, in both design and workmanship. As a result our estimate of the state of technology in the Hellenistic era may have to be revised.

Since 1950 much of the work of analysis and interpretation of the Antikythera mechanism has been done by Derek de Solla Price of Yale University [see "An Ancient Greek Computer," by Derek J. de Solla Price; SCIENTIFIC AMERICAN, June, 1959]. In 1972 Price was able for the first time to make X-ray and gamma-radiography studies of the fragments, and it is primarily through these techniques that many details of the construction and operation of the device have recently been revealed. Price's findings will be presented at the April meeting of the American Philosophical Society.

Although some parts of the theoretical reconstruction remain tentative, the main elements of the mechanism and their functions have been determined. The mechanism was apparently enclosed in a box about 12 inches by six by three, with hinged panels bearing inscriptions on the large faces. Inside were at least 30 gears, all made of bronze and all probably cut from a single sheet of metal. The input to the machine was through a small crown, or contrate, gear whose shaft extended through the side of the case. The power was transmitted to a large gear that had four spokes mortised into the rim, soldered and fixed in place by rivets.

The gears apparently drove pointers (which have not been found) on three large dials, each of which has a number of scales with rotating slip rings, and perhaps on two smaller, subsidiary dials as well. The information displayed was calendrical: the positions of the sun and moon as calculated by the 19-year cyclical calendar current in classical Greece. From these positions a third kind of information was derived by "the most spectacular mechanical feature of the Antikythera device." This feature was a differential turntable, a mechanism that is not attested to elsewhere in antiquity and that did not reappear in Europe until the 16th century.

Although a piece is missing from the gear train, the function of the turntable has been determined with virtual certainty. It appears that it subtracted the sidereal motion of the sun from that of the moon to produce the synodic month, the cycle of the phases of the moon.

The main cargo of the wrecked ship on which the Antikythera mechanism was found was apparently statuary, probably being transported from Rhodes or Cos to Rome. The mechanism itself was almost certainly part of the cargo, an *objet d'art*. It may have been intended as an exhibition piece, to be mounted on a pedestal, or perhaps held in the arms of a statue and driven by a water clock. An alternative possibility is that it may have been a portable instrument, a calendrical calculator, operated by a folding crank. Whatever the mode of operation, there is evidence that the mechanism was actually used (and therefore that it worked): two of the gears show marks of having been repaired.

The artifacts with which the Antikythera mechanism was found, the orthography of its inscriptions, its metallurgy and the context of astronomical knowledge in which it must have been used all indicate that the mechanism is of ancient origin, probably manufactured about 80 B.C. The mechanism itself may provide a more precise date; a mark on one of the dials may have been made to indicate the position of the sun in the zodiac at the time the dial was engraved. If it was, then the mechanism was almost certainly fabricated in 87 B.C.

It is the virtually certain antiquity of the mechanism that is its most intriguing and most important aspect. No similar relics of the period have been found; moreover, no simpler mechanisms that could have served as predecessors or models are known. Yet the sophistication of the Antikythera mechanism precludes the possibility that it could have been the unprecedented invention of a single man, however great his genius.

The only acceptable inference, Price concludes, is that a tradition of high technology did flourish in the Hellenistic world. There are literary references to support this view—Cicero, for example, describes a device as complicated as the Antikythera mechanism, and that may in fact have been the Antikythera mechanism—but without surviving artifacts they have rarely been credited. "A technological tradition," Price writes, "is something so much more fragile than anything that was encoded into a written book and transmitted in the orderly fashion of knowledge."

Chimeric Cats

The study of genetic mechanisms is facilitated when a sex-linked chromosomal variation manifests itself in a form that is readily perceived, for example in the color of an animal's hair. Such is the

case with tortoiseshell and calico cats. "Tortoiseshell" is a mixture of orange and black, usually blended together rather than in large patches; "calico" is tortoiseshell and white, usually in patches. (The "orange" ranges from cream and yellow to dark orange, and "black" covers gray, blue, brown, tabby and true black.) Most tortoiseshell and calico cats are female. Males are rare and tend to be infertile. It now appears that the existence of most, and perhaps all, tortoiseshell and calico male cats can be explained by abnormalities in their sex-chromosome complement.

Normal female cats have 38 chromosomes, including two *X* chromosomes. Tortoiseshell female cats are thought to have a gene for orange hair color on one *X* chromosome and a gene for non-orange on the other. Early in embryonic life one *X* chromosome or the other becomes active in each cell. The result is a random mixture of cell groups, some with the orange gene active and some with the non-orange one active. The genes for black and for white are believed to be on autosomal (nonsex) chromosomes. If the orange gene is active, it will mask the appearance of black. If the non-orange gene is active, it allows the autosomal black gene to manifest itself. The resulting coat color is a mixture of orange and black. If white genes are present, they will be dominant and mask the other colors, resulting in patches of white as well as orange and black.

Since normal male cats have an *XY* sex-chromosome complement and thus only one *X* chromosome, they cannot have such hair-color mixtures. Recent investigations by Willard R. Centerwall of Loma Linda University and Kurt Benirschke of the University of California at San Diego show that when a male cat is tortoiseshell or calico, the explanation is that it has more than one *X* chromosome. Writing in *The Journal of Heredity,* they report that when the chromosomes of 25 tortoiseshell and calico male cats were analyzed, all 25 were found to have chimeric, mosaic or trisomic sex-chromosome abnormalities. Chimerics result from the early embryonic fusion of what would have been two separate individuals. Male chimerics could be either *XX/XY* or *XY/XY,* and if two different *X* chromosomes are present, both orange and black could be manifested in the coat color. Such chimeric males could be fertile. The trisomics are animals with an *XXY* chromosome complement, which is strongly associated with sterility in most mammals, including man. Centerwall and Benirschke also found mosaics with an *XY/XXY* chromosome complement and chimeric-mosaic combinations (for example *XX/XY/XXY/XXYY*).

Hanging in There

It is well known that if one is obliged to perform a monotonous task, such as monitoring a radar screen or driving a car, one's attention deteriorates with the passage of time. Many efforts have been made to correlate the loss of attention with changes in the electroencephalogram (EEG). The only consistent finding is that when performance drops, owing to fatigue or monotony, there is an increase in the volume of theta waves from the occipital (visual) region of the brain. Theta waves have a frequency of three to seven cycles per second. (Alpha waves have a frequency of eight to 12 cycles and beta waves 13 to 30 cycles.)

It has been demonstrated in recent years that unconscious activities such as brain rhythms can be altered by operant conditioning, which involves supplying the subject with a tangible signal (for example an audible tone) driven by the activity one is trying to alter. It occurred to Jackson Beatty of the University of California at Los Angeles that if a person could be trained to suppress his brain's theta activity while he was simultaneously engaged in a monotonous task, the usual deterioration in performance might be prevented. Beatty and his colleagues report in *Science* that suppression of the theta rhythm indeed sustains alertness during performance of a dull task. His colleagues were Arana Greenberg, also of U.C.L.A., and W. Phillip Deibler and James F. O'Hanlon of Human Factors Research, Inc.

The task selected by Beatty was the familiar one of detecting "targets" simulating aircraft on the cathode ray screen of a radar plan-position indicator, similar to the screens monitored by air-traffic controllers at most airports. Targets, painted by an electron beam rotating at 10 revolutions per minute, appeared at random times and locations against a typical "noisy" background. Targets were repainted on successive sweeps until detected by the observer. In the first half-hour of a two-hour experiment a target was usually spotted in fewer than six sweeps (36 seconds). By the third half-hour the time that elapsed before detection often increased by 30 or 40 percent. In the last half-hour subjects typically showed an "end spurt" in which detection time improved almost to the original level.

Beatty and his co-workers trained 12 volunteers to suppress their theta activity by operant conditioning and seven to increase their theta activity. The subjects were then retested on the radar-monitoring task under two conditions: one in which they concurrently regulated their theta activity (either up or down according to their previous conditioning) and one in which they performed the task with their theta activity unregulated.

In the unregulated condition both groups exhibited the usual decline in attention for the first three-quarters of the two-hour test, followed by a final spurt in the last quarter. Under the theta-regulated condition the subjects who had been taught to increase their theta activity showed a continuous deterioration in performance, so that in the last half-hour they required, on the average, more than a minute to detect a target. The group that had been trained to suppress theta activity showed virtually no loss in vigilance over the entire two hours, rarely requiring more than 36 seconds to detect a target. "This is the first demonstration, to our knowledge," the authors write, "of a lawful relationship between operantly regulated cortical phenomena and performance in man."

Old Story

Acupuncture, the ancient therapy of Asia, is widely viewed as being new to the Western world. Not so, comments an editorial in *British Medical Journal.* The article cites a report by a medical officer of the East India Company who observed the use of acupuncture in Japan in the 1670's; application of the technique by a British surgeon, William Coley, in treating an infant in 1797, and the widespread practice of acupuncture in France and England in the 19th century.

The principal complaint for which acupuncture was prescribed early in the 19th century was distention caused by gas or body fluids. As the decades passed the procedure also came into use for relieving the pain of sciatica and muscle spasm. Exactly why it brought relief was not understood, although one practitioner in the 1870's suggested that the insertion of the needles increased the circulation of blood in "impoverished tissues."

At the end of the 19th century the increasing use of acupuncture by quacks and the attribution of its benefits to such phenomena as the "conductive" needles' ability to draw off the "electric fluid" produced by "animal magnetism" caused British physicians to abandon the procedure. Only in France did a number of acupuncturists continue in practice. They do so to this day, largely because many of their patients are former residents of French territories in the Far East.

WOOD PULP

This major commodity consists of wood fibers separated mechanically, thermally or chemically. It is mainly used to make paper, but it is also the source of such chemical products as rayon and acetate film

by F. Keith Hall

It is difficult to imagine how a modern industrial society could function without paper, which in the U.S. is consumed at a rate of 640 pounds per person per year. It is scarcely less difficult to imagine how the supply of paper for such a society could be maintained without wood pulp. Therefore one may expect the pulp, paper and paperboard industry to be a substantial factor in any developed economy. In the U.S. last year the industry employed 700,000 people, whose output in terms of products sold was valued at $26 billion. On a worldwide basis the industry produces annually some 127 million tons of wood pulp, of which the U.S. accounts for 47 million tons.

Paper does not have to be made from wood pulp; it has been and is made from rags, straw, cotton linters, bagasse (the fiber of the sugarcane plant) and flax. Wood, however, is a far more plentiful and less expensive source of the fibers that paper and paperboard are made from. Wood consists primarily of cellulose, hemicellulose and lignin. Cellulose is a crystalline linear polymer of glucose, that is, it is a polysaccharide whose molecules consist of glucose units strung together in long chains that tend to lie parallel to one another in a regular array. Hemicellulose is the various polysaccharide polymers of wood other than cellulose. Lignin is an amorphous polymer, consisting of aromatic (benzene ring) units, that serves as the cement holding the fibers of wood together.

Softwood, which is the term arbitrarily applied to the wood of coniferous, needle-bearing trees, consists of about 43 percent cellulose, 28 percent hemicellulose and 29 percent lignin. Hardwood, which is the term arbitrarily applied to the wood of broad-leafed trees, consists of 43 percent cellulose, 35 percent hemicellulose and 22 percent lignin. (Both types of wood vary enormously in hardness and density from species to species.) More pulp is produced from softwood than from hardwood; the ratio is roughly 4 : 1.

Pulping is the process by which wood is reduced to a fibrous mass. Put another way, it is the means of rupturing the bonds between the fibers of wood. The task can be accomplished mechanically, thermally or chemically. Although pulp can be obtained in many ways, the commercial processes are generally classified as mechanical, chemical or semichemical.

Mechanical Pulping

Mechanical pulping is frequently called the groundwood process because in its original form it entails literally grinding the fibers out of the wood. A debarked log (bark is a waste product in pulping, since it reduces the quality of pulp) is pressed lengthwise against a roughened grinding stone that is revolving at a speed of from 3,500 to 5,000 feet per minute at its outside edge. Fibers are thereby torn out of the wood and rubbed down to a fine powder, which is washed away from the stone with water. The thin slurry of fibers and fiber fragments thus obtained is put through screens to remove slivers and other oversize particles and then is thickened by the removal of water to form a pulp suitable for a papermaking process. The principle is simple, but the efficient production of a uniform pulp of good quality requires careful control of such variables as the volume and temperature of the water, the pressure on the stone and the surface roughness of the stone.

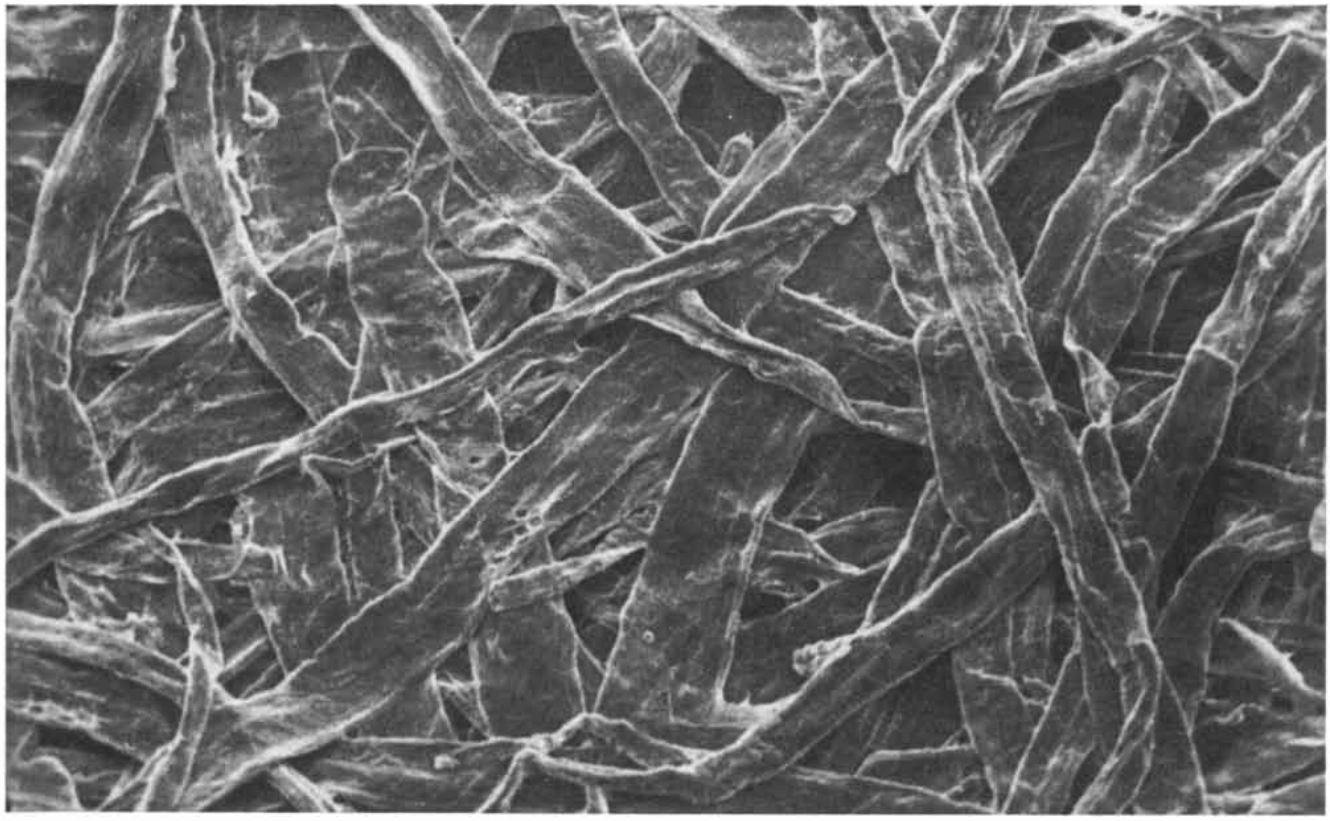
An entirely different method of producing groundwood pulp involves shredding and grinding chips of wood between metal shearing disks in a machine called a refiner; the product is known as refiner groundwood. Refiner groundwood is generally preferred over stone groundwood in the industry because it contains more long fibers and yields a stronger paper. Sometimes the wood is thermally or chemically presoftened to reduce the power needed for grinding.

Groundwood processes have the advantage of converting about 95 percent of the dry weight of the wood into pulp. They also yield a paper that is highly opaque. On the other hand, they have the disadvantage of producing a relatively weak paper and one that discolors easily on exposure to light. To increase the strength it is customary to add long-fibered chemical pulp to groundwood pulp. Newspapers the world over are printed on paper that is made from about 70 to 85 percent groundwood pulp and 15 to 30 percent chemical pulp. For the most part softwood is the source of groundwood pulp in North America. The much smaller, thin fibers of hardwood are severely damaged by grinding and yield a fine, flourlike material that forms an exceedingly weak sheet.

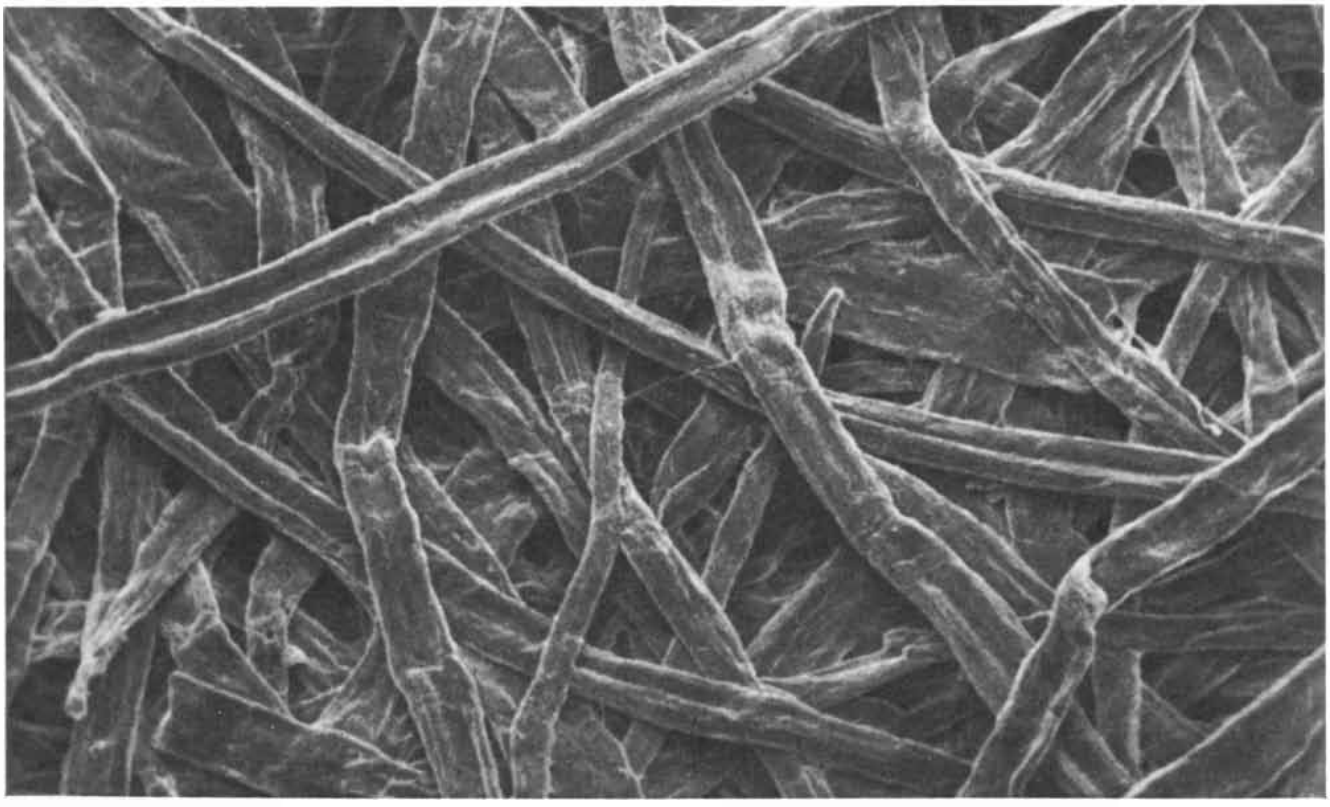
Chemical Pulping

The principle underlying chemical pulping is that the lignin in wood can be degraded and dissolved by various chemical reagents, leaving most of the cellulose and hemicellulose in the form of fibers. Chemical pulping always entails degrading a certain amount of the hemicellulose and cellulose as well as nearly all the lignin, however, so that the yield of pulp is markedly lower than the yield from the groundwood method. Nonetheless, a chemical pulping process on a commercial scale can be expected to yield as pulp about 40 to 50 percent of the weight of the wood.

In chemical pulping the debarked log



WOOD PULP made from southern pine appears at an enlargement of 250 diameters in a micrograph made with a scanning electron microscope. This pulp was made by the kraft process, in which the lignin that holds the cellulose fibers of wood in place is dissolved chemically in an alkaline liquor. The fibers came mainly from wood that grew in the spring and therefore grew rapidly. Such wood has cells with thin walls, so that after pulping the fibers collapse readily into ribbons and form a dense and well-bonded sheet.



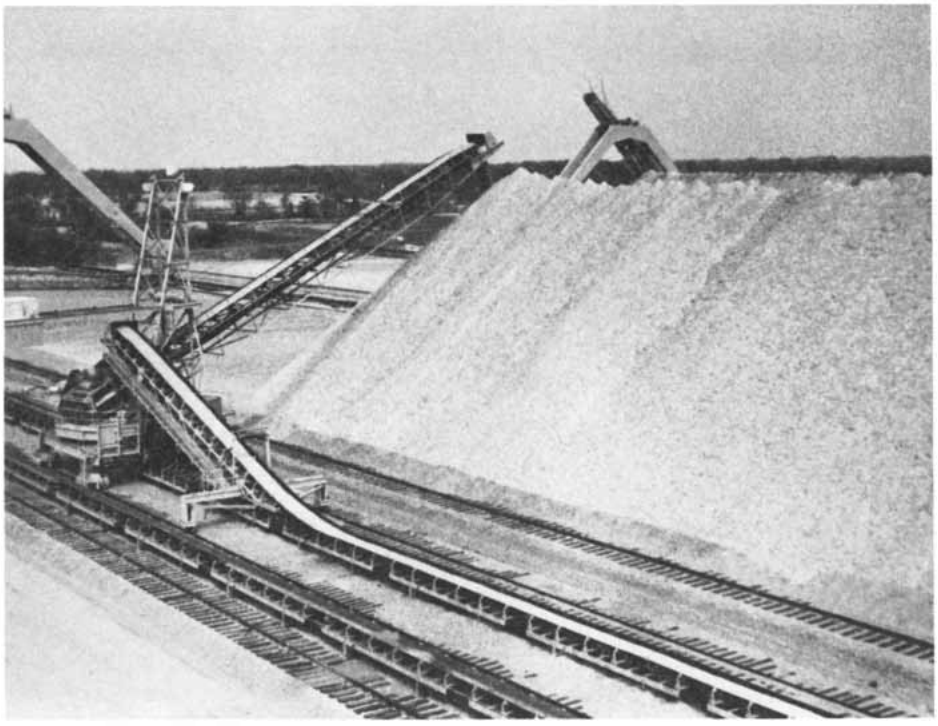
KRAFT PULP made primarily from the slow-growing summerwood of southern pine appears in a scanning electron micrograph at the same enlargement as in the micrograph at the top of the page. Here the fibers resist collapse and form a relatively open sheet. All scanning electron micrographs accompanying this article were made by R. A. Parham of the Institute of Paper Chemistry.

is first put through a chipper to reduce the wood to chips ranging in size from 3/8 inch to one inch in length, 5/8 to 7/8 inch in width and 1/16 to 3/16 inch in thickness. The chips are cooked with the appropriate chemicals in an aqueous solution, usually at an elevated pressure and an elevated temperature. The two chief methods of chemical pulping are the kraft process and the sulfite process.

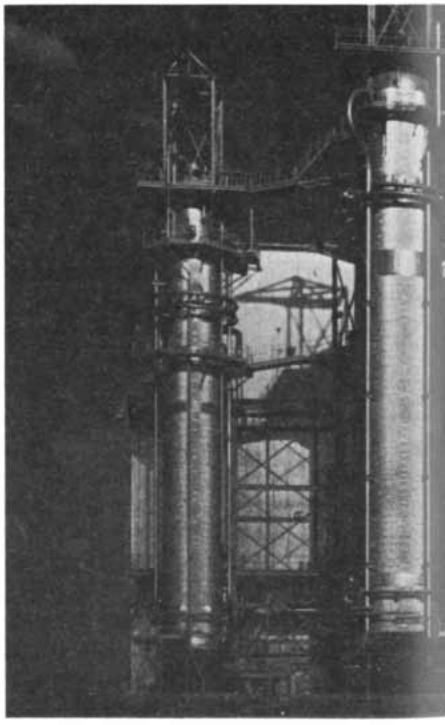
Of the two the kraft process has come to occupy the dominant position since its fortuitous discovery in Germany 95 years ago by Carl S. Dahl, who in attempting to reduce the cost of recovering the chemicals from the soda pulping process then in use introduced sodium sulfate into the recovery system. The pulp quality improved adventitiously as a result of the formation of sodium sulfide from sulfate during combustion.

PINE TREES ready to be cut for pulping stand on a plantation of the International Paper Company in Georgia. They are genetically improved slash pines, ready for harvesting at age 25.

RIVER SHIPPING conveys felled trees, which have been cut into logs,



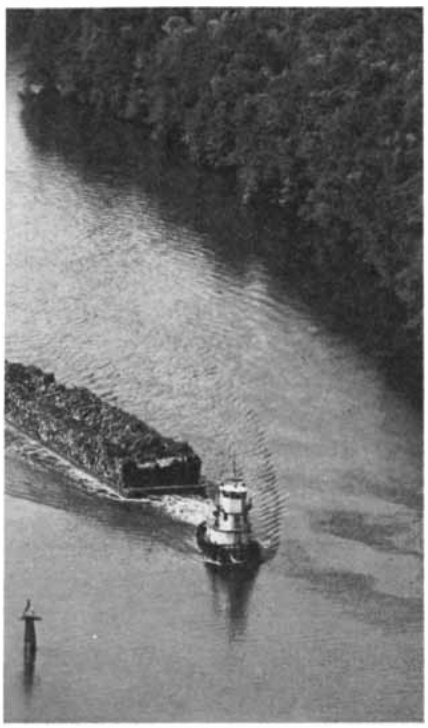
WOOD CHIPS are the form in which chemical pulping is done. Debarked logs are put into a chipper, which reduces them to the material seen here being moved from chipper to storage pile.



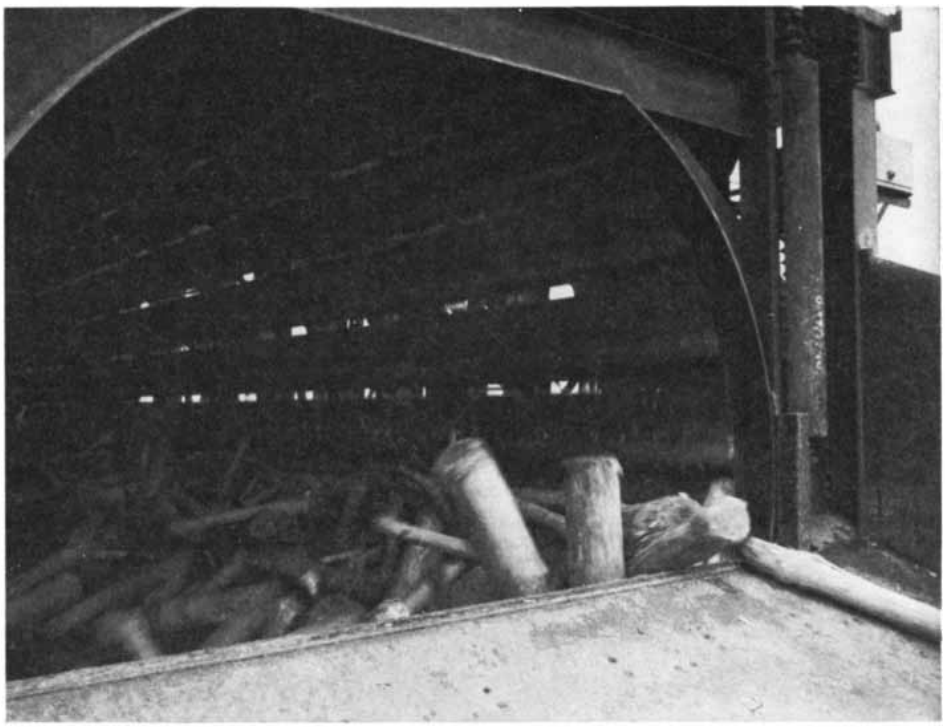
CONTINUOUS DIGESTERS accept wood chips at top and cook them in

Most soda pulp mills changed over to the kraft process (which is also often called the sulfate process). The process has evolved into the sophisticated and efficient kraft mills of today.

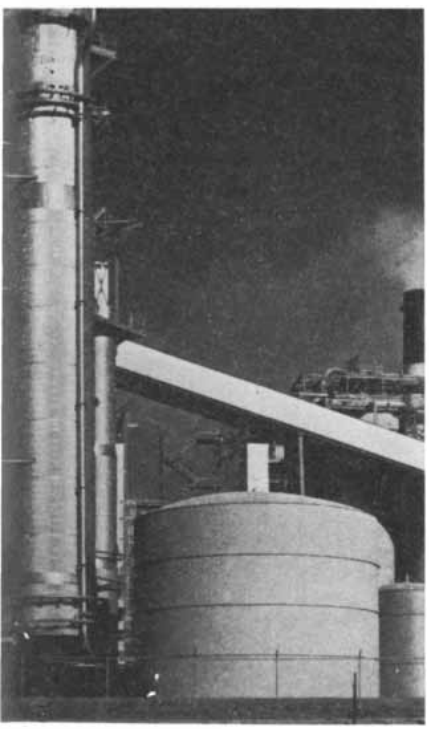
The kraft process involves cooking the wood chips in a solution of sodium hydroxide, sodium carbonate and sodium sulfide. Heating in a pressure vessel called a digester to the maximum temperature of from 160 to 180 degrees Celsius takes from one hour to two and a half hours, and the mixture is held at that temperature for another hour or two. By then most of the lignin has been degraded and dissolved out, although the wood still has the physical form of chips. At the end of cooking the chips are expelled into a blow tank, where a rapid drop in pressure breaks up or defiberizes them and reduces them to



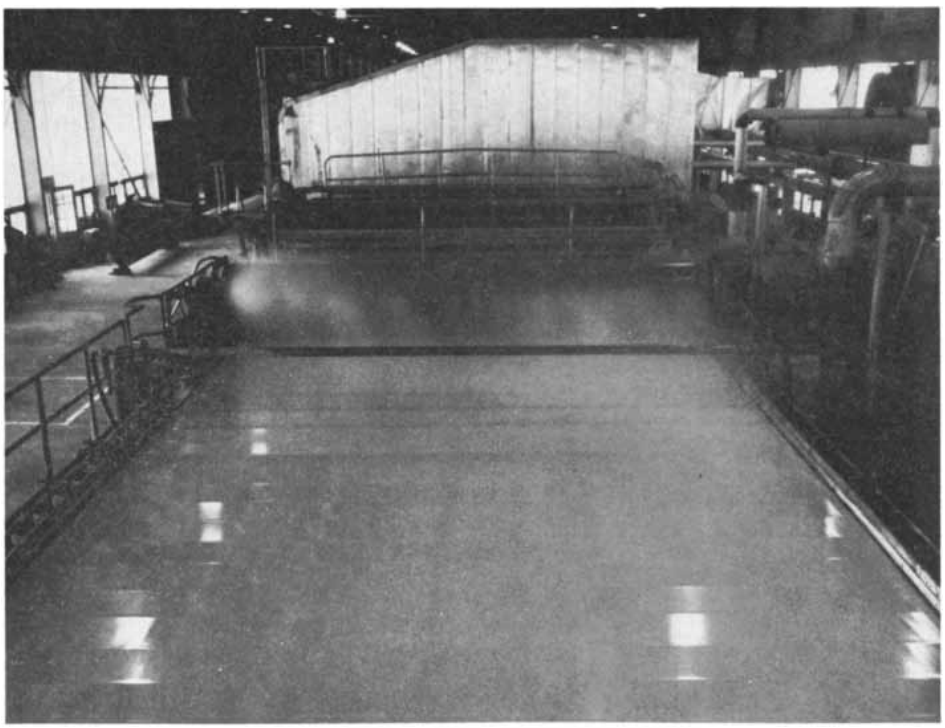
to the pulp mill. Each of the three barges transports 400 cords of wood.



DEBARKING DRUM removes the bark from logs by revolving and causing them to strike one another repeatedly. The bark, which is a waste product in pulping, falls between openings of drum.



an alkaline or acid solution to dissolve lignin out of the wood fibers.



SLURRY OF PULP flows onto a machine for first stage of papermaking. Material is 99.5 percent water. As water is removed, self-bonding wood fibers adhere to form a continuous sheet of paper.

pulp. In kraft pulping the alkaline attack damages the lignin so severely that much of the original aromatic ring structure is destroyed.

Kraft pulps yield a strong paper; indeed, the kraft process takes its name from the German word for strength. They also emerge in a dark brown color that is not easily bleached.

An environmental drawback to kraft pulping is that the process gives rise to gases that contain malodorous substances such as mercaptans and organic sulfides. The various gases from the process are condensed to recover turpentine as an important by-product of softwood cooking. The noncondensable portion of the gas is often treated further to reduce the odor and remove particulates.

Several different grades of kraft pulp are made for different end products. Unbleached grades, which make packaging material, contain more lignin and have a higher yield from wood than pulps destined to be bleached into grades that can be made into white paper.

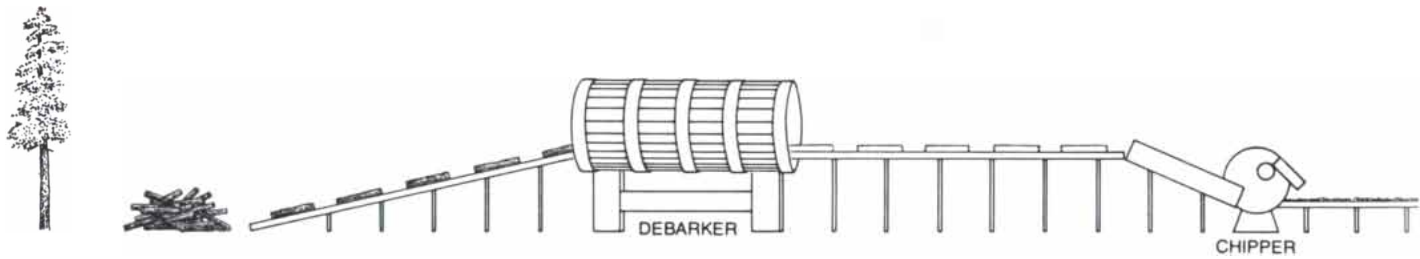
In the sulfite process an acidic mixture of sulfur dioxide, water and a chemical base material is the means of attacking and solubilizing the lignin. Here the mechanism of chemical attack removes the lignin as salts of lignosulfonic acid, and the aromatic ring structure is left largely intact. The chemical base can be ionic calcium, magnesium, sodium or ammonium. The solution also contains sulfurous acid and sulfite and bisulfite ions. Usually the cycle consists of a cooking period of from six to eight hours, during which the temperature rises to 140 degrees C.

Sulfite pulping works well for spruce, fir and hemlock. The so-called soft hardwoods such as the poplar and some species of eucalyptus also pulp well in the sulfite process. Resinous softwoods such as pine and tannin-containing hardwoods such as oak are much more difficult to handle in this process. Sulfite pulps are fairly light in color and can be bleached easily, so that they serve well in making fine papers, although the sheets are weaker than the equivalent sheets of kraft pulp. It is relatively difficult to recover the cooking chemicals from the sulfite process, which is one reason the process has lost ground to kraft pulping after having been dominant in the industry until about 1930. The other major reason for the decline in sulfite pulping has been the sensitivity of the process to variations in the species of wood used, since it has become increasingly difficult to supply a mill with only one or two selected species of wood.

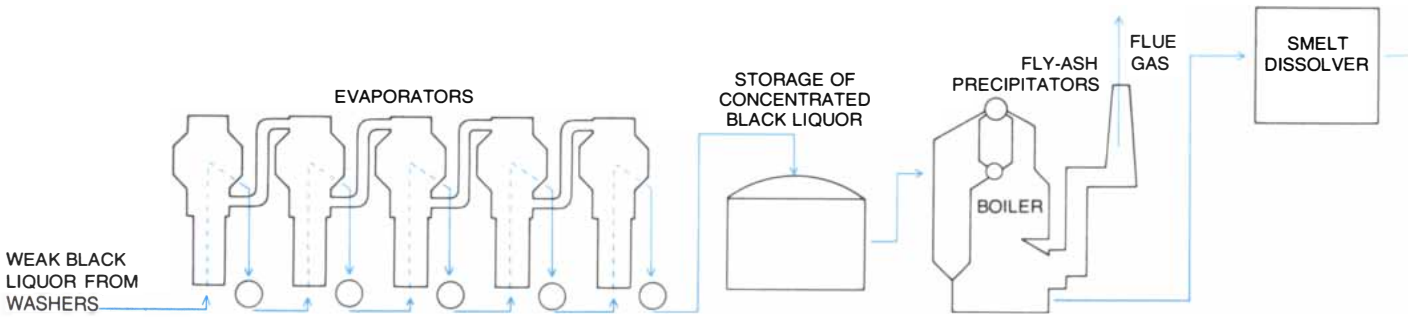
In view of the fact that the groundwood process yields weaker pulp than the chemical processes and also requires large amounts of power, the notion of combining chemical and mechanical treatment had a natural appeal for the industry. In pursuit of this approach the industry has developed the group of related processes called semichemical pulping. Essentially they involve softening wood chips or logs with such chemicals as a neutral solution of sodium sulfite, an alkaline solution of sulfide or an acid solution of sulfite. The remainder of the pulping action is supplied mechanically, most often in disk refiners. Pulp yields range from 65 to 90 percent of the weight of the wood. Semichemical processes are applied mainly to hardwoods because of the lower content of lignin in such wood. The pulps serve well in making magazine paper, newsprint and the center layer in corrugated board for containers.

Semichemical Pulping

The early efforts to develop semichemical pulping included the introduction of chemicals during grinding. That technique is in fact best regarded as a



CHEMICAL PULPING by the kraft process is depicted. Felled trees are cut into logs for convenient handling. At the pulping plant the logs are fed into a machine that removes the bark. The debarked logs proceed to a chipper, which reduces the wood to small chips. The chips are loaded into a digester, in which they are cooked for from two to four hours in an aqueous solution of sodium



RECOVERY OF CHEMICALS is an important part of a kraft pulping operation. Weak black liquor (*color*) that is separated from the wood pulp after cooking and washing is concentrated by a series of evaporators, becoming a thick liquid consisting of lignin fragments, hemicellulose and cellulose removed from the wood and of chemicals employed in pulping. Concentrated black liquor serves as fuel in a boiler from which the smelt, or molten organic chemical, is removed for further processing. Smelt is dissolved in water to form

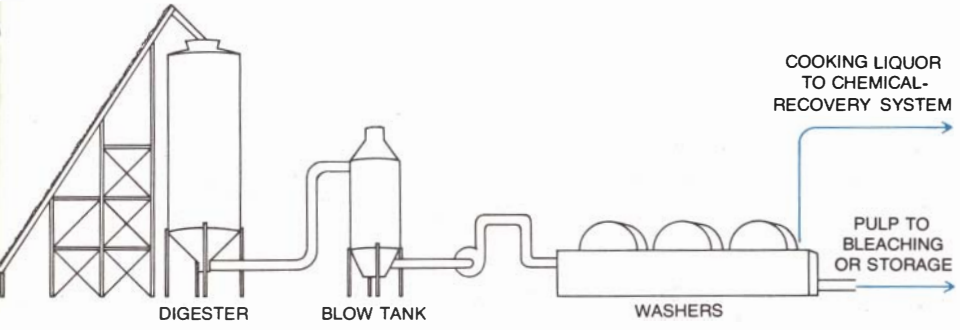
modification of the normal groundwood process. It still finds applications, as when alum is employed to minimize problems of stickiness when species of pine with a high resin content are pulped.

Another early development in semichemical pulping involved pretreatment of the entire log with chemicals. The results were rather poor until the "chemigroundwood" process was developed for hardwoods. In this process whole logs are dumped in a digester, from which the air is then removed. Cooking liquor containing sodium sulfite buffered to a *p*H of 9.5 (that is, made moderately alkaline) is introduced, and the wood is cooked at a temperature of from 135 to 150 degrees C. for five or six hours. When the wood is ground following this treatment, only about half the normal grinding energy is required. The yield of pulp is from 85 to 90 percent of the weight of the wood, and the quality of the pulp is greatly improved over the pulp resulting from stone grinding or refiner grinding. Chemigroundwood from hardwood is also considerably stronger than normal spruce groundwood. Among the products made from chemigroundwood pulp are newsprint, tissues, toweling and paper for books and magazines.

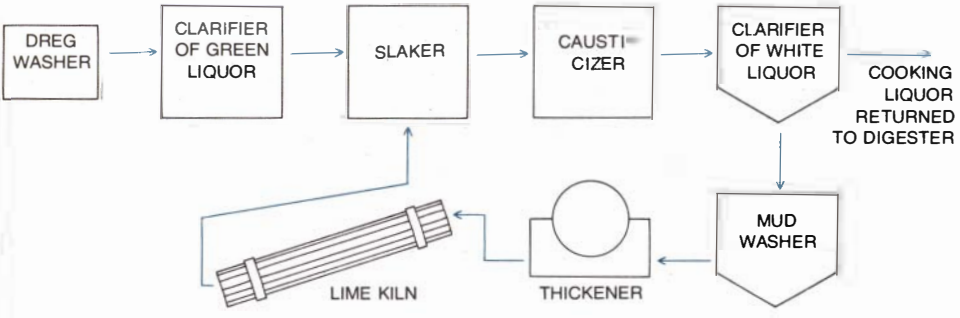
The group of semichemical pulping processes includes the so-called high-yield kraft and sulfite systems, which are designed to improve the ratio of pulp to wood. They work with both softwoods and hardwoods. The most important of the new high-yield systems are the cold-caustic process and the neutral-sulfite process.

The cold-caustic process, which takes its name from the fact that it employs caustic soda (sodium hydroxide) at low temperatures, is valued for its swelling power rather than for its delignifying action. Hardwood swollen in 10 percent caustic soda loses from 70 to 80 percent of its strength and therefore can be defiberized easily in a mechanical pulping operation. Yields are good in the cold-caustic process, and operating costs are lower than in other semichemical processes. All cold-caustic pulps are inferior to kraft pulps in physical properties but are stronger than softwood groundwood pulps.

In the neutral-sulfite process, the most successful of the semichemical pulping operations, hardwood chips are impregnated under pressure with the cooking liquor. The cooking chemicals are almost neutral at the start and are buffered to maintain neutrality as considerable quantities of organic acids are liberated from the wood during cooking. The cooked pulp is defiberized mechanically with refiners. This pulp serves mainly in making the fluted middle layer of corrugated board, where its resistance to crushing is important.



hydroxide and sodium sulfide. The cooking dissolves most of the lignin in the wood. After cooking the chips proceed to a blow tank, where a drop in pressure reduces them to fibers. After a washing and screening step the pulp is ready to be put into a papermaking process.



green liquor (sodium carbonate and sodium sulfide), clarified and converted to white liquor (sodium hydroxide and sodium sulfide) by treatment with slaked lime. Calcium carbonate is precipitated as lime mud, removed and heated in a kiln to regenerate quicklime, which is slaked for reuse. Other steps, not shown, recover such materials as turpentine and rosins.

Dissolving Pulp

Another pulping variation, which has given rise to a substantial industry in its own right, comes under the heading of dissolving pulp. In the dissolving-pulp industry pulps made by either the kraft process or the sulfite process are further purified chemically to remove all hemicellulose and leave only pure cellulose, which can then be transformed into such products as viscose rayon, cellophane and acetate fibers and film. The largest single use for dissolving pulp is as the raw material for the viscose-rayon process, which takes pure cellulose, chemically converts it into soluble sodium cellulose xanthate (called viscose) and then regenerates pure cellulose again as a continuous filament or a staple fiber by decomposing the extruded liquid xanthate in sulfuric acid.

The properties of the regenerated cellulose fibers can be varied widely because of the chemical flexibility of the viscose process. Rayon in continuous filaments of high tenacity serves as tire cord; in spite of severe competition from nylon and (more recently) from polyester, glass and steel, it has maintained a significant position in the tire-cord market. Cellophane is made from the same process as rayon except that the viscose is extruded through a wide slot in the form of a film rather than through small holes in the form of filaments.

Cellulose converted into the acetyl ester can be spun from an acetone or a methylene chloride dope to yield two distinct fibers: cellulose acetate and cellulose triacetate. Cellulose nitrate, one of the first synthetic fibers, is still produced in large quantities as an explosive and a propellant. Cellulose acetate butyrate is the material often found in automobile steering wheels and in screwdriver handles.

The sodium salt of carboxymethyl cellulose serves as a thickener in ice cream and in many other food products. It is also used in detergents, where it helps to hold in suspension the dirt the detergent has removed from clothes or dishes. In addition the ethyl and methyl ethers of cellulose are widely employed in industry as lacquers.

Continuous cellulose filaments can be pyrolyzed under controlled conditions to

yield carbon fibers of extremely high strength and stiffness. Embedded in a protective matrix, they form materials that are attractive to the aerospace industry because they are not only strong and stiff but also stable and temperature-resistant. Cellulose can also be hydrolyzed with acid under controlled conditions so that only microcrystalline particles remain. In aqueous suspension these highly structured microcrystals of pure cellulose have interesting thixotropic properties, that is, they behave as a gel until they are shaken or otherwise disturbed, whereupon they flow easily until they come to rest again. They are therefore useful in products such as hand lotion and mayonnaise, where thickening combined with ease of flow is desired.

Papermaking

Until about 1900, when the viscose-rayon industry began its period of rapid growth, the sole purpose of the various pulping processes was to provide fiber for the paper industry, which remains today by far the largest consumer of pulp. Papermaking entails forming sheets from the fibers yielded by pulping. Sheets ranging from thin tissue paper to heavy paperboard are all made in the same general way: by draining excess water from a dilute aqueous slurry of pulp fibers on a moving mesh of fine wire. The mesh is called a Fourdrinier after the brothers Henry and Sealy Fourdrinier, who developed it in England early in the 19th century after the basic idea of a continuous moving screen had been conceived by Nicolas Louis Robert, a French schoolteacher.

Initially the slurry is about 99.5 percent water and .5 percent pulp, which means that more than 200 cubic yards of water have to be removed for every ton of paper produced. Most of the water is removed while the continuous paper "web" is taking form on the Fourdrinier mesh. The sheet then passes to the second main section of the papermaking machine, the press section. Here large felted press rolls expel more water. Finally the sheet enters the dryer section, where the rest of the water is removed by evaporation.

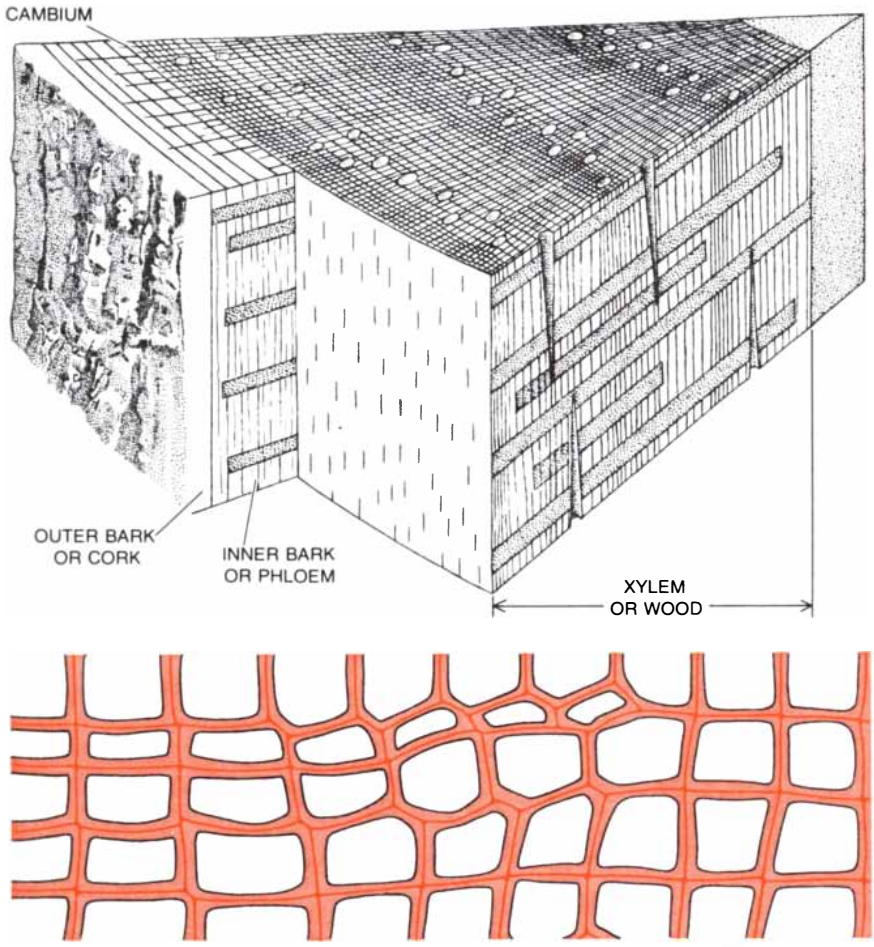
The speeds attained by modern papermaking machines are remarkable. On a newsprint machine the web can move up to 3,000 feet per minute, which is more than 34 miles per hour. For tissue paper the speed is considerably higher. A papermaking machine can be as much as 30 feet wide.

Various other operations can be performed on the papermaking machine or on separate machines later, depending on what kind of paper is wanted at the end of the manufacturing process. Such operations include adding pigment to increase the opacity of the paper and coating the paper with various substances to make it brighter or smoother or even electrically conductive. Most papers are calendered, that is, compressed and polished to develop stiffness and a final smooth finish. Finally the sheet is cut to make rolls or stacks of the shape and size required for shipment.

A substantial part of the industry is devoted to producing paperboard, the stiff and relatively heavy material that is found in such products as computer cards and the containers in which many foods and beverages are delivered to retailers. According to the American Paper Institute, the production of paper and paperboard in the U.S. in 1973 amounted to 61.9 million tons, of which paper of all grades accounted for 26.6 million tons, paperboard 29.7 million tons and construction paper and board 5.5 million tons. Insulating board was at one time made from material rejected in groundwood pulping processes; later the Masonite and Asplund processes for producing hardboard from wood chips were developed. The production of hardboard now totals more than three million tons per year.

Strong paper has great stiffness and high strength per unit of mass. It is these characteristics that can be turned to advantage in construction, provided that the sensitivity of paper to moisture is controlled in circumstances where moisture is a factor. A few years ago the International Paper Company built a 9,000-pound bridge solely from high-yield lignocellulosic papers (such as kraft liner board) and adhesive to demonstrate the merits of strong paper in structural engineering. The bridge easily supported a 12,000-pound truck.

A pulping process is markedly affected by the structure of the wood fed into it. The degree of variation in the anatomy of the wood from different species of



STRUCTURE OF WOOD is portrayed as it appears in a pie-shaped segment of a coniferous tree (*top*) and under a microscope revealing individual cells (*bottom*), with the areas having the greatest concentration of lignin shown in color. In the spring the wood grows rapidly, so that the cells are large and have thin walls. In the summer the rate of growth is slower, so that the cells are smaller and have thicker walls. They are also densely packed.

trees determines the commercial success or failure of any particular forest tract as a source of wood for pulp. The similarity in the anatomy of several kinds of northern softwood makes it possible to pulp them together in a single sulfite operation. It is difficult to pulp several kinds of tropical hardwood together even in the more versatile kraft process.

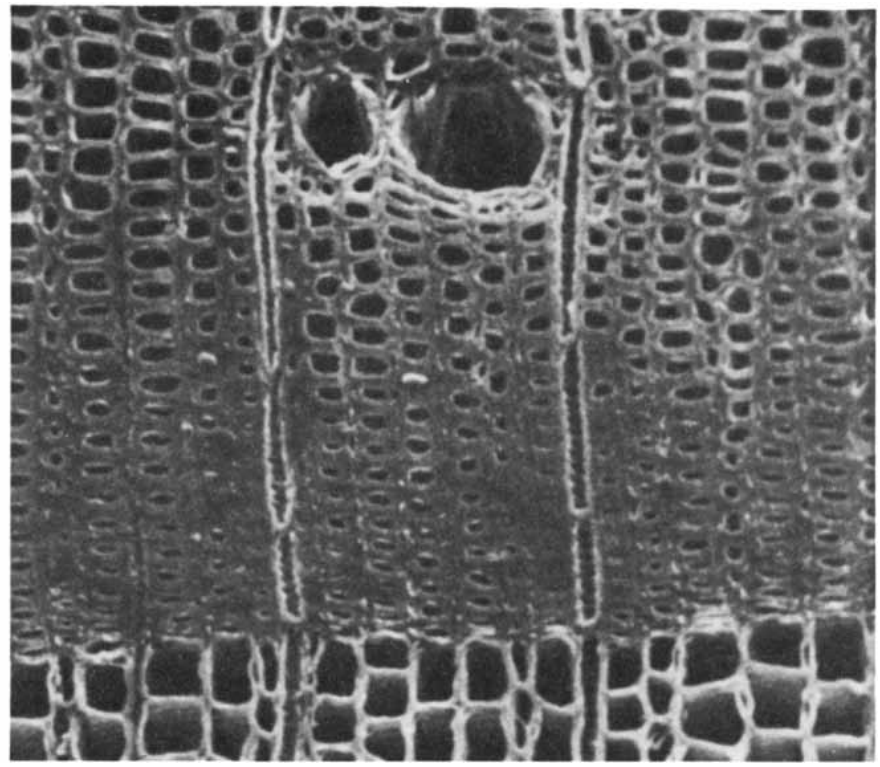
Usually the part of the tree that goes into pulping is the trunk. In nature the trunk has three main physiological functions: transporting water and salts from the roots of the tree to the crown, carrying the crown aloft into the light and storing reserves of nutrients. The trunk has three constituents: the xylem, which is the woody tissue; the cambium, which is a layer of growing cells between the xylem and the bark, and the bark, which is subdivided into the white inner bark (the phloem) and the dark outer bark (the cork).

The cambial cells grow by dividing. When such a cell divides, one daughter cell becomes a new xylem cell, and these longitudinally arranged xylem cells form in concentric layers outside the older wood. In the spring in temperate zones the xylem cells grow rapidly and so have a thin wall and a large lumen (inner cavity); this is the springwood. In the summer the cells have a thicker wall and a smaller lumen; this is the summerwood. In most tree species a year's growth is easily seen as a ring in which the springwood is different in appearance from the summerwood.

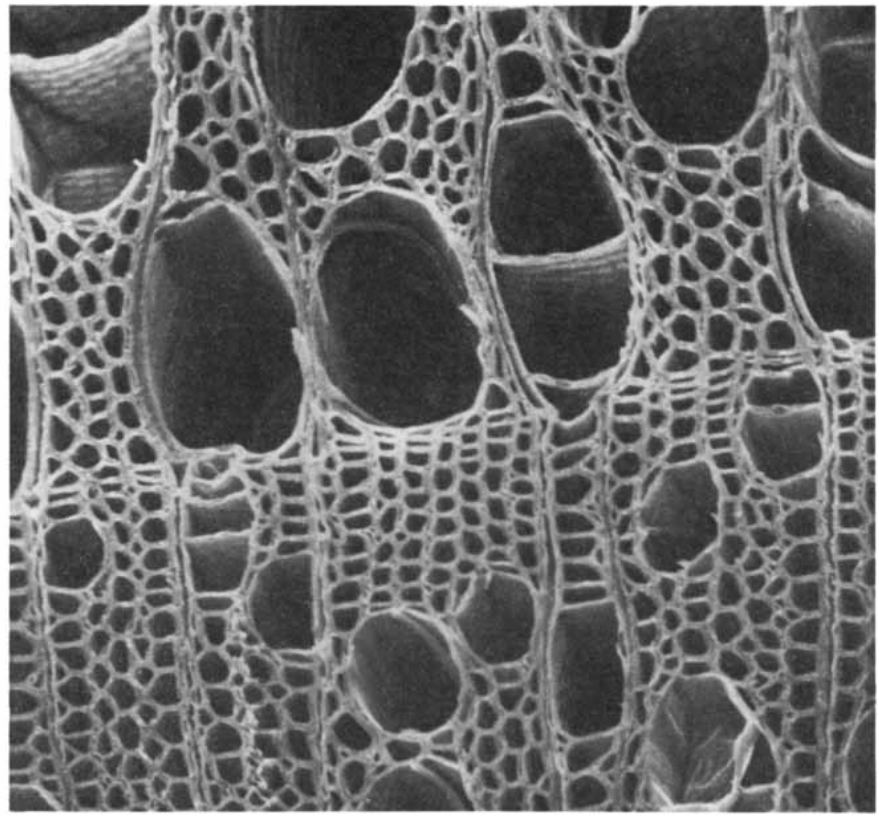
Cells of Wood

Four main types of cell are formed from the cambium. They are fiber, tracheid, vessel and parenchyma cells. Fiber cells are the supporting element, giving rigidity to the structure. The tracheids and the tubular vessel cells serve both to conduct water and to give strength to the structure. Parenchyma cells are thin-walled and store nutrients; these cells have no value in pulping. Other cell types are intermediate in structure between the "libriform" fiber cells, which serve a purely mechanical function, and the tracheids.

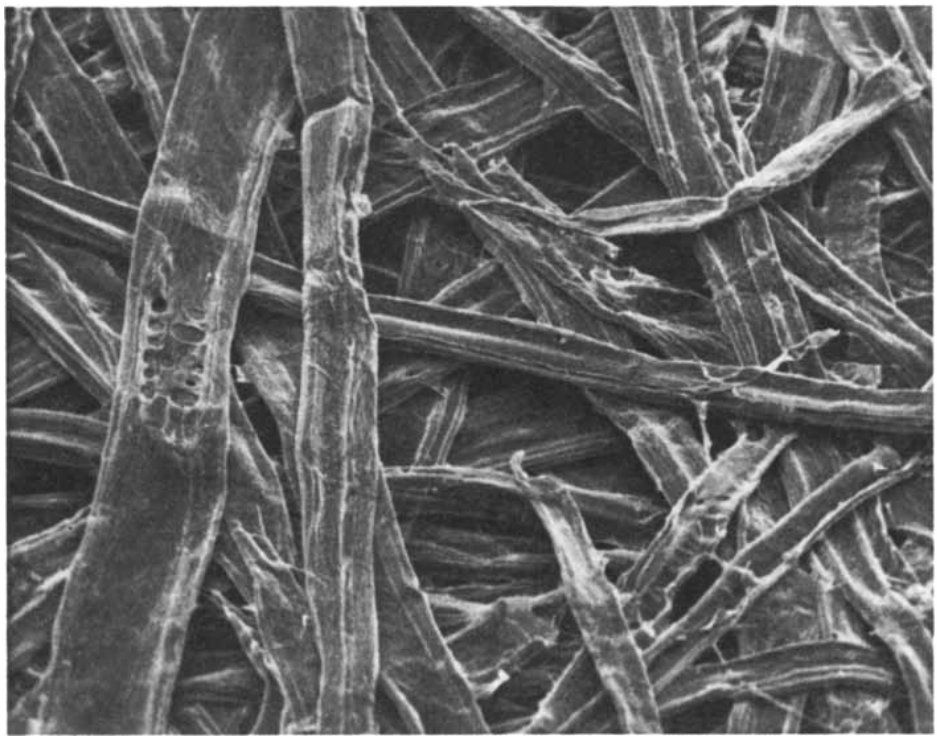
Tracheids predominate in softwoods, which have no vessel cells or libriform fiber cells. In hardwoods, which evolved later than softwoods, vessel and fiber cells predominate. The difference gives rise to the ability of the hardwood tree to transport large quantities of water rapidly. Softwoods have longer and broader fibers than hardwoods and are generally more uniform from species to species. The softwood species grown in western



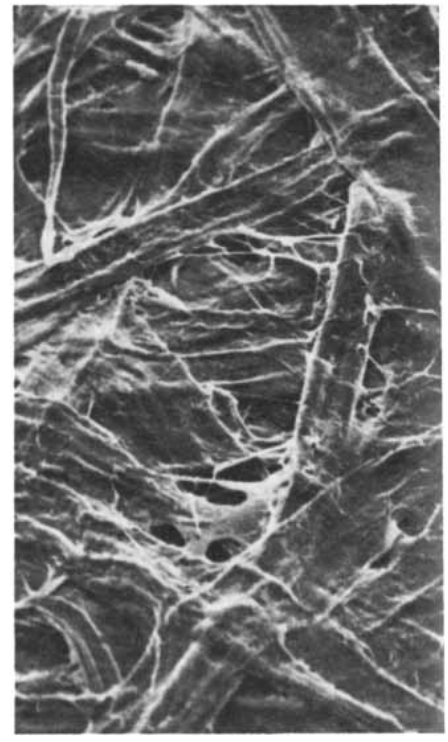
FIBERS OF SOFTWOOD appear at an enlargement of 250 diameters in this scanning electron micrograph of a cross section of a piece of Douglas fir. The densely packed, thick-walled cells in the upper part of the micrograph are the kind of cells that make up the fibers of summerwood. They differ in shape and wall thickness from the springwood cells in the lower part of the micrograph. The two large openings at top center are resin ducts.



FIBERS OF HARDWOOD appear in a scanning electron micrograph of a cross section of a piece of Eastern cottonwood. The enlargement is about 250 diameters. The larger structures are vessels that conduct water in the tree. The objects of smaller diameter are fibers.



SHEET OF PAPER is viewed under a scanning electron micrograph at an enlargement of 150 diameters. The fibers were obtained from a coniferous, or softwood, tree by chemical pulping.



BEATEN FIBERS appear in a sheet of paper made from the same pulp as

North America have exceptionally long fibers.

A measure of structure known in the industry as "percent wall fraction" is indicative of the plasticity of the fiber and hence is related to the fiber's papermaking characteristics. The term refers to the percentage of the radius of a fiber cell that is represented by wall. Hardwoods of dense structure have a high percent wall fraction and make poor paper; the reverse is true of fibers with a low percent wall fraction. Springwood has a wall fraction under 20 percent, whereas the wall fraction in summerwood is more than 40 percent. Ideally wood pulp of good papermaking quality should be made from wood with a wall fraction below 40 percent. Tropical and subtropical woods, which do not contain rapidly grown springwood fibers, have poorer papermaking characteristics than Temperate Zone trees.

Characteristics of Polymers

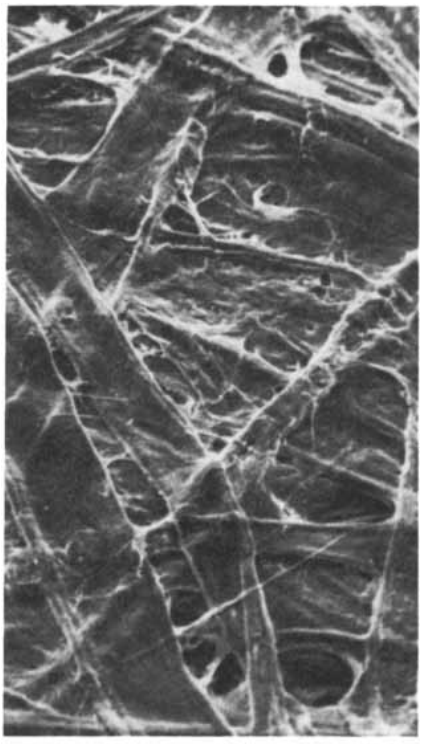
Natural fiber-forming polymers in general, and cellulose in particular, have several characteristics that differentiate them from synthetic polymers. Most natural fiber-forming polymers are hydrophilic, that is, they have a strong affinity for water; the equivalent synthetic polymers are largely hydrophobic. Cellulose is highly hydrophilic. Cellulose fibers are also flexible, strong, physically and chemically durable and have a high self-bonding, sheet-forming ability if they are brought into intimate contact with one another. This combination of features gives paper many of its distinctive properties.

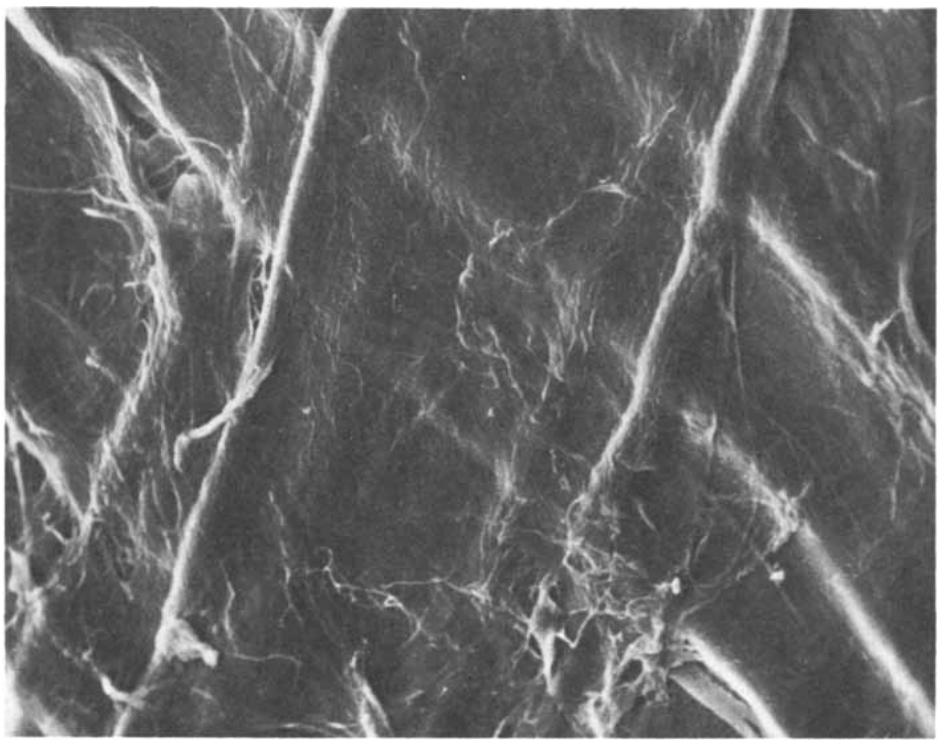
Hydrogen bonding between the molecules of cellulose is the reason for the strength and stability of cellulose fibers. The large number of intermolecular hydrogen bonds also explains why cellulose cannot be melted by the application of thermal energy. When cellulose is heated, the molecules themselves break down before the intermolecular hydrogen bonds break. In cellulose fibers the linear molecules are arranged roughly parallel to one another and are oriented in the direction of the long axis of the fiber. The structure has both crystalline and amorphous regions, and a single polymer chain may pass through several such regions.

In hemicellulose many nonlinear, branched-chain polymers are found. They are generally amorphous rather than crystalline. Hydrogen bonds are formed more easily in hemicellulose than in cellulose because the molecules of hemicellulose are not constrained by crystallized regions. For this reason hemicelluloses provide a conformable adhesive within and between fibers in papermaking, and pulp with a high content of hemicellulose makes strong paper. Molecules of lignin are largely amorphous and hydrophobic; they do not form many intermolecular hydrogen bonds. Therefore the presence of a large amount of lignin in wood pulp reduces the strength of the paper.

Pulping, as I have indicated, consists in rupturing the bonds between fibers by mechanical, thermal or chemical processes or a combination of them. It is the middle lamella, or layer, between the fibers that must be ruptured, and it is composed mostly of lignin. The middle lamella is analogous to the adhesive between the plies of plywood. With paper pulps it is desirable to leave the hemicellulose in place because of its role as an interfiber and intrafiber bonding agent during papermaking. With pulps for other purposes it is desirable to remove as much of the hemicellulose as possible, since the material interferes with both the solubilization of the cellulose and the reprecipitation of the filaments in the regeneration phase of the viscose and acetate processes.

The strength of the middle lamella in wood is such that mechanical force alone rarely causes rupture. An increase in thermal energy causes a rapid decrease in the strength of wood, particularly when the wood is wet. Over the temperature range from 60 to 120 degrees C. wet wood loses 80 percent of its strength; if the wood is dry, the loss is lower and higher temperatures are required. Chemically delignified wood





one in micrograph at left. Mechanical beating increases adhesion of fibers.

WELL-BONDED FIBERS appear in a scanning electron micrograph of a sheet of paper made from fibers that have been beaten after pulping. Paper is seen at an enlargement of 900 diameters.

loses strength rapidly. At the end of cooking, wood cooked in a neutral sulfite process with a yield of 75 percent has only 10 percent of its original strength, a condition that greatly assists the mechanical defiberization process that follows.

Most natural polymers, including cellulose, are biodegradable; synthetic polymers are not. The complex material of wood decomposes fairly rapidly in a damp, warm environment because of attack by bacteria, fungi and insects. Pure cellulose is attacked even more rapidly than wood. Therefore all types of pulp, from highly purified cellulose to groundwood, which still has all its lignin, are rapidly decomposed, as are the paper products made from them.

Handling the Wood

The cost of wood constitutes from 40 to 75 percent of the cost of manufacturing pulp. The industry therefore lays a great deal of emphasis on maximizing the efficiency of the procedures whereby wood is obtained and handled. One area of emphasis is the mechanical harvesting of trees. A development in this area is the Busch combine, which grasps a tree, slices through the trunk, delimbs the fallen tree, cuts the logs and stacks them. A more recent development is a machine that reduces the entire tree, limbs and all, to chips in the forest. In this way the yield of fiber per acre is increased, since tops, branches, bark and even leaves and needles are transported to the mill for pulping. The machine reduces the cost of preparing the site for replanting, but it also reduces the quality of certain grades of pulp. The development is so recent that its impact on the industry cannot yet be accurately estimated.

Machines of this kind are unsuitable for trees in mountainous or marshy terrain and in forests where the trees are large or vary widely in size. In such places much felling is still done with hand-held chain saws. This type of felling leaves the problem of disposing of the "slash," consisting of the crowns and branches. It is often burned.

The logs are transported to the mill by floating or by trains and trucks. In the northern woods of Canada and Scandinavia river drives remain a common method, which has the advantage of costing only about a tenth as much as land transportation. On the other hand, the amount of wood tied up in a river drive is substantial and involves a considerable capital investment. Moreover, the seasonal nature of the system gives rise to technical and production problems in the mills.

The wood yards of mills often contain thousands of cords of wood, stored in random piles that are difficult and expensive to handle. Because of increasing demands for lumber, pulp mills are receiving more of their wood supply in the form of residual chips from sawmills. This trend has resulted in an increase in the storage of wood in the form of chips, either in a silo or in an outdoor pile. Prolonged storage of chips outdoors leads to significant losses from fungal attack, and better methods of outdoor storage are therefore the subject of intensive research.

Logs are often debarked by rolling them in slotted drums; the tumbling action takes off the bark, which escapes through the slots. Other methods include stationary and rotating knives. Debarking is also done hydraulically with jets of water.

Trends in the Industry

Pulp mills have a reputation for malodorousness, which originates with the sulfur employed in pulpmaking. The objectionable odor around most kraft mills can be virtually eliminated by suitable engineering, but an even more effective scheme would be to do away with sulfur altogether. Several nonsulfur processes are under investigation. Most of them are based on removing lignin with an oxidative reaction or a solvent, not necessarily in an alkaline medium. For example, it is known that chlorine dioxide selectively attacks lignin in a mild acid condition without

damaging the cellulose and hemicellulose. Molecular oxygen itself is under investigation as an oxidative agent.

With regard to water pollution, the kraft process is already markedly superior to the sulfite process because the cooking liquors employed in the process are almost completely recovered. It is the effluent from the bleaching plant that is troublesome in kraft processing, because the material cannot be fed to the evaporators and furnaces of the chemical-recovery system without severe corrosion problems caused by the buildup of the chloride ion. Hence the industry has been seeking ways to reduce the reliance on chlorine in the bleaching plant, and such processes are now forthcoming.

The engineering incorporated into pulping plants and bleaching plants is steadily being improved. For example, the digesting of wood chips by the traditional method of separate batches is now being replaced by continuous digesters: tall pressure vessels that accept wood chips at the top and discharge delignified chips ready for blowing at the bottom. Batch digesters that operate automatically are also coming on the scene. Both continuous digesters and automatic batch digesters can be controlled by computer. Computer control of bleaching plants is fairly well established and is highly effective.

It seems certain that more mechanical equipment for removing fibers will come into play. A thermomechanical method of producing groundwood of superior quality has recently been developed. Chips are steamed at a pressure of 30 pounds per square inch for a few minutes to soften the lignin. They are then defiberized in either a pressurized refiner or one operating at atmospheric pressure. The process liberates a large number of substantially undamaged fibers from unmodified wood, resulting in a superior groundwood that appears to have a good future, since its whole fibers are suitable in many applications where normal groundwood is inadequate. It is possible that this new groundwood will even serve as a substitute for chemical pulp in certain applications.

Further engineering changes can be expected to result from efforts to minimize the effect of pulping on the environment. For example, the direct-contact evaporators now common in the U.S. will be replaced by indirect-contact equipment, resulting in less entrainment of chemicals in the streams of waste gas. Secondary and even tertiary treatment of aqueous effluents will become commonplace, with the result that water of high quality will be returned to streams.

Since the disposal of solid waste is becoming a problem in many areas and from 40 to 60 percent of municipal waste is paper-based material, the question of recycling paper attracts much attention. The fact that only 20 percent of the paper produced in 1972 was made from recycled fiber suggests that recycling is at present not a major means of disposing of waste paper. It also reflects the fact that a number of technical and economic problems must be faced in recycling paper.

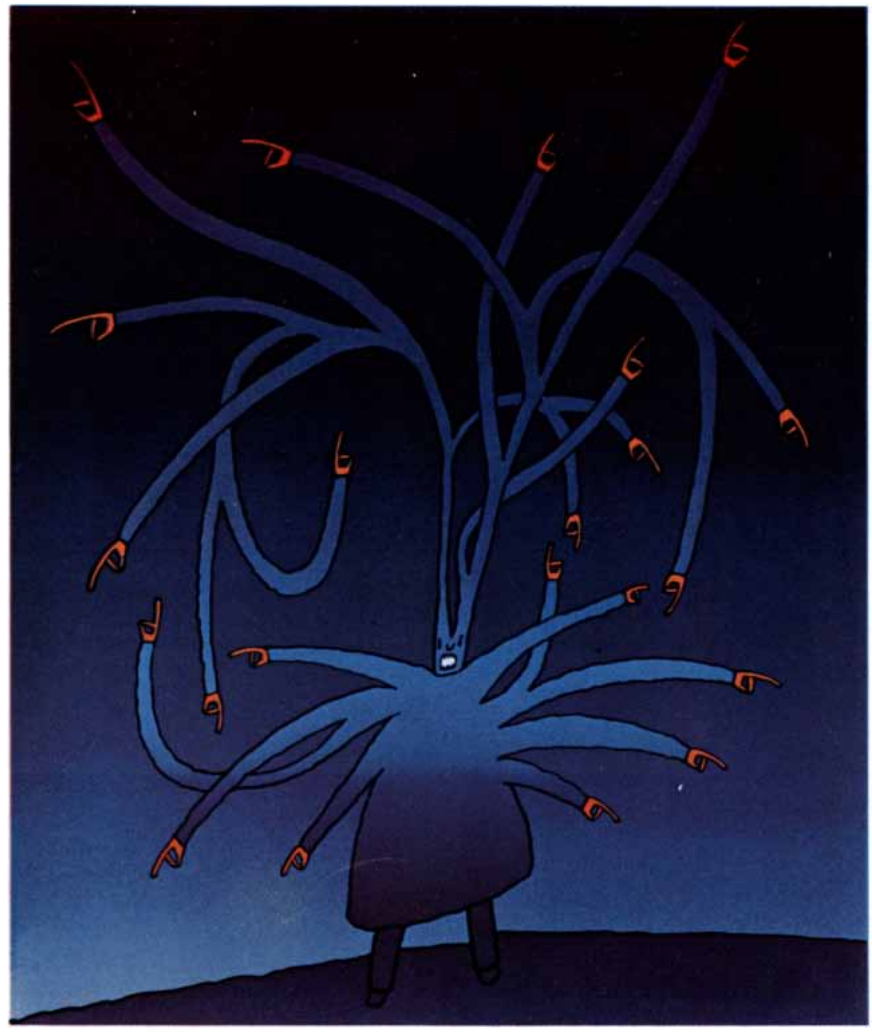
Problems of Recycling

The technical problems include the recovery of clean fiber from mixed wastes, the removal of ink from printed sheets, the separation of clay and other fillers and coatings and the recovery of a reasonable yield of acceptable fiber from a large quantity of unusable fines, which themselves present a serious disposal problem. Since the recovered material varies in quality, the quality of the paper products inevitably fluctuates, and it is not yet clear that consumers are willing to accept such fluctuations as part of the price of recycling.

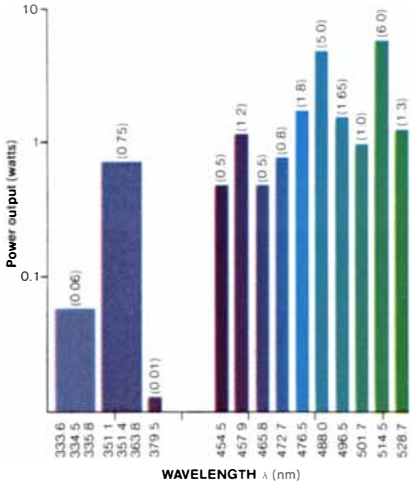
The economic problems include the separation of waste into broad categories in the home, for which there is at present little economic or moral incentive, and the cost of collecting such wastes and transporting them to recycling plants. Recycling paper requires a specialized pulp and paper mill, which must be large to be economic. In addition a paper mill is product-specific, that is, it can produce newsprint or corrugated material or tissue but not all three.

An alternative to this kind of recycling is what can be termed natural recycling, in which waste paper is converted to carbon dioxide and water by either fire or biological processes. Carbon dioxide is the basic material out of which a tree makes itself, so that in the end the natural recycling process converts paper products into new cellulose. Since, according to the American Forest Institute, the U.S. is growing about 8 percent more wood than it is consuming, the requirement for carbon dioxide is substantial. Young, vigorous, growing forests produce about four tons of oxygen per acre per year for each six tons of carbon dioxide they consume. Therefore one can say that the ultimate answer to the problem of disposing of cellulosic waste will be a compromise: the recycling of used fiber will increase and so will the production of new fiber by way of natural recycling.

ION LASERS WITH THE POWER TO TAKE YOU PLACES YOU'VE NEVER BEEN.



Spectra-Physics had tomorrow in mind while developing the first commercially available ion laser to operate reliably at beam powers beyond 10 watts.



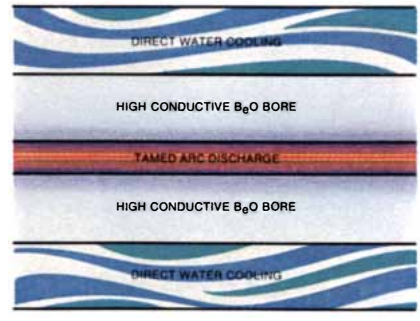
The Model 170 has power in reserve to explore future directions in such fields as Raman spectroscopy, molecular excitation with ultra-violet light, high-resolution holography, data storage and microelectronics fabrication.

In continuous operation, the Model 170 generates from 300 milliwatts to over 20 watts across the spectrum, including watts of ultraviolet emission. Optional optics divide the spectrum into TEM_{OO} outputs at 17 discrete lines or into several wavelength combinations.

Single-line powers range from 5 milliwatts at the weakest UV line to over 6 watts at the strongest visible line. And you may select multi-line outputs to 740 milliwatts at UV to 15 watts visible.

Yet noise can be held to 1% RMS or less and long-term amplitude drift within ±0.5% — important considerations when experiments must be run time and again.

Like Spectra-Physics' medium-power ion lasers, the Model 170 is based on a long life beryllium-oxide plasma tube in a kinematically isolated, heat-shielded Stabilite™ resonant cavity.



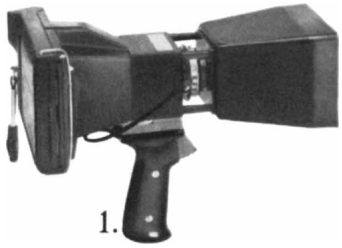
B_eO PLASMA TUBE

For all its power, this laser is a precision, reliable instrument — a source of pure optical power that will give maximum experimental value tomorrow as well as today.

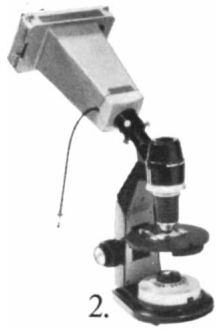
Spectra-Physics

1250 West Middlefield Road
Mountain View, California 94040
(415) 961-2550

As you can see from the list on your left, there's no limit to the kind of material that people need to record. Fortunately, as you can see from the list of Polaroid Land cameras below, there's almost no limit to the kind of photographic records you can get. And, since you get them on the spot, if anything's missing, you can re-shoot on the spot. You can get them in color, or almost *any* form including black-and-white slides or prints with negatives. Because we make 14 different instant-picture films for our lab cameras. And 4 different lab cameras for our films.

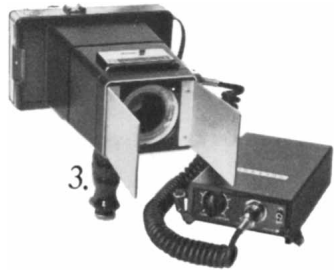

1.

1. *CR-9 Oscilloscope Camera* is hand-held and can be used with almost any oscilloscope. Just change the hood to fit the scope and get distortion-free trace recordings in seconds.


2.

2. *ED-10 Instrument Camera* costs less than $70. It makes excellent photomicrographs in seconds. Your picture is as good as the microscope's own image because the ED-10 uses the microscope's own optics.

3. *CU-5 Close-up Camera* is a hand-held unit that takes closeups from ¼:1 to 3:1. There's no troublesome calculation, because focusing and exposure settings are predetermined. A built-in electronic ring light on the camera is powered by a portable pack when no AC outlet is available.

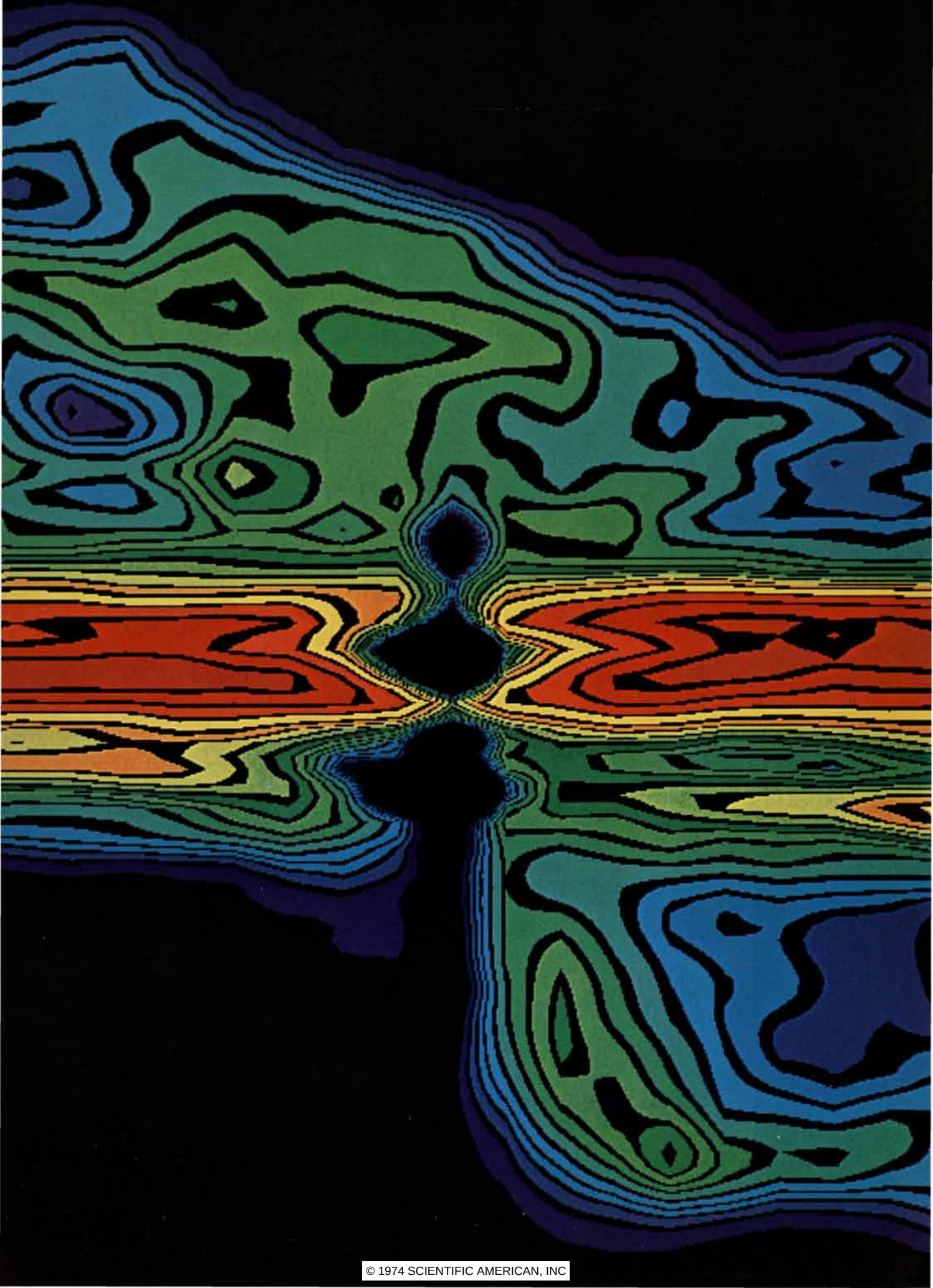

3.

4. *MP-4 Multipurpose Camera* is practically everything you'd expect in a photo studio. It makes photomicrographs, photomacrographs, copies, and photos of small objects or specimens. Remove the camera to a tripod and you have a studio view camera. You can even turn the MP-4 into a professional enlarger. And there are enough accessories available to photograph almost anything.


4.

For detailed information, or a demonstration, write to Polaroid Corporation, Dept. 57-288, Cambridge, Mass. 02139. Or call collect (617) 547-5176.

No matter what you need to record, there's a Polaroid Land camera that will photograph it.



THE CENTER OF THE GALAXY

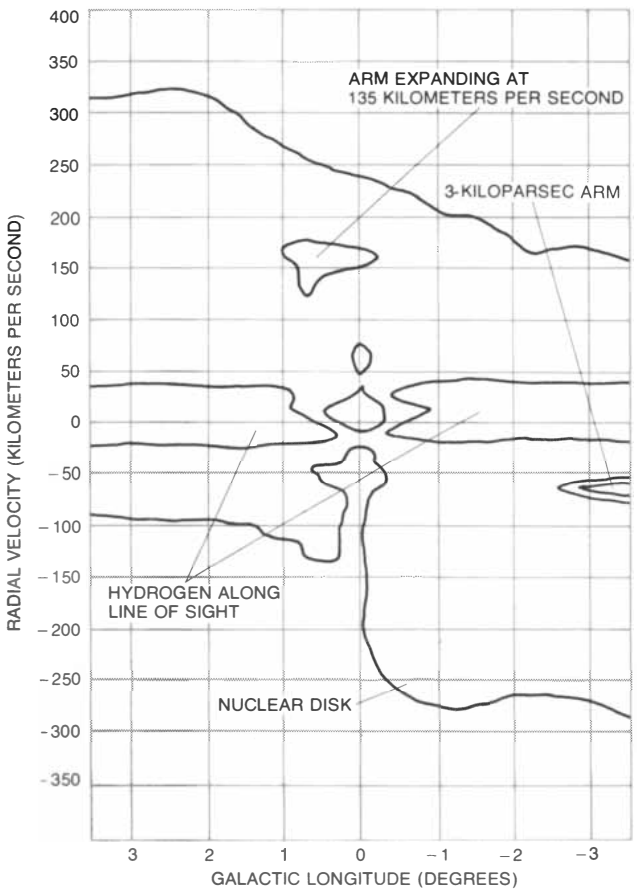
Coded in the radio, infrared and X-ray emissions from the invisible nucleus of our galaxy is mounting evidence that it is periodically the scene of titanic explosions

by R. H. Sanders and G. T. Wrixon

Our galaxy is a disk-shaped collection of stars, gas and dust whose components are all bound together by their mutual gravitational attraction. Like an enormous pinwheel it rotates majestically around its mysterious nucleus, or central region. From our vantage near the edge of the disk the nucleus has until quite recently been hidden by clouds of obscuring dust. Within the past 20 years, however, new techniques of "looking" through the dust have been developed, providing us with tantalizing glimpses of the nuclear region. Piecing together observations made through "windows" at wavelengths both longer and shorter than those of visible light, astronomers have become aware of striking similarities between our galactic center and the nuclei of certain bizarre objects called Seyfert galaxies.

Seyfert galaxies are named for Carl K. Seyfert, who in 1943 first classified a group of galaxies according to the unusual properties of their nucleus. They are spiral galaxies (disk-shaped galaxies with luminous spiral arms like our own galaxy) characterized by a small, very bright nucleus embedded in rapidly moving masses of gas that are apparently being ejected from the central region. Seyfert nuclei are also a rich source of infrared radiation and radio waves. Indeed, Seyfert nuclei emit such large amounts of energy and matter that our present laws of physics may be inadequate to explain them.

Quasars, which were discovered in the early 1960's, may be even more extreme examples of the phenomenon seen in Seyfert galaxies. Quasars are very



GAS MOTIONS AT GALACTIC CENTER are displayed in the computer-generated contour map on the opposite page. Based on observations made with the 140-foot radio telescope of the National Radio Astronomy Observatory at Green Bank, W.Va., the map shows the distribution of emission from un-ionized hydrogen at a wavelength of 21 centimeters. The different colors indicate the intensity of the 21-centimeter line, ranging from violet (lowest level) to red (highest level). The vertical axis, as shown at the left, does not indicate galactic latitude but rather the velocity of hydrogen lying in the galactic plane: velocities toward us are negative, velocities away from us are positive. The horizontal red ridge centered at zero velocity represents all the un-ionized hydrogen along the line of sight on either side of the galactic center. The deep hole at the center identifies where 21-centimeter radiation from the powerful radio source known as Sagittarius A is absorbed by un-ionized hydrogen lying between it and us. To the right and just below the central ridge, at a negative velocity of more than 50 kilometers per second, is a fainter ridge produced by emission from the "three-kiloparsec arm." Since the three-kiloparsec arm also produces an absorption line in the output of Sagittarius A, the arm is on this side of the center, and since its velocity is negative it is expanding away from the center and toward us. At the lower right there is a ridge of emission (*blue-and-green feature*) whose maximum velocity exceeds 200 kilometers per second. This ridge is emission from the nuclear disk; its high velocity means that it is very near the galactic center and rotating rapidly. Notice that the velocity approaches zero toward the galactic center. Above and to the left of the center the picture is somewhat ambiguous. The dominant feature is a ridge of emission (*green band*) that crosses the position of the galactic center at +135 kilometers per second; hence it is an expanding arm of gas moving away from us. The absence of an absorption hole at the position of Sagittarius A means that the expanding arm is on the other side of the nucleus. The program for generating the color contour map was devised by Thomas Cram and David Ehnebuske.

bright starlike objects whose spectra exhibit large red shifts, indicating that they are receding at high velocity. The prevailing view is that objects with red shifts as large as those of quasars are among the most distant objects in the universe. If that is so, quasars must be emitting gigantic amounts of electromagnetic radiation, perhaps 1,000 times as much as Seyfert nuclei. According to one hypothesis, the pointlike quasars are simply Seyfert-like nuclei so distant that the surrounding galaxy is too faint to register on a photographic plate.

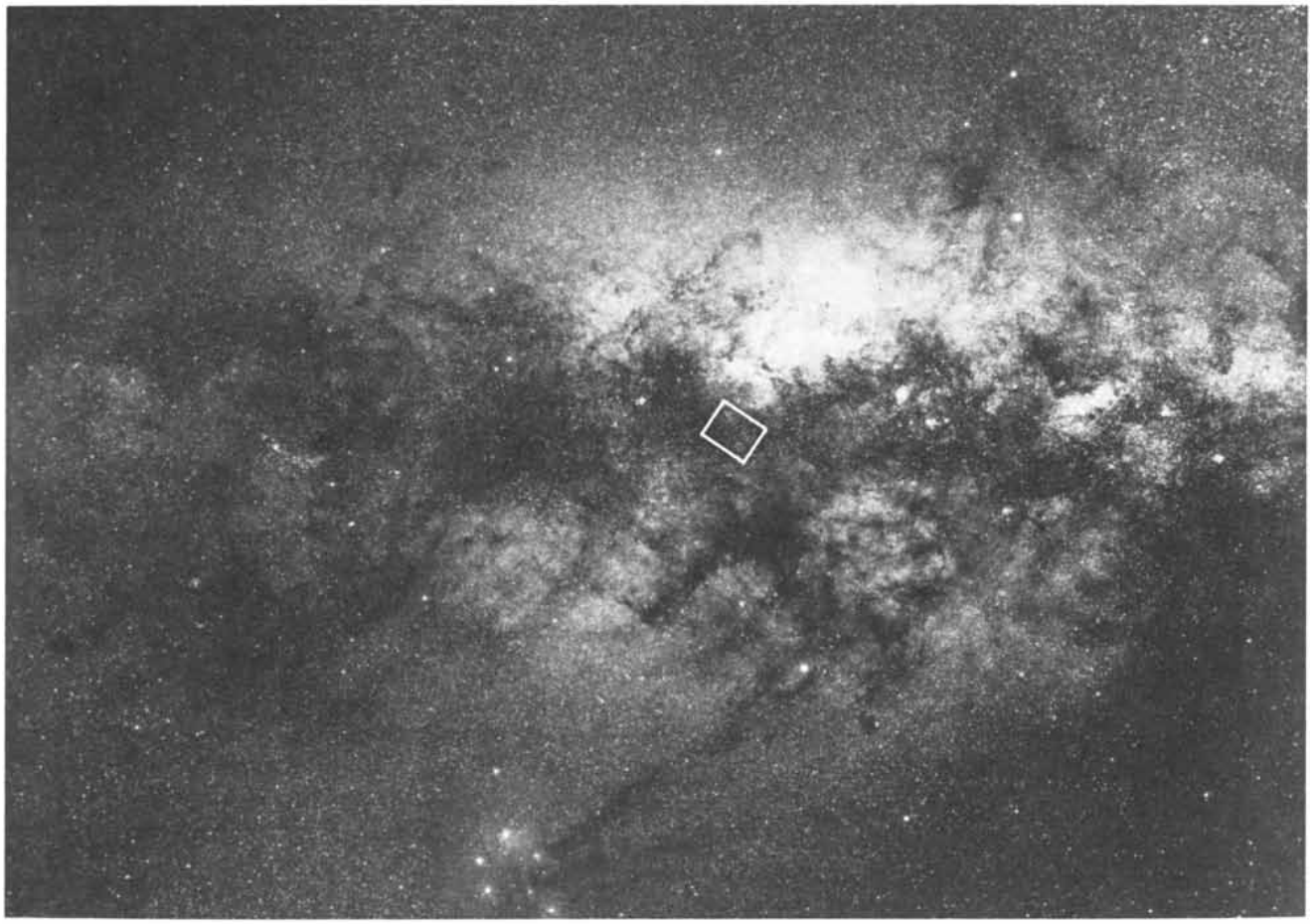
If one compares the energy output and the calculated lifetime of our galactic nucleus with the corresponding values for Seyfert galaxies and quasars, it may appear that the center of our galaxy is a rather peaceful place, at least at the moment in cosmic time when we are privileged to observe it. For example, the power output of our galactic nucleus is less than that of a Seyfert nucleus by a factor of at least 10,000 and less than that of a quasar by a factor of at least 10 million [*see top illustration on page 71*].

It now appears, however, that the present quiet appearance of our galactic center is misleading: the nucleus shows unmistakable evidence of a violent history. The impression is therefore growing that the nuclei of ordinary galaxies, including our own, share a number of significant characteristics with the nuclei of Seyfert galaxies and possibly with quasars as well. The chief similarities are the presence of strong sources of radio and infrared emission and, most important, an energetic flow of matter outward from the core.

These observations suggest that the nuclei of galaxies may exhibit a continuum of energetic events consistent with the phenomena observed in the entire range of objects we classify as normal galaxies, Seyfert galaxies and quasars. It may be that all galaxies pass through epochs of energetic activity, some more violent than others. Therefore by studying, and if possible understanding, what is happening at the center of our own galaxy we may also gain some understanding of the workings of the more energetic Seyfert galaxies and quasars.

The advantage of observing our own galactic nucleus, of course, is that it is relatively close to us. The center of our galaxy is only 10,000 parsecs from the sun. (One parsec is 3.26 light-years.) The nucleus of the Great Nebula in Andromeda, the nearest normal spiral galaxy, is more than 60 times farther away, and the nucleus of NGC 4151, a typical Seyfert galaxy, is more than 1,000 times more distant. Obviously one can observe much finer detail in the nucleus of our own galaxy than one can hope to observe in the nucleus of any other.

In the usual kind of astronomical photograph the center of our galaxy looks like almost any other region of the Milky Way. Vast clouds of dust in the plane of the galaxy prevent any light from the center from reaching us. The dust appears as dark lanes in wide-angle photographs and robs us of what would otherwise be a spectacular view of the nucleus [*see illustration below*]. Visible light, however, is only a tiny segment of the entire electromagnetic spectrum. The

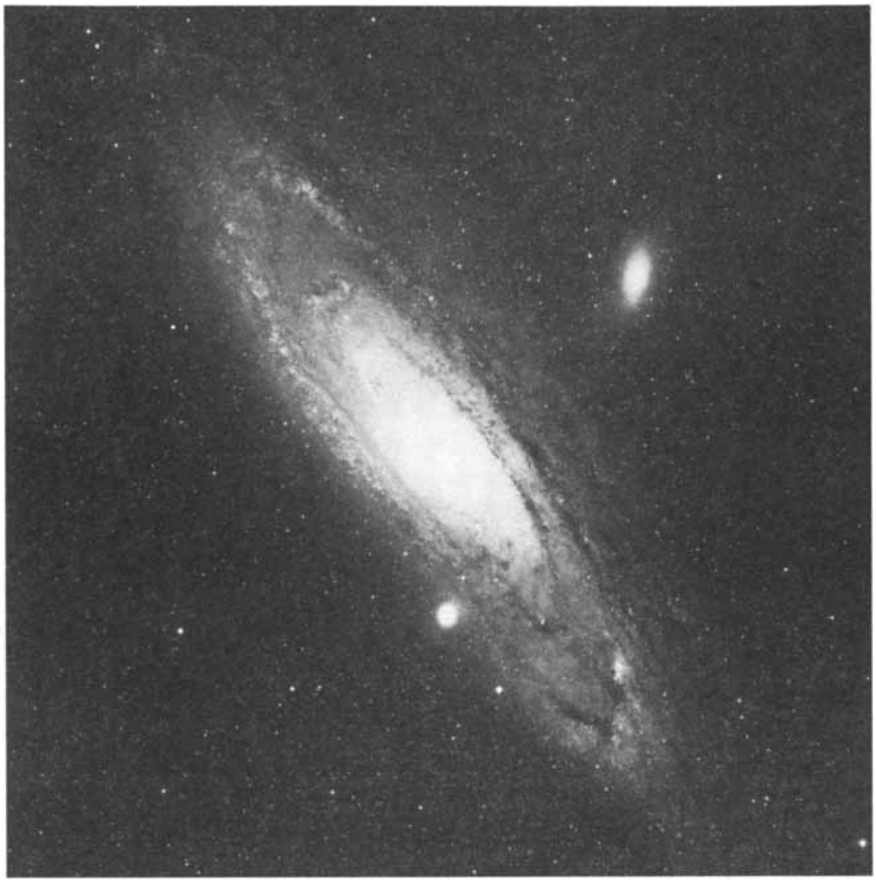


PHOTOGRAPH OF CENTER OF GALAXY indicates how great dust clouds embedded in the Milky Way hide what otherwise would be a spectacular view of hundreds of millions of stars concentrated in the galactic nucleus. The center of the galaxy is some 33,000 light-years (10,000 parsecs) away in the constellation Sagittarius. The rectangle represents the area depicted in the radio-emission map of the galactic center at the top of page 72. The brilliant mass of stars near the rectangle gives a hint of what the nucleus might look like. The photograph was made by the Hale Observatories with a Tessar lens of 10 inches focal length and an emulsion sensitive to red light.

visible portion of the spectrum consists of radiation with wavelengths between 4,000 and 7,000 angstroms (.4 to .7 micron). Because the dust particles in interstellar space have about the same dimensions as the visible wavelengths, they interfere primarily with the passage of visible light. Radiation of longer wavelengths, such as infrared radiation (.7 micron to roughly 1,000 microns) and radio waves (from about 1,000 microns, or one millimeter, up), in effect flows around the dust particles and thus travels relatively unobstructed from the galactic center to us. X rays and gamma rays, being highly energetic, pass right through the intervening dust.

The inner region of the galaxy was first observed with the aid of radio telescopes in the early 1950's. The galactic nucleus has four mechanisms for emitting radio waves; each contributes its share toward our current understanding of the physical and dynamical processes at work. The four mechanisms are (1) emission over a broad continuum of wavelengths by energetic electrons held in orbit by a magnetic field (known as synchrotron radiation); (2) the emission of radiation at a "line" wavelength of 21 centimeters by hydrogen atoms flipping from a higher energy state to a lower one; (3) similar lines emitted at specific wavelengths by various molecules in interstellar space, and (4) both line and continuum emissions from H II regions. These are the regions around hot stars where protons and electrons resulting from the ionization of hydrogen occasionally recombine to form hydrogen atoms. They do so with the emission of photons of specific wavelengths, and at the same time fast-moving electrons passing close to protons give rise to a continuum of radio emissions.

The most powerful source of synchrotron radiation in the galactic nucleus is known as Sagittarius A. It was recognized as being one of the strongest radio sources in the sky before it was realized that it lay at the galactic center. All radio sources outside the solar system are so far away (like most stars) that trigonometry, with the earth's orbit as a baseline, is useless for determining their distance. Conceivably Sagittarius A might lie almost anywhere on a line drawn between the earth and the galactic center. Moreover, because Sagittarius A radiates over a very broad band of wavelengths devoid of identifiable emission lines, there is no way of telling whether it is approaching, receding or standing still. The evidence that it is located at the galactic center is somewhat indirect.



NEAREST SPIRAL GALAXY, the Andromeda galaxy, or M31, shows what our own galaxy would look like if it could be seen from the outside. M31 contains more than 100 billion stars in a great flattened disk roughly 100,000 light-years in diameter. It is two million light-years away. Photograph was taken with 48-inch Schmidt telescope on Palomar Mountain.

First of all, it does lie precisely in the same direction as the dynamical center of the galaxy: the point around which all the stars and gas in the galaxy are rotating. Second, we know from radio observations of other normal spiral galaxies like our own that most of them have bright radio sources in their nucleus. Therefore it would seem highly unlikely that one of the brightest radio sources in the sky lying exactly in the direction of the galactic center was not in fact at the galactic center.

Detailed observations with radio telescopes reveal that Sagittarius A is a source about 12 parsecs in diameter whose continuum emission is generated by highly energetic electrons spiraling in a magnetic field. This immediately tells us that there must be a source of energetic electrons at the galactic center. The synchrotron emission also suggests that the center might be the site of explosive events, since such radio sources are observed at the site of known explosions in the galaxy, such as supernovas. In Seyfert galaxies the synchrotron emission exceeds that of Sagittarius A by a factor of 100.

Embedded in Sagittarius A, and very near the actual center of the galaxy, are several small, bright knots of thermal radio emission about a parsec or somewhat less in size [*see illustrations on page 72*]. Thermal radiation is produced when electrons in a hot ionized gas pass close to protons. Thermal continuum radiation is distinguished from nonthermal, or synchrotron, radiation by the fact that the intensity of the radio emission varies with wavelength in one way for the thermal radiation and in another for the nonthermal. Observations of thermal emissions imply that there are clouds of hot gas in the very core of the nucleus.

The second source of radio emission, the 21-centimeter line radiation from clouds of un-ionized hydrogen, has probably supplied the greatest amount of information about the galactic center and its immediate vicinity. A radio wavelength of 21 centimeters corresponds to a frequency of 1,421 megahertz. Any observed variation from this frequency is assumed to represent a Doppler shift caused by movement of the emitting source either toward the observer or away from him. Therefore observation

of the 21-centimeter line gives us two important kinds of information about the source: the mass of hydrogen present (which is proportional to the signal intensity) and the distribution of velocities in it [*see illustrations on page 74*]. With this information one can construct spatial maps of the distribution of un-ionized hydrogen throughout the galaxy.

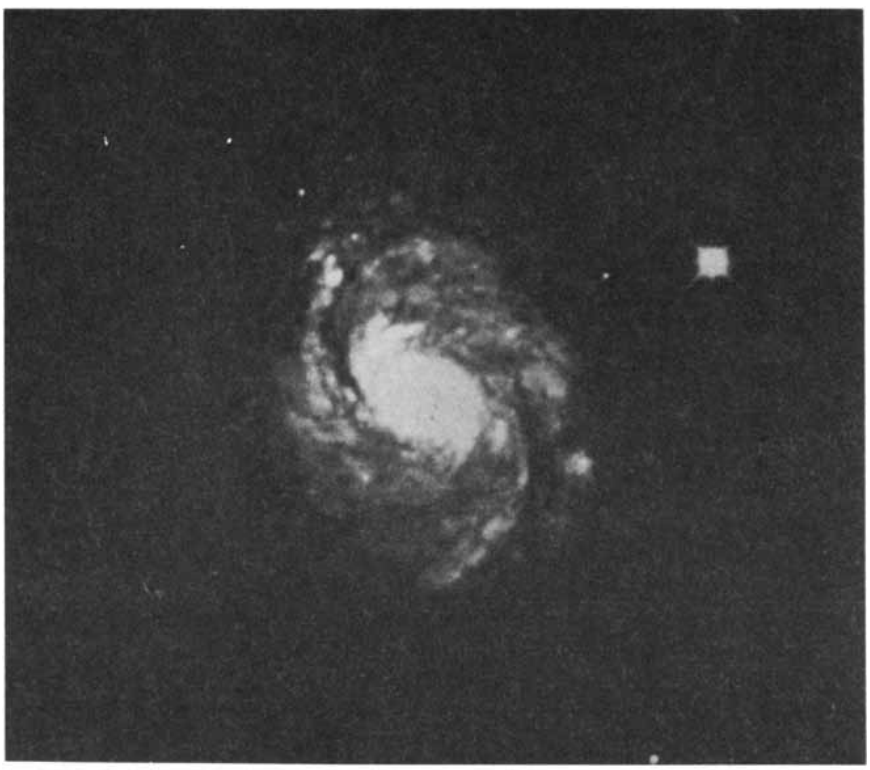
The pioneering 21-centimeter observations carried out by J. H. Oort and G. W. Rougoor of the University of Leiden in the early 1960's showed not only that hydrogen partakes of the normal galactic rotation shared by stars but also that some of the hydrogen in the vicinity of the center is moving outward. This was the first indication that gas was being pushed out of the galactic nucleus, conceivably by unknown sources of energy.

Oort and Rougoor deduced from their observations that there are at least two expanding arms of hydrogen, one lying this side of the center and moving in our direction with a velocity of about 50 kilometers per second and the other lying on the other side of the center and moving away from it at about 135 kilometers per second. The first arm lies a little more than 3,000 parsecs this side of the center and thus is known as the three-kiloparsec arm; the second is the "expanding arm at 135 kilometers per second." The second arm lies somewhere between 2.5 and four kiloparsecs from the center, and therefore it is known by its velocity rather than by its distance [*see top illustration on page 77*].

The evidence for the fact that the arms are expanding away from the center is provided by Sagittarius A. If we study the 21-centimeter emission of the arms, we can perceive that the three-kiloparsec arm is moving toward us and that the expanding arm at 135 kilometers per second is moving away from us. But how do we know that the three-kiloparsec arm is not on the opposite side of the galactic center from us and also moving toward the center? Similarly, how do we know that the expanding arm at 135 kilometers per second is not on our side and moving away from us toward the center? In short, how do we observationally distinguish between gas falling into the center and gas flowing out of the center? The answer is that some of the continuum radiation emitted by Sagittarius A is absorbed at the wavelength of 21 centimeters by un-ionized hydrogen in the three-kiloparsec arm, whereas none is absorbed by the expanding arm at 135 kilometers per second. This means that the three-kiloparsec arm is between us and the galactic center, and since it is moving toward us it is moving away from the center; the other arm is on the opposite side of the center, and since it is moving away from us it is moving away from the center. Therefore we have firm observational evidence for the outflow of gas from the galactic center [*see illustration on page 67*].

The third source of radio emission from the center of the galaxy—line radiation from clouds of molecules in interstellar space—has been discovered only in the past few years. Among the molecules so far identified in the nuclear region are the hydroxyl radical (OH), water vapor (H_2O), ammonio (NH_3), carbon monoxide (CO) and formaldehyde (H_2CO). Molecules are less widely distributed than un-ionized hydrogen, and they are usually observed grouped together in large clouds whose density is much higher than the average density of interstellar gas. Nicholas Z. Scoville of the University of Minnesota has concluded that the molecules form a partial, if not complete, ring around the galactic center at a radius of about 300 parsecs. The ring is also expanding from the center at a velocity of nearly 100 kilometers per second. This provides additional evidence that gas is being blown away from the center, perhaps as the result of a recent explosion.

The fourth source of radio emission from the center of the galaxy, the emission from H II regions consisting of both line and continuum radiation, is a strong indication that new stars are currently being born near the nucleus. Several H II regions have been identified within a few hundred parsecs of the galactic center. Moreover, they seem to be physically related to the large molecular clouds.



SEYFERT GALAXY NGC 1068 is typical of a type of galaxy with a bright starlike nucleus that is also an intense source of radio waves and infrared radiation. Gas has been observed flowing out of the nucleus of NGC 1068 at a velocity that approaches 600 kilometers per second. Photograph was made with the 120-inch reflecting telescope at Lick Observatory.

Some of the most recent information about the galactic center has come to us through the infrared window at wavelengths of between one micron and three microns. Eric E. Becklin, G. Neugebauer and Allan R. Sandage of the California Institute of Technology have found that the distribution of 2.2-micron radiation from the nucleus of our galaxy closely follows the distribution of visible light coming from stars in the unobscured nucleus of the Andromeda galaxy. One can therefore conclude that the 2.2-micron radiation is a good index of the density of stars in the nucleus of our own galaxy. The infrared measurements indicate that the galactic core contains about a million stars per cubic parsec, a stellar density about a million times greater than that in the neighborhood of the sun. This implies that any beings liv-

OBJECT	POWER OUTPUT (ERGS PER SECOND)	LIFETIME OF PHENOMENON (YEARS)	TOTAL ENERGY EMITTED OVER LIFETIME (ERGS)	TOTAL OBSERVED ENERGY IN MASS MOTION (ERGS)
QUASARS	10^{47}–10^{48}	10^{6}	10^{62}	10^{60}
NUCLEI OF SEYFERT GALAXIES	10^{44}–10^{45}	10^{8}	10^{61}	10^{55}–10^{60}
GALACTIC CENTER	10^{39}–10^{40}	10^{10}	10^{58}	10^{53}–10^{54}

ENERGY EMISSION OF OUR GALACTIC NUCLEUS appears to be significantly less than that estimated for Seyfert galaxies and for the still more powerful starlike objects known as quasars. The power output given for quasars is based on the assumption that the large red shift in their spectra is a valid clue to their distance. Since quasars produce spectra with the greatest red shifts known, they are inferred to be the most remote objects in the universe. The estimated and inferred distances of Seyfert galaxies are less controversial than the distances assigned to quasars. Thus the fact that some Seyfert nuclei are nearly as energetic as quasars tends to support the red-shift distances for quasars. The power output given for the center of our galaxy, between 10^{39} and 10^{40} ergs per second, is for sources other than the normal stars present in the core. In spite of the differences in the table, the authors suggest that the nuclei of normal galaxies, such as our own, are capable of violent activity qualitatively similar to that shown by Seyfert nuclei and quasars.

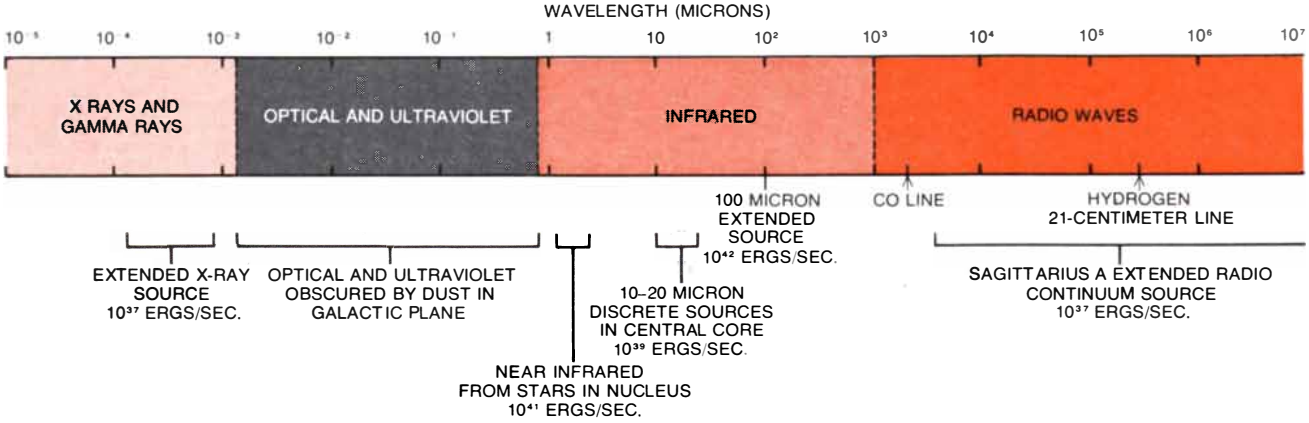
ing on a planet circling a star in the galactic nucleus would see a million stars as bright as Sirius, the brightest star in our sky. The integrated intensity of all the stars in the night sky of such a planet would equal that of about 200 full moons. Under such conditions optical astronomers would be limited to studying the brightest nearby objects; the light of even the nearest galaxies would be drowned out. (It is doubtful, however, that there would be any life on planets in the galactic nucleus, since with such high stellar densities close encounters between stars would be so frequent that planets would be ripped out of their orbit every few hundred million years.)

The infrared radiation emitted by the galactic nucleus between four and 20 microns yields a quite different picture from the one observed at 2.2 microns. George H. Rieke and Frank J. Low of the Lunar and Planetary Laboratory of the University of Arizona find that at the longer wavelengths the radiation seems to come primarily from several discrete sources within about three parsecs of the galactic center. These sources are similar to the bright infrared sources found in the nuclei of other galaxies and may be radiation from dust particles surrounding very hot massive stars. Each of these sources emits a little less than a million times more in the infrared than the sun emits at all wavelengths, indicating that the central star heating the dust particles is perhaps 50 times more massive than the sun. Even allowing for uncertainties in this estimate it is unlikely to be low by a factor of more than two. This places an upper limit of about 100 solar masses on any starlike object so far observed, even indirectly, in the galactic core.

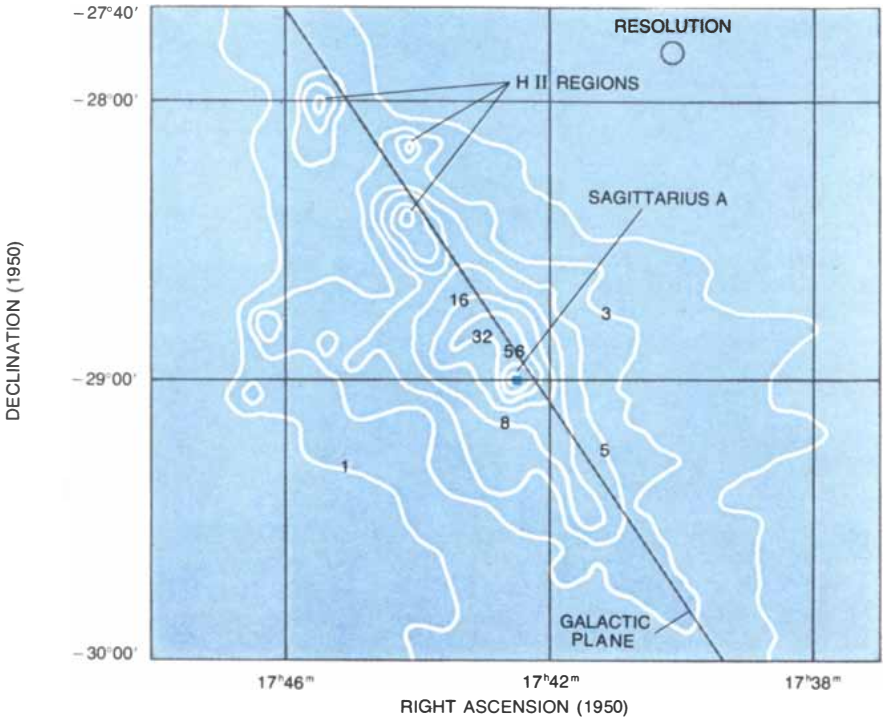
William F. Hoffmann of the Goddard Institute for Space Studies of the National Aeronautics and Space Administration has found that most of the detectable radiation coming from the galactic center is emitted in the far infrared at a wavelength of about 100 microns. This vast outpouring of energy, amounting to between 10^{42} and 10^{43} ergs per second, originates in an extended region about two degrees across near the center of the galaxy. (The full moon is half a degree across.) Since the emission is evidently produced by dust particles heated by starlight, the ultimate source of the energy is no mystery. It must originate in the enormous aggregate of normal stars packed into the galactic nucleus.

The most recently opened window on the galactic center is at X-ray wavelengths. Investigators at American Science and Engineering, observing with X-ray detectors aboard the artificial satellite UHURU, have detected an extended X-ray source that roughly coincides with the 100-micron infrared source. The mechanism for the emission of X rays is not known, but it may be that they are simply part of the thermal emission from a very hot extended gas in the central few hundred parsecs of the galaxy.

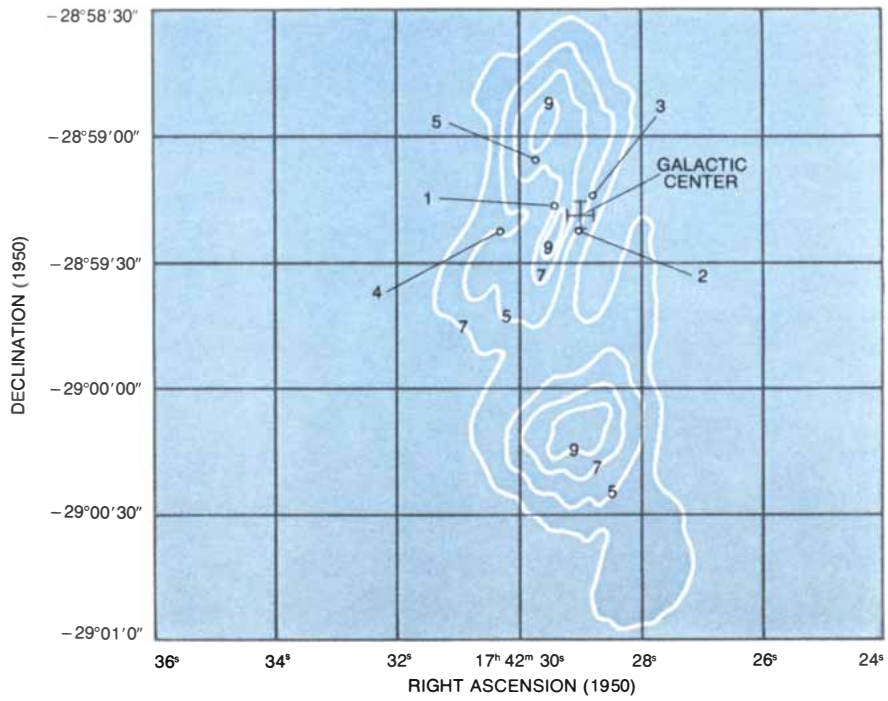
Thus we see that the picture of the galactic center that reaches us through the various electromagnetic windows is a complex one. The high density of stars is to be expected. We have also discovered within a few hundred parsecs of the center, however, large clouds of interstellar molecules, stars in the process of birth and H II regions. We observe in the



GALACTIC NUCLEUS CAN BE OBSERVED by selecting wavelengths of electromagnetic radiation capable of penetrating the clouds of dust that absorb visible light, ultraviolet radiation and long-wavelength X rays. Although some of this penetrating radiation is ultimately blocked by the earth's atmosphere, much of it can be observed from the ground. The illustration describes the kinds of source in the galactic nucleus that have been identified at various wavelengths and gives the total power emitted by each source.



CONTINUUM RADIO EMISSION FROM GALACTIC CENTER was mapped at six centimeters by D. B. Whiteoak and F. F. Gardner using the 64-meter radio telescope at Parkes in Australia. Three H II regions, envelopes of ionized gas around hot stars, are sources of both thermal continuum and line emission. The nonthermal source Sagittarius A is evidently located at the center. A detailed map of the radio emission inside the rectangle appears below.



CORE OF SAGITTARIUS A appears in this detailed radio map representing the area within the rectangle in the map at the top of the page. The core was mapped by Bruce Balik with the interferometer at the National Radio Astronomy Observatory. It shows the small but intense thermal radio emission from hot ionized gas that occupies the galactic center. The cross marks the point of maximum stellar density as determined from infrared observations at 2.2 microns by Eric E. Becklin and G. Neugebauer. The five numbered regions are point sources that shine brightly in the infrared at 10 to 20 microns. They were discovered by Frank J. Low and George H. Rieke. The numbers on contours indicate intensity of emission.

inner few parsecs of the nucleus thermal emission from hot ionized gas and nonthermal emission from high-energy electrons spiraling in magnetic fields. In the same inner region we find discrete sources of infrared radiation. Perhaps most striking of all, we observe gas streaming out of the center at high velocity. Can all these seemingly diverse facts be combined into one coherent picture? Are the various phenomena indicative of outbursts of energy in our galactic nucleus similar to the violent events we see in the nuclei of Seyfert galaxies?

The outflow of gas from the galactic center is the most direct evidence that an explosion of some kind has occurred. The present kinetic energy of expansion in the three-kiloparsec arm is 10^{53} ergs, but the energy of the assumed explosion that caused the expansion must have been very much larger. Moreover, a considerable mass of gas was probably ejected from the nuclear region by the explosion.

In order to learn something about the character of the explosions that would be needed in the center of the galaxy to form the three-kiloparsec arm we have numerically simulated such explosions on a large electronic computer. This work was done at Columbia University in collaboration with Kevin H. Prendergast.

Although the real galaxy consists of both stars and gas (chiefly un-ionized hydrogen), most of its mass resides in the stars. Therefore it is the stars that determine the overall gravitational field of the galaxy. An explosion, however, affects only the gas, leaving the gravitational field unaltered. This greatly simplifies the computational problem, because we need only to follow the motion of the gas in a constant gravitational field.

Modeled on the real galaxy, the computer galaxy assumes that the gas is distributed in a disk slightly less than 100 parsecs thick that is rotating around the center in an overall gravitational field determined by the stars. The explosion itself is simulated by placing a very hot concentration of gas right in the center of the disk. The hot gas immediately starts expanding in all directions.

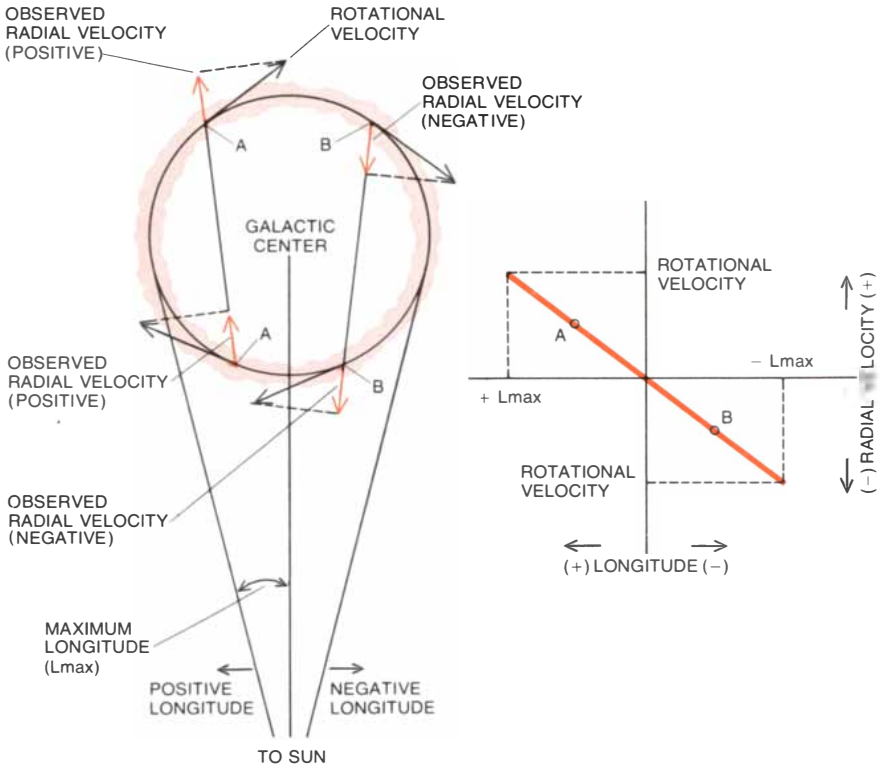
The computer studies show that most of the expanding gas flies out of the plane of the galaxy and never affects the gas in the disk. Nevertheless, a small fraction of the hot gas does push the disk outward. Ten million years after the explosion the matter that was blown out of the plane has effectively escaped from the galaxy, but the matter in the plane

has been pushed out into an expanding ring of gas about three kiloparsecs from the center. In the computer model there is virtually no gas between the galactic center and the ring. Moreover, the gas in the ring has cooled below the point where the hydrogen is ionized (thus it will radiate only the 21-centimeter line of un-ionized hydrogen), and the ring itself is rotating.

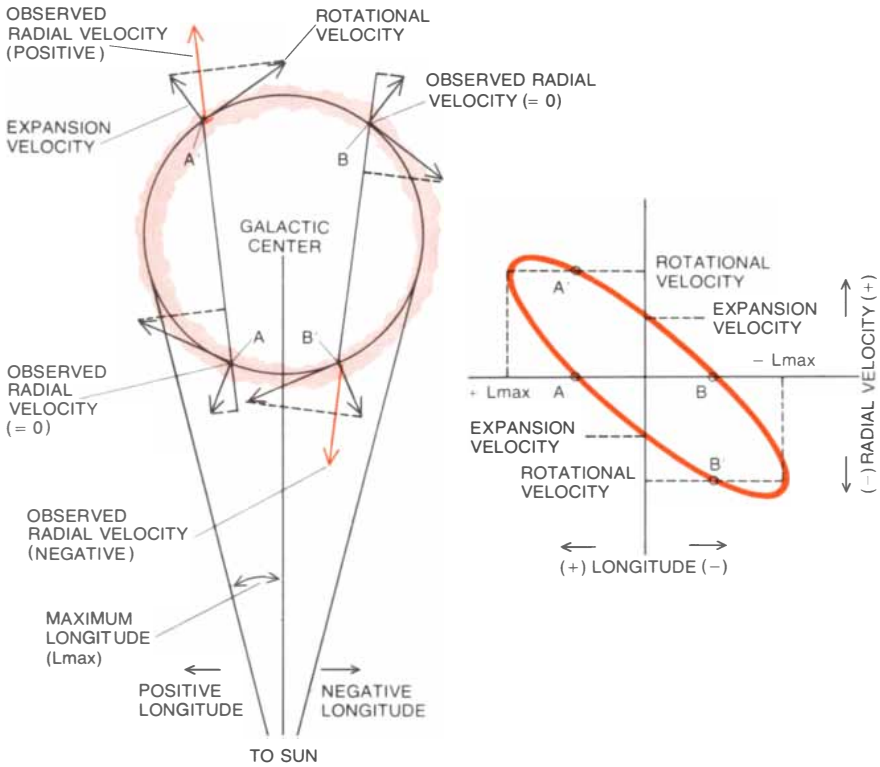
Since the gas forming the ring was originally lying close to the galactic center, however, the ring itself is rotating more slowly than the gas from which it was created (just as a spinning figure skater slows down when he extends his arms). Indeed, the ring is rotating so slowly that its centrifugal force is unable to balance the gravitational force pulling it back into the center. Eventually the outward motion of the ring stops at a distance from the center that depends on the original explosive energy and on the mass ejected from the center by the explosion. In order to produce a ring at the observed distance of the three-kiloparsec arm, moving outward from the center with the observed velocity of 50 kilometers per second, the explosion energy must be 3×10^{58} ergs, the equivalent of converting the mass of more than 10,000 suns completely into energy. The ejected mass must be more than 100 million times the mass of the sun. In this case the ring matches the observed conditions in 10 million years.

Does this mean that a titanic explosion rocked the galactic center only 10 million years ago and that the three-kiloparsec arm is the observable remnant? If this were the case, then the presence of other expanding features in the central region (such as the ring of molecular clouds) would suggest that such explosions happen every five million years or so. Therefore over the lifetime of the galaxy (10^{10} years) more than 10^{62} ergs of energy have been released by explosions, with the ejection of more than 10^{11} solar masses. This would seem highly improbable because 10^{62} ergs exceeds the estimated amount of energy released by the most active Seyfert nuclei known, and the ejected mass is equivalent to the mass of the entire galaxy. Can there be some other explanation for the three-kiloparsec arm, particularly since the region between the arm and the center is not swept clean, as the computer model predicts, but actually contains a good bit of gas?

Fortunately there is a way out of the numerical improbabilities of the explosion hypothesis. If we continue to follow the evolution of the expanding ring on the computer, we find that it continues



MOTIONS OF HYDROGEN AT GALACTIC CENTER show up as displacements in the wavelength of 21-centimeter line radiation, provided the gas has a radial component of velocity, that is, a component either toward or away from the observer. If the gas is distributed in a ring (*left*) that is simply rotating, neither expanding nor contracting, the radial component will be zero at zero longitude and will reach a maximum at the longitude corresponding to the radius of the ring. In the diagram at right radial velocities of points *A* and *B* on the ring are marked on a curve showing the entire range of observed radial velocities.



EXPANDING AND ROTATING RING OF GAS exhibits radial components of velocity even at zero longitude. When the velocity components for various longitudes are plotted, one obtains an elliptical curve. Zero radial velocities are recorded at two points (*A*, *B*) on opposite sides of the ring where the velocity of expansion exactly cancels the radial velocity.

outward to more than four kiloparsecs from the galactic center before it is stopped by gravitation. It then begins to fall back into the center, rotating faster and faster (like the figure skater pulling his arms into his body). Eventually the ring reaches an equilibrium radius where its increasing rotational velocity again balances gravity. At this point, however, the ring has acquired inward momentum, so that it continues to move inward until it reaches a radius of about 2.5 kiloparsecs. Thereupon the centrifugal force finally overcomes the inward motion and forces the ring back out again [*see bottom illustration on page 77*].

The ring will oscillate in and out for several hundred million years before reaching an equilibrium position. It first reaches the observed three-kiloparsec position 10 million years after the explosion; it returns to the same position 57 million years later and again 67 million years later (that is, 135 million years after the initial explosion). Thus the explosions that caused the various expanding features near the center of the galaxy may be occurring at intervals of 500 million years instead of five million years. This reduces the total energy requirement needed for explosions over the lifetime of the galaxy by a factor of 100 to 10^{60} ergs and the total mass ejected from the center to 10^9 solar masses. These reductions would make the total energy of the explosive events in the galaxy comparable to the total energy estimated for such events in the nuclei of Seyfert galaxies. In other words, the nucleus of our galaxy is periodically a Seyfert-galaxy nucleus, and by implication the same is true of the nuclei of other spiral galaxies.

This, of course, is all highly speculative, and there are other possible explanations for the observed three-kiloparsec arm. For example, the spiral arms of our galaxy are regions where the density of stars and gas is significantly higher than average. The extra gravitational attraction of the spiral arms might also excite the radial oscillation of a gas ring near three kiloparsecs. Further theoretical calculations should reveal if this mechanism is more plausible than the one we have analyzed.

Even if the three-kiloparsec arm were not created by an explosion, the inner expanding ring of dense clouds, which reveal their presence by molecular line radiation, may well have been caused by a recent explosion in or near the galactic center. The total energy needed for such an event is only about 10^{55} ergs (equivalent to the total conversion of 10 suns

SCIENCE/SCOPE

An automated telemetry and command station has been built by Hughes for Western Union Telegraph Company to provide telemetry processing spacecraft command and ranging for its Hughes-built Westar satellites. It will be in operation when the first of three Westars is launched this month. The station, first of its kind in the U.S. for domestic satellites, features a unique instant-readout display panel that shows the status of up to three satellites in orbit and requires only one engineer on duty at a time.

Microwave testing of a maritime satellite and its antenna system for the U.S. Navy is now under way in a new anechoic test chamber at Hughes. Several distinctive features enable the chamber to test the wide frequency range of the satellite's independent transponders in L-band, C-band, and UHF. The chamber's absorption materials vary from 26 to 48 inches in depth. It has seven antennas for measuring spacecraft systems and providing telemetry and command functions between spacecraft and systems test equipment. Three satellites are being built under contract with Comsat, the first scheduled for delivery next fall. They will provide communications for both the U.S. Navy and the maritime industry.

First of a new line of Hughes laser subsystems, aimed at the commercial market, is an ultra-ruggedized model providing high performance for such applications as construction alignment, pipe laying, surveying, and guidance systems for construction equipment. Similar design concepts, integrating laser components and optics, are under development for data processing and general systems markets.

Day and night cloud-cover pictures of the entire earth are now being beamed twice a day directly to nations around the world by the NOAA-3 satellite, new primary weather watcher of the National Oceanic and Atmospheric Administration's worldwide environmental satellite system. This is made possible by the scanning radiometer built by Santa Barbara Research Center, a Hughes subsidiary. The picures can be received in real time by 550 simple, inexpensive APT (Automatic Picture Transmission) stations located in 80 countries. NOAA-3 is the 24th in the series of weather satellites launched since 1960.

Hughes Research Laboratories has openings for semiconductor scientists with experience in device technology and processing. Duties will be in area of compound semiconductor devices, exploiting advanced materials technologies presently under development at Hughes. U.S. citizenship required. Please write: Mr. A. J. Simone, Hughes Research Laboratories, 3011 S. Malibu Canyon Rd., Malibu, CA 90265. An equal opportunity M/F employer.

A new series of axial heat pipes for highly efficient thermal control has been introduced by Hughes. Typical applications: electronics, chemical and mechanical equipment, industrial processing, medical equipment. The heat pipes use copper as the envelope and wick material and distilled water as the sealed-in working fluid. They have a thermal transport capability between 35 and 6,000 watts maximum over a recommended operating temperature range of +50° to +150°C.

into energy). The exploding object and the ejected mass would have to be 100,000 solar masses. Although such an explosion is much smaller than the one needed to produce the three-kiloparsec arm, it is nonetheless an extremely violent event.

Granting that violent events occur in the center of galaxies, where might the energy come from? One conceivable source, examined by several theorists, is a supermassive star: a star at least a million times more massive than the sun. Such a star would be 10,000 times heavier than any ever observed. If such a supermassive object could be formed in the galactic center, it would be unstable and would eventually collapse under the weight of its own gravitational attraction. As the star collapses it converts its gravitational energy into kinetic and thermal energy, and the collapsing star gets hotter and hotter. The thermonuclear reactions occurring inside the star are strongly dependent on the temperature; therefore under certain conditions (if the material comprising the star has more than a critical concentration of elements heavier than helium) a tremendous thermonuclear explosion could result.

If the conditions for the explosion are not met, however, the collapse of the star never stops and the end result is a black hole. A black hole is formed when matter becomes so dense that not even photons can escape from the grip of its gravitational attraction. Thus if a supermassive star could be formed in the galactic nucleus, the result would be either a great explosion or a black hole.

Two mechanisms have been proposed for the formation of supermassive stars. The first envisions a series of head-on stellar collisions. Detailed calculations show that the density of stars must be on the order of 100 million stars per cubic light-year (or three billion stars per cubic parsec) for the collisions to be frequent enough to create massive stars by coalescence. Since the required density is about 3,000 times the peak stellar density in the galactic nucleus, stellar collisions are an unlikely source of supermassive stars, at least in our galaxy.

The second mechanism proposed for creating a supermassive star invokes the gravitational accretion of gas. Although every star exerts a gravitational force on every other star, the strongest force felt by any single star is directed toward the galactic center. The stars are kept from falling into the center, however, by their high random velocity of about 100 kilometers per second. The gas that all stars emit in the course of their evolution also has a high random velocity; unlike the stars, however, the gas can lose its random energy, that is, it can cool by radiation. As a result the cooling gas will be slowly drawn into the center of the galaxy, where it could conceivably form a single massive object.

One way or another, to account for gigantic explosions in the center of the galaxy it is necessary to have a large concentration of mass in a very small volume of the galactic core. It is therefore important to determine the total distribution of mass in the very center of the galaxy to see if there is evidence for the existence of sufficient material to create an explosion capable of producing features such as the three-kiloparsec arm. From infrared observations we know that there are a million solar masses of ordinary stars present in the innermost one parsec of the galactic nucleus. Could there be much more mass there that was not in the form of ordinary stars but some other form we have not yet been able to observe directly?

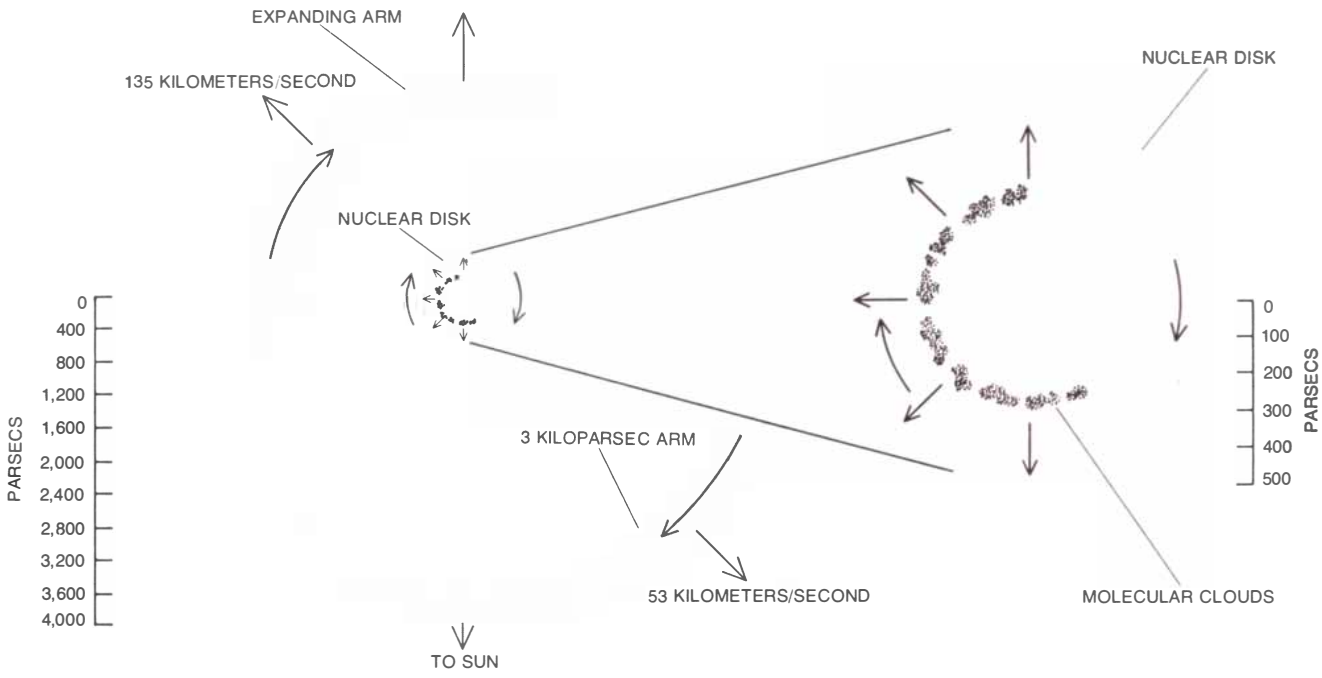
If there were a large unseen concentration of mass in the core of the galactic nucleus, the gas in the central part of the galaxy would be rapidly rotating around the center (so that the centrifugal force would be adequate to balance the strong gravitational force). This high rotational velocity would be observable in the Doppler shift of the 21-centimeter line emission of un-ionized hydrogen; therefore the 21-centimeter line profiles (the distribution of line intensity with gas velocity) in the direction of the center are an indication of the total mass concentration near the center.

We have made sensitive measurements of the 21-centimeter line in the direction of the galactic center, using the 140-foot radio telescope of the National Radio Astronomy Observatory in Green Bank, W.Va. We analyzed our data utilizing synthesized model galaxies on the computer in the following way. We simulate "observing" the model galaxy numerically in the same way we observe the real galaxy with the 140-foot telescope and generate artificial profiles of the 21-centimeter line. We then change the model galaxy to see how much mass we can add to the center before the artificial profiles no longer resemble the actual profiles. We find that the largest mass we can place in the galactic center without violating the stated condition is 200 million solar masses. Thus if there is a single condensed object in the center (either a supermassive star to explode or a black hole), its mass cannot exceed 200 million times the mass of the sun. We do not say that such a mass actually exists, only that its presence would not violate the observed gas motions.

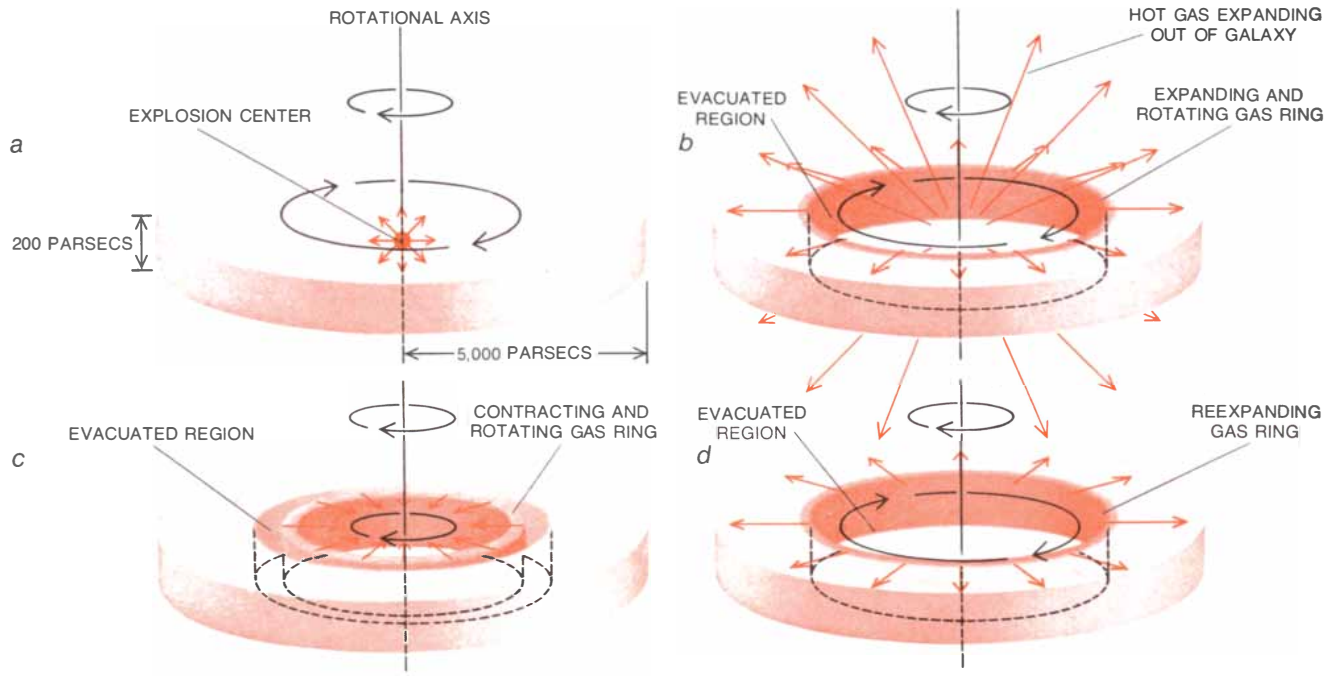
What are the consequences of this upper limit on the mass at the center? The mass is something less than twice the mass that must be expelled in an explosion energetic enough to create the three-kiloparsec arm. Therefore the galactic center contains at most only enough mass for one more titanic explosion. Either the next one will be the last or there must be a mechanism for replacing 100 million solar masses in the galactic center over a time scale of from 100 million to 500 million years. In other words, there must be a net flow of mass into the center on the order of one solar mass per year if superexplosions are to recur periodically into the future.

All the observational evidence points to an upper limit of perhaps two million solar masses (stars and gas combined) inside a core one parsec in diameter at the center of the galaxy. Moreover, there is no direct evidence for single objects with more than 100 times the mass of the sun. Therefore there are no stars massive enough to account for the relatively mild explosion needed to produce the 300-parsec expanding ring of molecular clouds (which requires an exploding object about 100,000 times the mass of the sun), let alone the three-kiloparsec arm.

There seems to be no way to avoid the conclusion that the core as we currently observe it may have been depleted of massive objects by the explosion a million years ago that created the 300-parsec expanding ring. If there is to be another such explosion within roughly a million years, matter must be flowing into the galactic center at the rate of about a tenth of a solar mass per year. The mechanism of this hypothetical flow is not known. The gas may flow into the central core from the galactic plane, or it may come from outside the galaxy. One provocative suggestion is that mass may spontaneously appear in the nuclei of all galaxies through some process that lies outside our present understanding of physics. One does not, however, tamper with the known laws of physics lightly. This alternative should only be considered if further theoretical and observational investigations reveal no other reasonable explanation of the events that are occurring in the active center of our galaxy.



CURRENT PICTURE OF GALACTIC CENTER is based on observed motions of neutral hydrogen and clouds containing assorted molecules (such as ammonia and formaldehyde). Four main features are depicted. There is a rotating and expanding arm of gas lying somewhat more than three kiloparsecs from the center, moving toward us at about 53 kilometers per second. On the other side, at a distance of roughly 2.5 kiloparsecs, another rotating arm is expanding at about 135 kilometers per second. Recent evidence suggests that both arms are parts of complete rings. The third kinematic feature is the nuclear disk, a rotating disk of gas extending about 500 parsecs from the center. The disk is evidently not expanding, but its rotation velocity decreases slowly inward from a maximum of about 250 kilometers per second at the edge. The last feature is a partial ring of molecular clouds inside the nuclear disk, some 300 parsecs from the center and apparently expanding at about 100 kilometers per second. Since its rotational velocity of 50 kilometers per second is insufficient to balance the gravitational force at its present location, the ring may have been ejected from a point closer to the center. The apparent absence of un-ionized hydrogen inside ring suggests that it was swept up as ring expanded.



SUPEREXPLOSION IN GALACTIC CORE would have the effects illustrated, according to computer studies made by the authors in collaboration with Kevin H. Prendergast at Columbia University. Gas intensely heated by the explosion expands in all directions (*a*). About 10 million years later (*b*) the gas expanding into the disk has swept up and compressed most of the gas present, forming an expanding and rotating ring about three kiloparsecs from the center. The ring expands beyond the point where its rotational velocity can balance the gravitational force pulling it back to the center, with the result that it begins to contract. Some 44 million years after the explosion (*c*) it has contracted to 2.5 kiloparsecs from the center. Now the ring's rotational velocity exceeds the equilibrium value and it starts to expand outward again. Sixty-seven million years after the explosion (*d*) the ring is again about three kiloparsecs from the center, moving outward at 50 kilometers per second. The ring may oscillate in this manner for several hundred million years.

Another small miracle: the first fully programmable pocket-sized computer calculator.

In January 1972, we introduced the first pocket-sized computer calculator, the HP-35. With transcendental functions, a full two-hundred-decade range and four register operational stack the HP-35 revolutionized the process of doing complex calculations.

In June 1973, we introduced the HP-45 designed for twice the computational power in the same size package.

Now we're introducing the HP-65, the first fully programmable hand-held computer calculator. Like its two antecedents the HP-65 makes single complex calculations automatically. In addition, because it is fully programmable, it can automatically perform a complete *sequence* of calculations to give you the answer. All you do is enter the raw data and a prerecorded program. You are relieved of remembering and executing the correct sequence of keystrokes manually.

The program is recorded on a magnetic card the size of a small stick of gum. Each program can consist of as many as 100 steps, and each step can be any of the calculator's 51 keystroke functions. These can be combined with branching, logical comparison and conditional skip instructions. The HP-65 also has a 4-register operational stack and nine addressable memory registers.

You can write, edit and record your own programs on the HP-65 keyboard. It couldn't be easier; you push the keys in the same sequence you would use to solve the problem manually, with the programming switch in the W/PRGM position. Then you record these steps on a small magnetic card. Next time you want to solve the same problem turn the switch to RUN and insert the card.

If you don't want to write your own programs, use ours. Six prerecorded packages, called Application Pacs, are available now; more are on the way. Each Pac contains up to 40 programs in disciplines such as mathematics, statistics, surveying, medicine, and electrical engineering. In the Math Pac II, for example, there are prerecorded programs for base conversion, complete elliptic integrals, intersection of a line and a conic section, as well as 30 other calculations. All Pacs available for $45.*

The HP-65 is priced at $795.* This includes one Application Pac, one Pac of blank program cards, handbook, battery pack, AC adapter/recharger, carrying case and travel case plus a one-year subscription to the User Library Catalog.

The new Logic Lab: a practical way to learn digital electronics hands-on.

Like many other aspects of modern life, electronics isn't what it used to be. With practically every new electronic product introduction the change becomes clearer: the trend is toward digital and away from analog. And the whole scene is confusing until you learn what the 1's and 0's of digital circuits are all about. The new HP Logic Lab clears away the confusion; in many cases, it can help you learn all you'll ever need to know about digital circuits. Designed for hands-on introduction, the Logic Lab includes a textbook and lab workbook as well as a fully equipped portable laboratory for digital experimentation.

The student learns to assemble a variety of functioning digital circuits using the breadboard, components and inter-connecting wires supplied. This takes a minimum of effort because the Logic Lab utilizes a solderless, plug-in connection technique for components and wires. No longer is assembly a gruelling process of soldering and unsoldering IC's and wires.

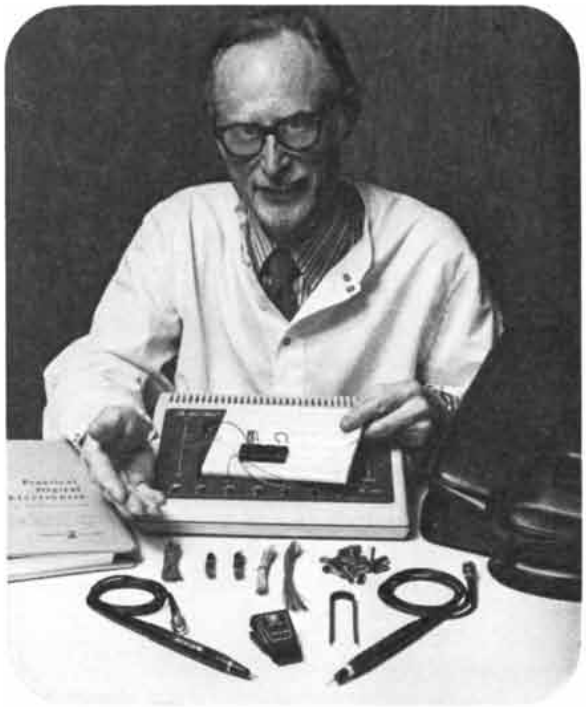
The laboratory station mainframe has the essentials built-in (power supply, logic state indicators and programmers, and pulse sources to provide active stimulus for the student's circuits). When the student has assembled a circuit, he can verify its operation using the professional-quality digital test instruments — the HP Logic Probe, Logic Pulser and Logic Clip. In doing so he not only learns how to troubleshoot a digital circuit, but also sees visual evidence of circuit component behavior and can make sense of their interactions with other components. In the final analysis, this is where he learns what the 1's and 0's mean.

The Logic Lab is ready for the next class, even if the student isn't. Caught by the bell in the middle of assembly or testing, the student's breadboard is easily removed and replaced by another one without disturbing the circuits.

The Logic Lab is an effective self-directed learning package. Written by a team of experienced educators and industry engineers, all text is arranged in a modular format that allows the student to start at a level that suits his individual needs. The instructor can involve himself to the extent he desires.

Not just for students, engineers also find the Logic Lab a boon in breadboarding, analyzing and debugging new designs.

The complete Logic Lab package is available for $650.*



For more information on these products write to us. Hewlett-Packard, 1503 Page Mill Road, Palo Alto, California 94304.

*Domestic USA prices only. 00407

Mail to: Hewlett-Packard, 1503 Page Mill Road, Palo Alto, CA 94304
Please send me information on:
( ) HP 65 Computer/Calculator
( ) HP Logic Lab

Name_____ Title_____
Company_____
Address_____
City_____ State_____ Zip_____

HEWLETT hp PACKARD

Sales and service from 172 offices in 65 countries.
Palo Alto, California 94304

Plate Tectonics and the History of Life in the Oceans

The breakup of the ancient supercontinent of Pangaea triggered a long-term evolutionary trend that has led to the unprecedented variety of the present biosphere

by James W. Valentine and Eldridge M. Moores

During the 1960's a conceptual revolution swept the earth sciences. The new world view fundamentally altered long established notions about the permanency of the continents and the ocean basins and provided fresh perceptions of the underlying causes and significance of many major features of the earth's mantle and crust. As a consequence of this revolution it is now generally accepted that the continents have greatly altered their geographic position, their pattern of dispersal and even their size and number. These processes of continental drift, fragmentation and assembly have been going on for at least 700 million years and perhaps for more than two billion years.

Changes of such magnitude in the relative configuration of the continents and the oceans must have had far-reaching effects on the environment, repatterning the world's climate and influencing the composition and distribution of life in the biosphere. These more or less continual changes in the environment must also have had profound effects on the course of evolution and accordingly on the history of life.

Natural selection, the chief mechanism by which evolution proceeds, is a very complex process. Although it is constrained by the machinery of inheritance, natural selection is chiefly an ecological process based on the relation between organisms and their environment. For any species certain heritable variations are favored because they are particularly well suited to survive and to reproduce in their prevailing environment. To answer the question of why any given group of organisms has evolved, then, one needs to understand two main factors. First, it is necessary to know what the ancestral organisms were that formed the "raw material" on which selection worked. And second, one must have some idea of the sequence of environmental conditions that led the ancestral stock to evolve along a particular pathway to a descendant group. Given these factors, one can then infer the organism-environment interactions that gave rise to the evolutionary events. The study of the relations between ancient organisms and their environment is called paleoecology.

The new ideas of continental drift that came into prominence in the 1960's revolve around the theory of plate tectonics. According to this theory, new sea floor and underlying mantle are currently being added to the crust of the earth at spreading centers under deep-sea ridges and in small ocean basins at rates of up to 10 centimeters per year. The sea floor spreads laterally away from these centers and eventually sinks into the earth's interior at subduction zones, which are marked by deep-sea trenches. Volcanoes are created by the consumption process and flank the trenches. The lithosphere, or rocky outer shell of the earth, therefore comprises several major plates that are generated at spreading centers and consumed at subduction zones. Most lithospheric plates bear one continent or more, which passively move with the plate on which they rest. Because the continents are too light to sink into the trenches they remain on the surface. Continents can fragment at new ridges, however, and hence oceans may appear across them. Conversely, continents can be welded together when they collide at the site of a trench. Thus continents may be assembled into supercontinents, fragmented into small continents and generally moved about the earth's surface as passive riders on plates. In tens or hundreds of millions of years entire oceans may be created or destroyed, and the number, size and dispersal pattern of continents may be vastly altered.

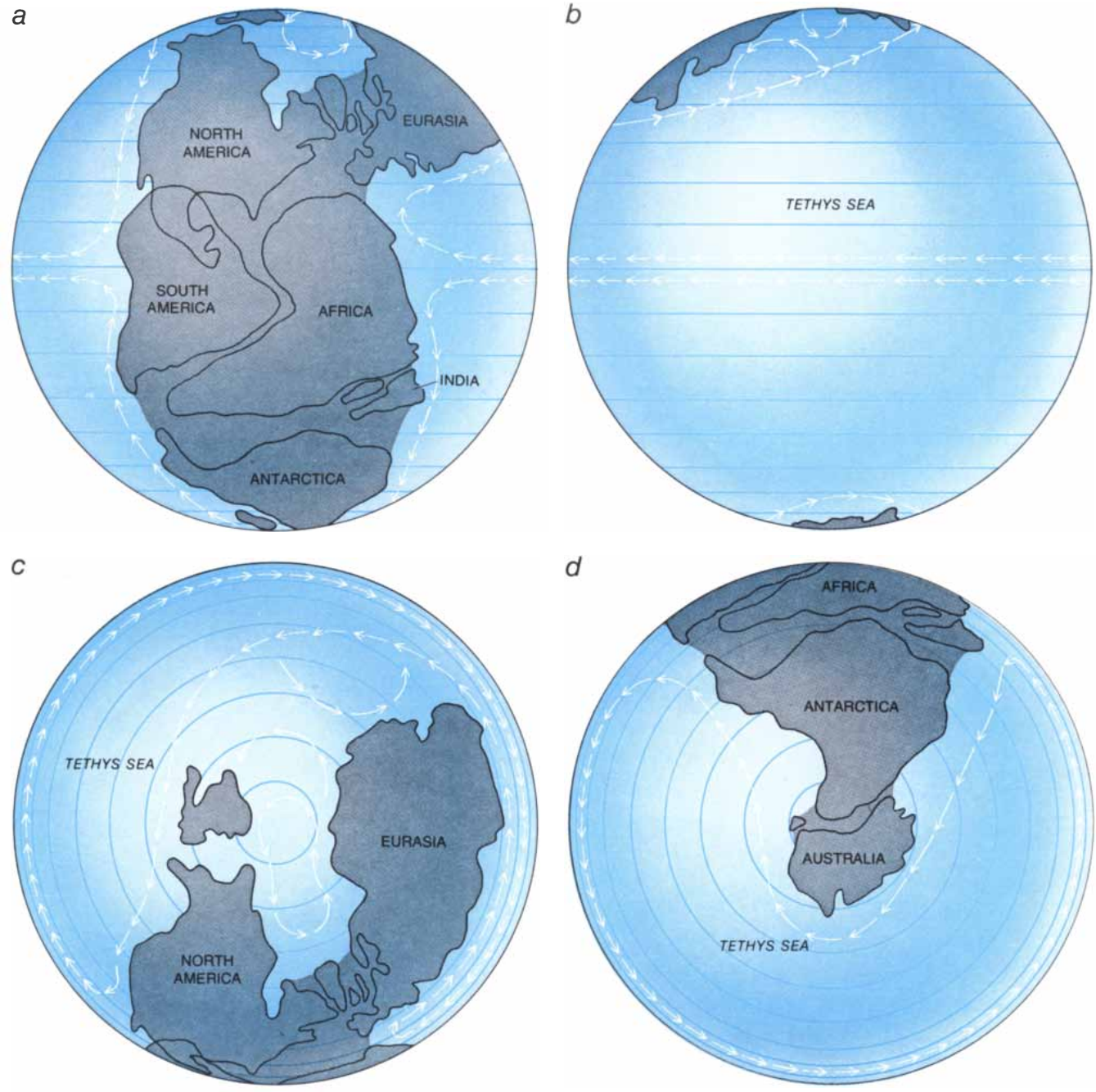
The record of such continental fragmentation and reassembly is evident as deformed regions in the earth's mountain belts, particularly those mountain belts that contain the rock formations known as ophiolites. These formations are characterized by a certain sequence of rocks consisting (from bottom to top) of ultramafic rock (a magnesium-rich rock composed mostly of olivine), gabbro (a coarse-grained basaltic rock), volcanic rocks and sedimentary rocks. The major ophiolite belts of the earth are believed to represent preserved fragments of vanished ocean basins [*see illustration on pages 82 and 83*]. The existence of such a belt within a continent (for example the Uralian belt in the U.S.S.R.) is evidence for the former presence there of an ocean basin separating two continental fragments that at some time in the past collided with each other and were welded into the single larger continent. The timing of such events as the opening of ocean basins, the dispersal of continents and the closing of oceans by continental collisions can accordingly be "read" from the geology of a given mountain system.

Of course, the biological environment is constantly being altered as well. For example, the changes in continental configuration will greatly affect the ocean currents, the temperature, the nature of seasonal fluctuations, the distribution of nutrients, the patterns of productivity and many other factors of

fundamental importance to living organisms. Therefore evolutionary trends in marine animals must have varied through geologic time in response to the major environmental changes, as natural selection acted to adapt organisms to the new conditions.

It should in principle be possible to detect these changes in the fossil record. Indeed, paleontologists have long recognized that vast changes in the composition, distribution and diversity of marine life are well documented by the fossil record. Now for the first time, however, it is possible to reconstruct the sequence of environmental changes based on the theory of plate tectonics, to determine their environmental consequences and to attempt to correlate them with the sequence of faunal changes that is seen in the fossil record. Such a thorough reconstruction ultimately may explain many of the enigmatic faunal changes known for many years. Even at this early stage paleontologists have succeeded in shedding much new light on a number of major extinctions and diversifications of the past.

As a first step toward understanding the relation between plate tectonics and the history of life it is helpful to investigate the relations that exist today between marine life, the present pattern of continental drift and plate-tectonic theory. The vast majority of marine species (about 90 percent) live on the continental shelves or on shallow-water portions of islands or subsurface "rises" at depths of less than about 200 meters

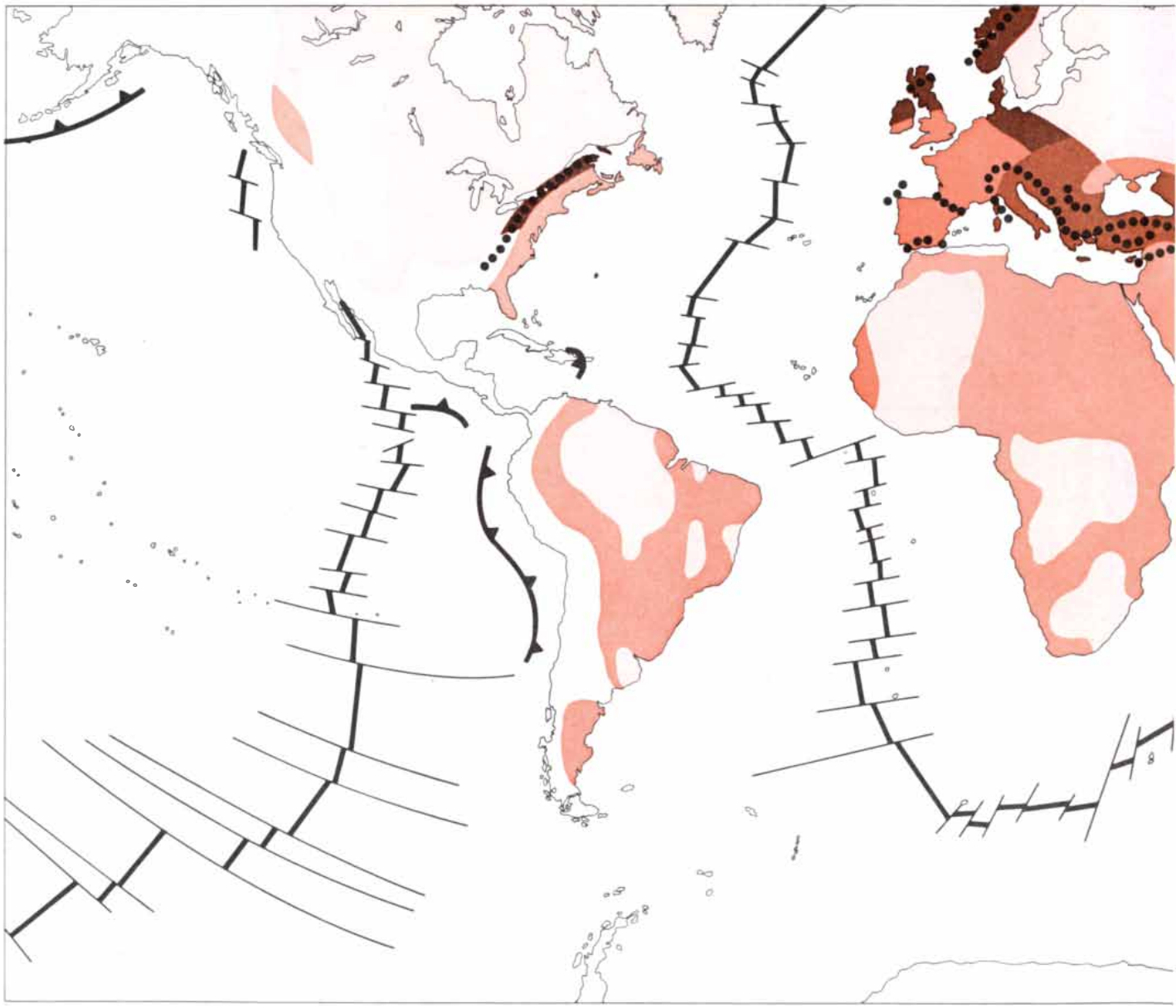


ANCIENT OCEAN CURRENTS in the vicinity of Pangaea, the single "supercontinent" that is believed to have existed near the beginning of the Triassic period some 225 million years ago, are indicated here in two equatorial views (*a*, *b*) and two polar views (*c*, *d*). Owing to a combination of geographic and environmental factors, including the predominantly warm-water currents shown, one would expect the continuous shallow-water margin that surrounded Pangaea to have been populated by comparatively few but widespread species. Such low species diversity combined with low provinciality is precisely what the fossil record indicates.

(660 feet); most of the fossil record also consists of these faunas. Therefore it is the pattern of shallow-water sea-floor animal life that is of particular interest here.

The richest shallow-water faunas are found today at low latitudes in the Tropics, where communities are packed with vast numbers of highly specialized species. Proceeding to higher latitudes, diversity gradually falls; in the Arctic or Antarctic regions less than a tenth as many animals are living as in the Tropics, when comparable regions are considered [*see illustration on pages 84 and 85*]. The diversity gradient correlates well with a gradient in the stability of food supplies; as the seasons become more pronounced, fluctuations in primary productivity become greater. Although this strong latitudinal gradient dominates the earth's overall diversity pattern, there are important longitudinal diversity trends as well. In regions of similar latitude, for example, diversity is lower where there are sharp seasonal changes (such as variations in the surface-current pattern or in the upwelling of cold water) that affect the nutrient supply by causing large fluctuations in productivity.

At any given latitude, therefore, diversity is highest off the shores of small islands or small continents in large oceans, where fluctuations in nutrient supplies are least affected by the seasonal effects of landmasses, whereas diversity is lowest off large continents, particularly when they face small oceans,



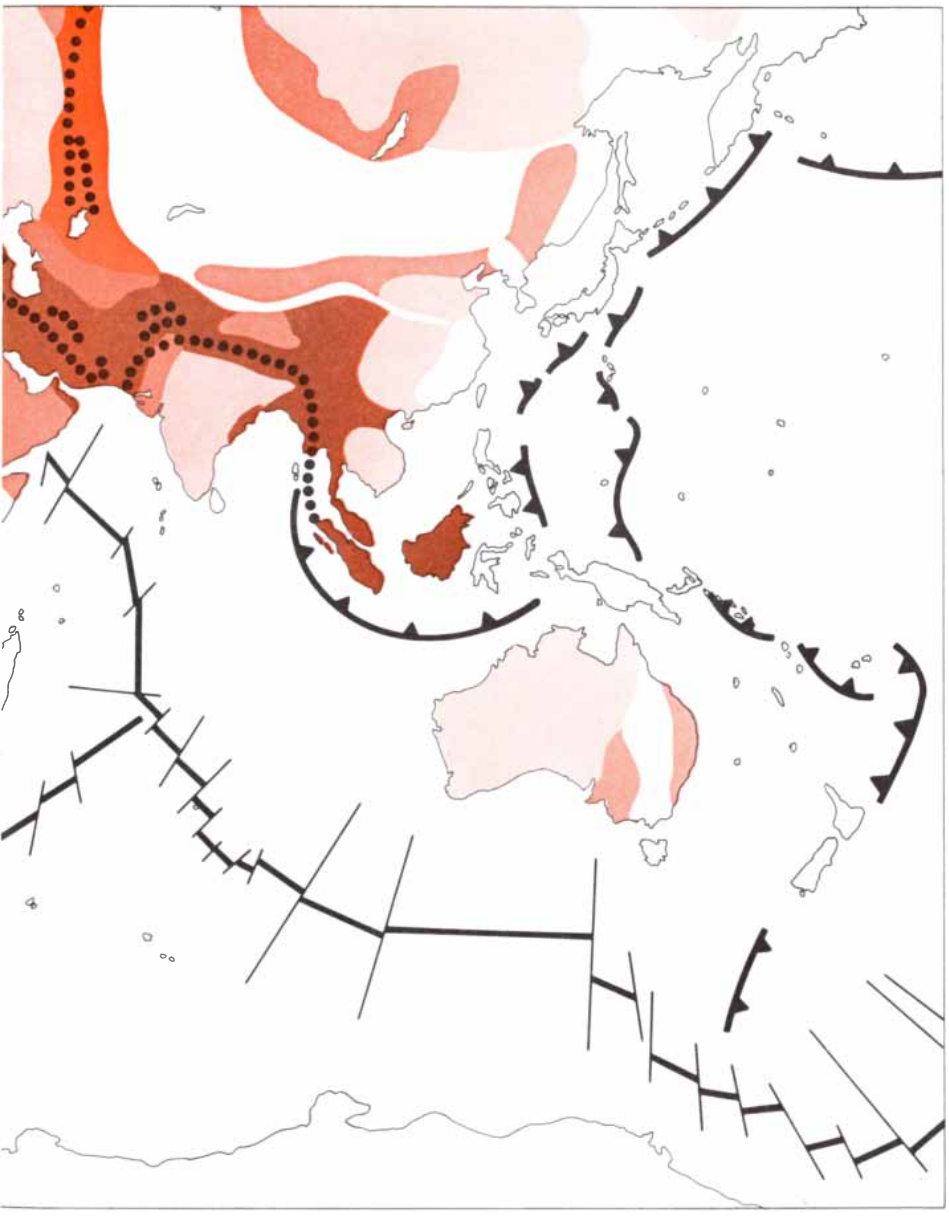
GEOLOGICAL RECORD of ancient plate-tectonic activity is preserved in certain deformed mountain belts (*color*), particularly those that contain the characteristic rock sequences known as ophiolites (*black dots*). The Pan-African–Baikalian belt, for example, is made up of rocks dating from 873 to 450 million years ago and may represent the assembly of all or nearly all the landmasses near the beginning of Phanerozoic time. This supercontinent may then have fragmented into four or more smaller continents, sometime just before and during the Cambrian period. The Caledonian mountain system may represent the collision of two continents at about late Silurian or early De-

where shallow-water seasonal variations are greatest. In short, whereas latitudinal diversity increases generally from high latitudes to low, longitudinal diversity increases generally with distance from large continental landmasses. In both of these trends the increase in diversity is correlated with increasing stability of food resources. The resource-stability pattern depends largely on the shape of the continents and should also be sensitive to the extent of inland seas and to the presence of coastal mountains. Seas lying on continental platforms are particularly important: not only do extensive shallow seas provide much habitat area for shallow-water faunas but also such seas tend to damp seasonal climatic changes and to have an ameliorating influence on the local environment.

Today shallow marine faunas are highly provincial, that is, the species living in different oceans or on opposite sides of the same ocean tend to be quite different. Even along continuous coastlines there are major changes in species composition from place to place that generally correspond to climatic changes. The deep-sea floor, generated at oceanic ridges, forms a significant barrier to the dispersal of shallow-water organisms, and latitudinal climatic changes clearly form other barriers. The present dominantly north-south series of ridges forms a pattern of longitudinally alternating oceans and continents, thereby creating a series of barriers to shallow-water marine organisms. The steep latitudinal climatic gradient, on the other hand, creates chains of provinces along north-south coastlines. As a result the marine faunas today are partitioned into more than 30 provinces, among which there is in general only a low percentage of common species [*see illustration on pages 86 and 87*]. It is estimated that the shallow-water marine fauna represents more than 10 times as many species today as would be present in a world with only a single province, even a highly diverse one.

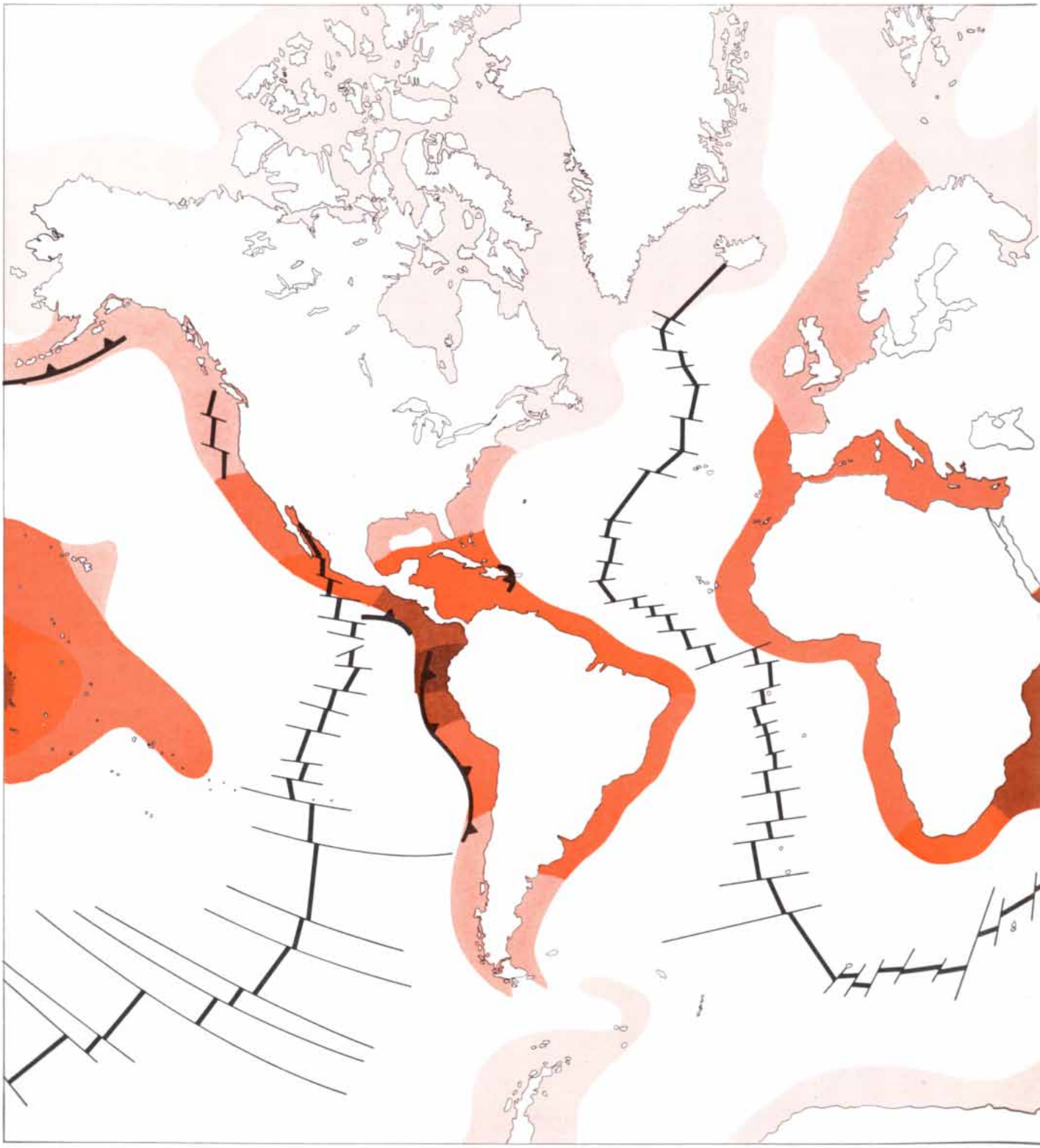
The volcanic arcs that appear over subduction zones form fairly continuous island chains and provide excellent dispersal routes. When long island chains are arranged in an east-west pattern so as to lie within the same climatic zone, they are inhabited by wide-ranging faunas that are highly diverse for their latitude. Indeed, the widest ranging marine province, and also by far the most diverse, is the Indo-Pacific province, which is based on island arcs in its central regions. The faunal life of this province spills from these arcs onto tropical continental shelves in the west (India and East Africa) and also onto tropical intraplate volcanoes (the Polynesian and Micronesian islands) that are reasonably close to them. This vast tropical biota is cut off from the western American mainland by the East Pacific Barrier, a zoogeographic obstruction formed by a spreading ridge.

Since current patterns of marine provinciality and diversity fit closely with the present oceanic and continental geography and the resulting environmental patterns, one would expect ancient provinces and ancient diversity patterns also to fit past geographies. One of the best-established of ancient geographies is the one that existed near the beginning of the Triassic period, about 225 million years ago. The continents were then assembled into a single supercontinent named Pangaea, which must have had a continuous shallow-water margin running all the way around it, with no major physical barriers to the dispersal of shallow-water marine ani-



vonian time (approximately 400 million years ago). The Appalachian-Hercynian system may represent a two-continent collision during the late Carboniferous period (300 million years ago). The Uralian mountains may represent a similar collision at about Permo-Triassic time (220 million years ago). The Cordilleran-Tethyan system represents regions of Mesozoic mountain-building and includes the continental collisions that resulted in the Alpine-Himalayan mountain system. The ophiolite belts shown are the preserved remnants of ocean floor exposed in the mountain systems in question. Spreading ridges such as the Mid-Atlantic Ridge are indicated by heavy lines cut by lighter lines, which correspond to transform faults. Subduction zones are marked by heavy black curved lines with triangles.

mals [*see illustration on page 81*]. Therefore provinciality must have been low compared with today, and it must have been attributable entirely to climatic effects. It is likely that the marine climate was quite mild and that even in high latitudes water temperatures were much warmer than they are today. As a result climatic provinciality must have been greatly reduced also. Furthermore, the seas at that time were largely confined to the ocean basins and did not extend significantly over the continental shelves. Thus the habitat area for shallow-water marine organisms was greatly reduced, first by the diminution of coastline that accompanies the creation of a supercontinent from smaller continents, and second by the general withdrawal of seas from continental platforms. The reduced habitat area would make for low species

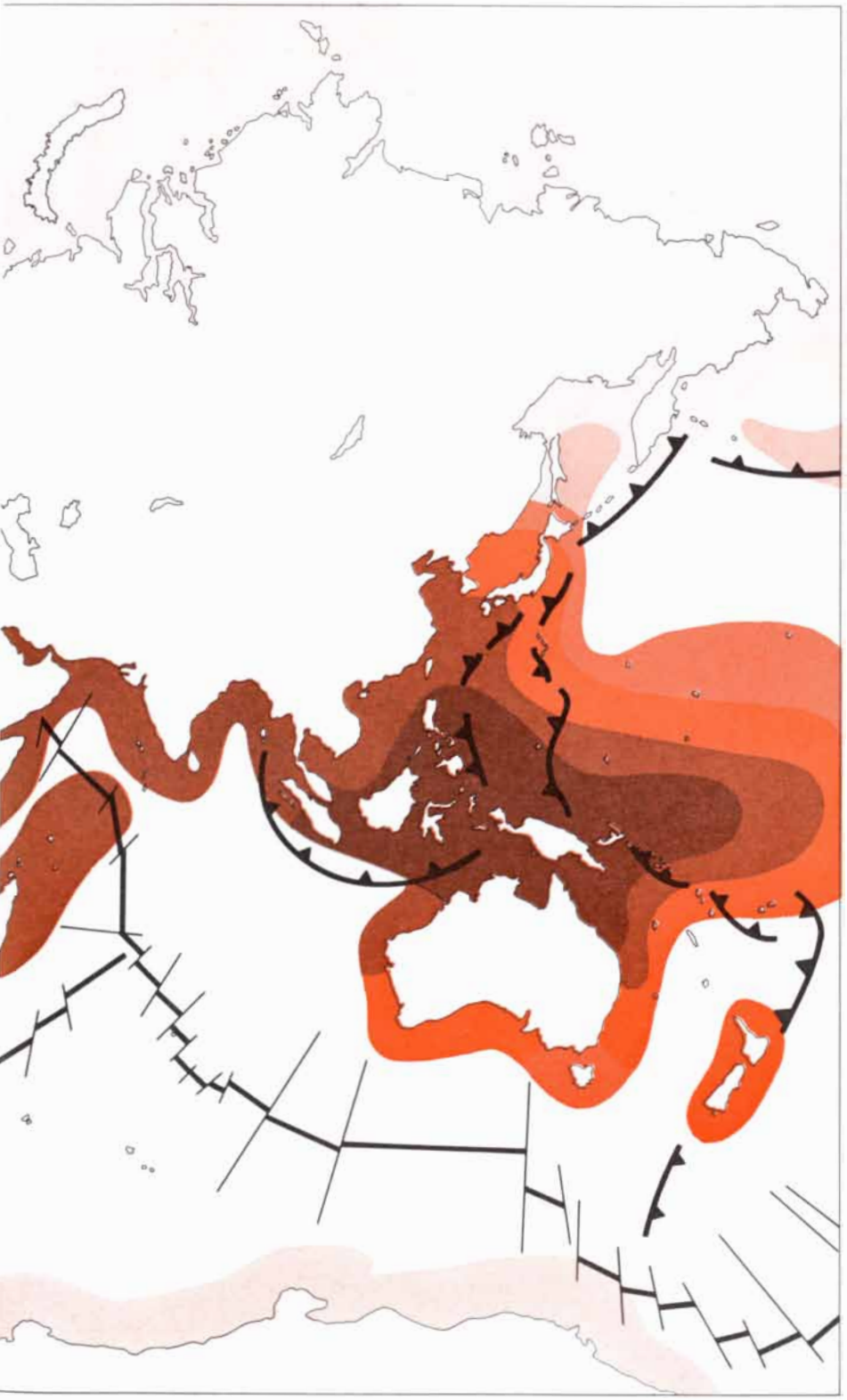


RELATIVE DIVERSITY of shallow-water, bottom-dwelling species in the present oceans is suggested by the colored patterns in this world map. The diversity classes are not based on absolute counts but are inferred from the diversity patterns of the best-

diversity. Finally, the extreme emergence of such a supercontinent would provide unstable nearshore conditions, with the result that food resources would have been very unstable compared with those of today. All these factors tend to reduce species diversity; hence one would expect to find that Triassic biotas were widespread and were made up of comparatively few species. That is precisely what the fossil record indicates.

Prior to the Triassic period, during the late Paleozoic, diversity appears to have been much higher [*see top illustration on page 89*]. It was sharply reduced again near the close of the Permian period during a vast wave of extinction that on balance is the most severe known to have been suffered by the marine fauna. The late Paleozoic species that were the more elaborately adapted specialists became virtually extinct, whereas the surviving descendants tended to have simple skeletons. A high proportion of these survivors appear to have been detritus feeders or suspension feeders that harvested the water layers just above the sea floor. These successful types seem to be ecologically similar to the populations found today in unstable environments, for instance in high latitudes; the unsuccessful specialists, on the other hand, seem ecologically similar to the populations found in stable environments, for instance in the Tropics. Thus the extinctions appear to have been caused by the reduced potential for diversity of the shallow seas, a trend associated with less provinciality, less habitat area and less stable environmental conditions.

In the period following the great extinction, as Pangaea broke up and the resulting continents themselves gradually fragmented and migrated to their present positions, provinciality increased, communities in stabilized regions became filled with numerous specialized animals and the overall diversity of species in the world ocean rose to unprecedented heights, even though occasional waves of extinctions interrupted this long-range trend.

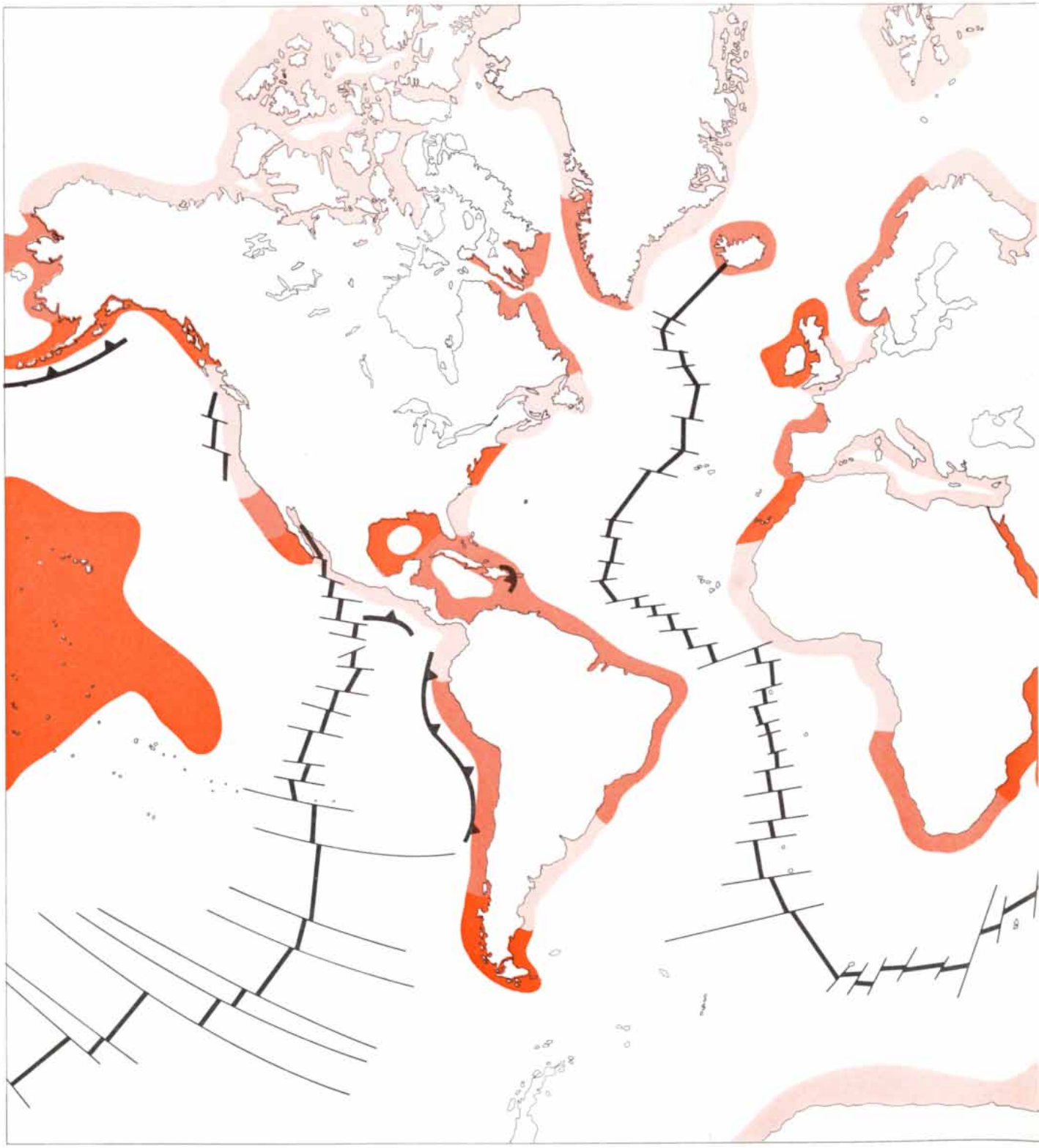


known skeletonized groups, chiefly the bivalves, gastropods, echinoids and corals. The highest class (*darkest color*) is about 20 times as diverse as the lowest (*lightest color*).

There is another time in the past besides the early Triassic period when low provinciality and low diversity were coupled with the presence of a high proportion of detritus feeders and near-bottom suspension feeders. That is in the late Precambrian and Cambrian periods, when a widespread, soft-bodied fauna of low diversity gave way to a slightly provincialized, skeletonized fauna of somewhat higher diversity. It seems likely that the late Precambrian environment was quite unstable and that there may well have been a supercontinent in existence, or at least that the continents then were collected into a more compact assemblage than at present. In the late Precambrian period one finds the first unequivocal records of invertebrate life, including burrowing forms that were probably coelomic, or hollow-bodied, worms. In the Cambrian four continents may have existed although they were not arranged in the present pattern. During the Cambrian a skeletonized fauna appears that is at first almost entirely surface-dwelling and that includes chiefly

detritus-feeding and suspension-feeding forms, probably with some browsers.

It seems possible, therefore, that the late Precambrian species were adapted to highly unstable conditions and became diversified chiefly as a bottom-living, detritus-feeding assemblage. The coelomic body cavity, evidently a primitive adaptation for burrowing, was developed and diversified into a variety of forms, perhaps as many as five basic ones: highly segmented worms that lived under the ocean floor and were detritus feeders; slightly segmented worms that lived attached to the ocean floor and were suspension feeders; slightly segmented worms that lived attached to the ocean floor and were detritus feeders; "pseudosegmented" worms that lived on the ocean floor and were detritus feeders or browsers, and nonsegmented worms



PRINCIPAL SHALLOW-WATER MARINE PROVINCES at present are indicated by the colored areas. The dominant north-south chains of provinces along the continental coastlines are created by the present high latitudinal gradient in ocean temperature

that lived under the ocean floor and fed by means of an "introvert." In addition to these coelomates there were a number of coelenterate stocks (such as corals, sea anemones and jellyfishes) and probably also flatworms and other noncoelomate worms.

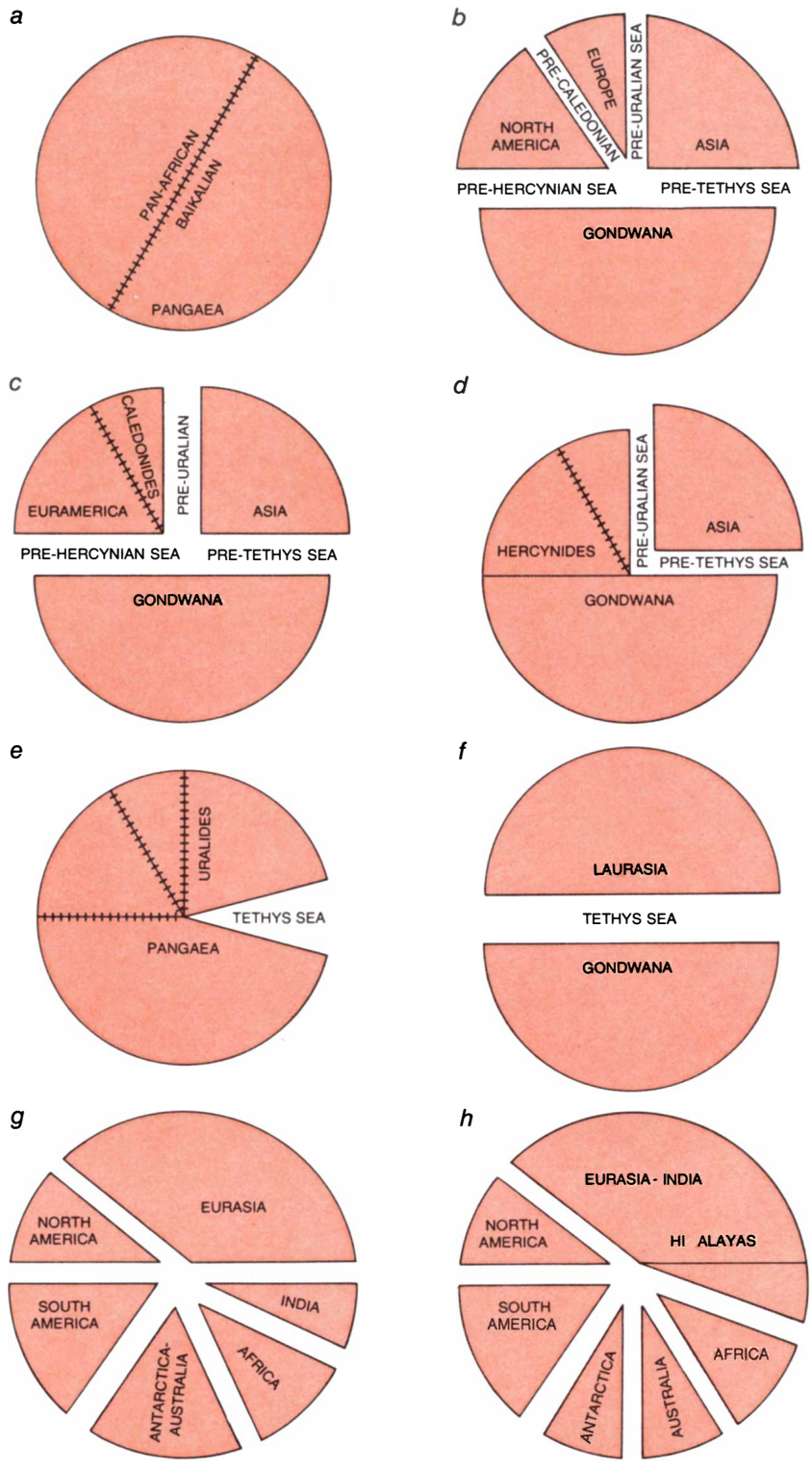
From the chiefly wormlike coelomate stocks higher forms of animal life have originated; many of them appear in the Cambrian period, when they evidently first became organized into the groups that characterize them today. Animals with skeletons appeared in the fossil record at that time. Presumably the invasion of the sea-floor surface by coelomates and the origin of numerous skeletonized species accompanied a general amelioration of environmental conditions as the continents became dispersed; the skeletons themselves can be viewed as adaptations required for worms to lead various modes of life on the surface of the sea floor rather than under it. The sudden appearance of skeletons in the fossil record therefore is associated with a generalized elaboration of the bottom-dwelling members of the marine ecosystem. Later, free-swimming and underground lineages developed from the skeletonized ocean-floor dwellers, with the result that skeletons became general in all marine environments.

and by the undersea barriers formed by spreading ridges. The vast Indo-Pacific province (*darkest color*) spills out onto scattered islands as indicated. There are 31 provinces shown.

The correlation of major events in the history of life with major environmental changes inferred from plate-tectonic processes is certainly striking. Even though details of the interpretation are still provisional, it seems certain that further work on this relation will prove fruitful. Indeed, the ability of geologists to determine past continental geographies should provide the basis for reconstructing the historical sequence of global environmental conditions for the first time. That sequence can then be compared with the sequence of organisms revealed in the fossil record. The following tentative account of such a comparison, on the broadest scale and without detail, will indicate the kind of history that is emerging; it is based on the examples reviewed above and on similar considerations.

Before about 700 million years ago bottom-dwelling, multicellular animals had developed that somewhat resembled flatworms. As yet no fossil evidence for their evolutionary pathways exists, but evidence from embryology and comparative anatomy suggests that they arose from swimming forms, possibly larval jellyfish, which in turn evolved from primitive single-celled animals.

Approximately 700 million years ago, perhaps in response to the onset of fluctuating environmental conditions brought about by continental clustering, a true coelomic body cavity was evolved to act as a hydrostatic skeleton in roundworms; this adaptation allowed burrowing in soft sea floors and led to the diversification of a host of worm architectures as that mode of life was explored. Burrows of this type are still preserved in some late Precambrian rocks. As the environment later became more stable, several of the worm lineages evolved more varied modes of life. The changes



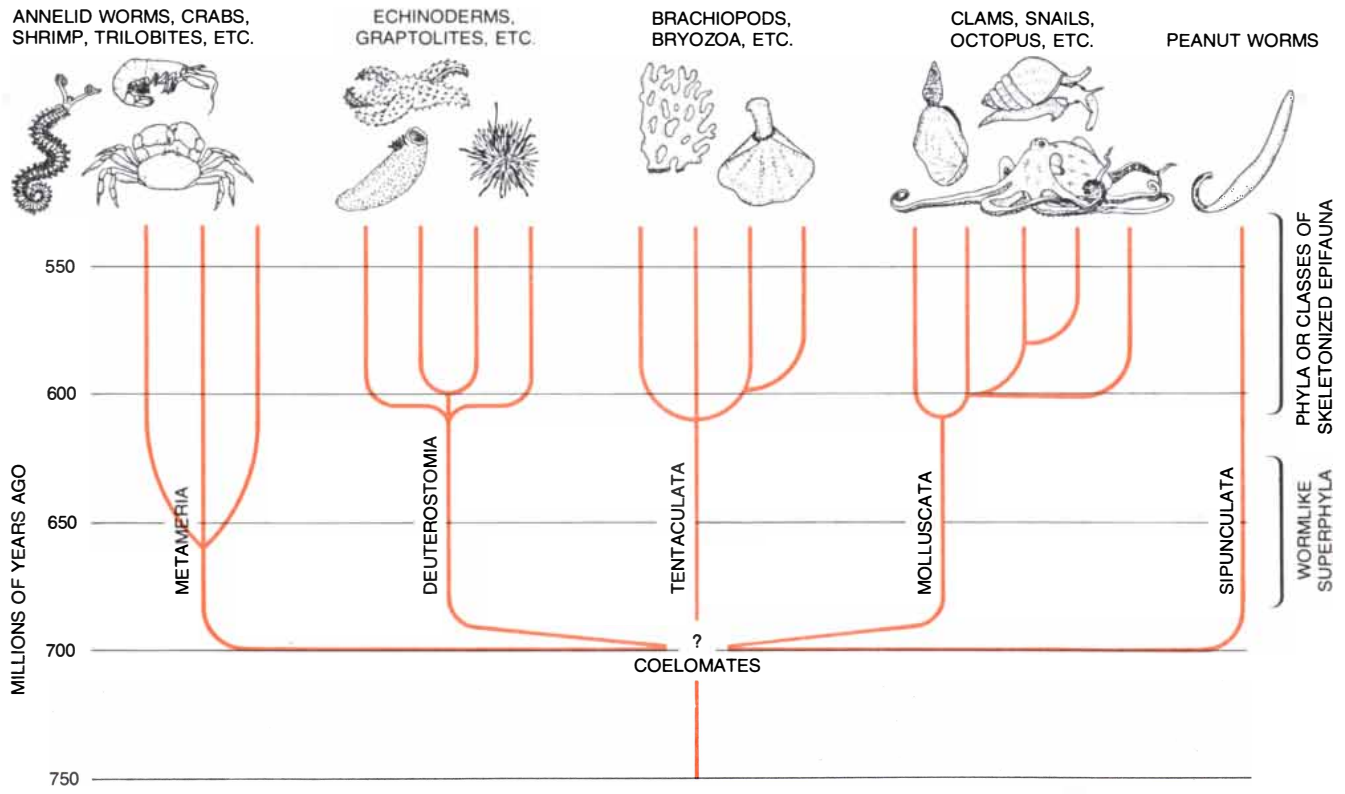
SIMPLIFIED DIAGRAMS are employed to suggest the relative configuration of the continents and the oceans during the past 700 million years. The late Precambrian supercontinent (*a*), which probably existed some 700 million years ago, may have been formed from previously separate continents. The Cambrian world (*b*) of about 570 million years ago consisted of four continents. The Devonian period (*c*) of about 390 million years ago was distinguished by three continents following the collapse of the pre-Caledonian Ocean and the collision of ancient Europe and North America. In the late Carboniferous period (*d*), about 300 million years ago, Euramerica became welded to Gondwana along the Hercynian belt. In the late Permian period (*e*), about 225 million years ago, Asia was welded to the remaining continents along the Uralian belt to form Pangaea. In early Mesozoic time (*f*), about 190 million years ago, Laurasia and Gondwana were more or less separate. In the late Cretaceous period (*g*), about 70 million years ago, Gondwana was highly fragmented and Laurasia partially so. The present continental pattern (*h*) shows India welded to Eurasia.

in body plan necessary to adapt to such a life commonly involved the development of a skeleton. There were evidently three or four main types of worms that are represented by skeletonized descendants today. One type was highly segmented like earthworms, and presumably burrowed incessantly for detrital food; these were represented in the Cambrian period by the trilobites and related species. A second type was segmented into two or three coelomic compartments and burrowed weakly for domicile, afterward filtering suspended food from the seawater just above the ocean floor; these evolved into such forms as brachiopods and bryozoans. A third type consisted of long-bodied creepers with a series of internal organs but without true segmentation; from these the classes of mollusks (such as snails, clams and cephalopods) have descended. Probably a fourth type consisted of unsegmented burrowers that fed on surface detritus and gave rise to the modern sipunculid worms. These may also have given rise to the echinoderms (which include the sea cucumber and the spiny sea urchin), and eventually to the chordates and to man. Although the lines of descent are still uncertain among these primitive and poorly known groups, the adaptive steps are becoming clearer.

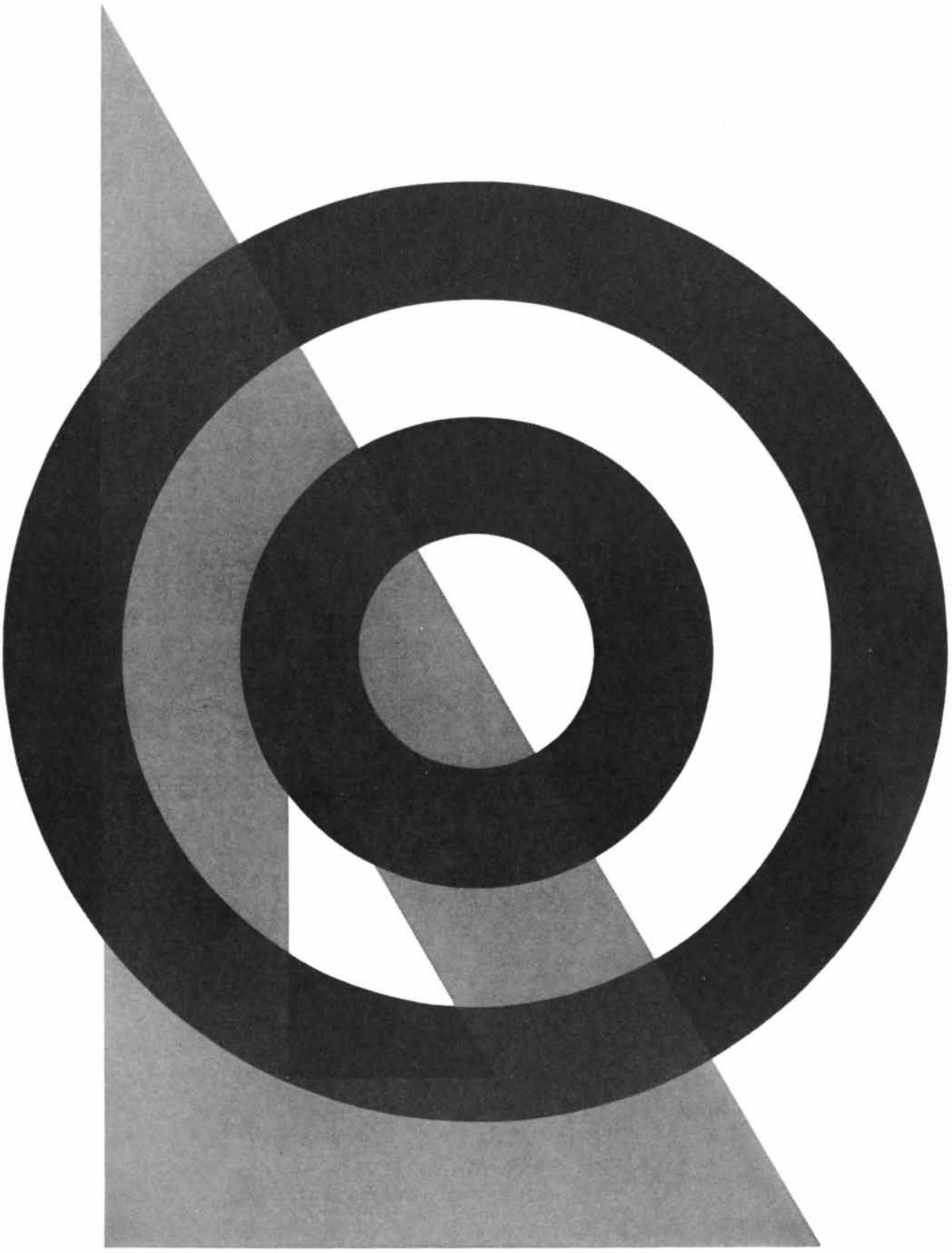
The major Cambrian radiation of the underground species into sea-floor surface habitats established the basic evolutionary lineages and occupied the major marine environments. Further evolutionary episodes tended to modify these basic animals into more elaborate structures. After the Cambrian period shallow-water marine animals became more highly specialized and richer in species, suggesting a continued trend toward resource stabilization. Suspension feeders proliferated and exploited higher parts of the water column, and predators also became more diversified. This trend seems to have reached a peak (or perhaps a plateau) in the Devonian period, some 375 million years ago. The characteristic Paleozoic fauna was finally swept away during the reduction in diversity that accompanied the great Permian-Triassic extinctions. Thus the rise of the Paleozoic fauna accompanied an amelioration in environmental conditions and increased provinciality, whereas the decline of the fauna accompanied a reestablishment of severe, unstable conditions and decreased provinciality. The subsequent breakup and dispersal of the continents has led to the present biosphere.

Today we live in a highly diverse world, probably harboring as many species as have ever lived at any time, associated in a rich variety of communities and a large number of provinces, probably the richest and largest ever to have existed at one time. We have been furnished with an enviably diverse and interesting biosphere; it would be a tragedy if we were to so perturb the environment as to return the biosphere to a low-diversity state, with the concomitant extinction of vast arrays of species. Of course, natural processes might eventually recoup the lost diversity, if we waited patiently for perhaps a few tens of millions of years. Alternatively we can work to preserve the environment in its present state and therefore to preserve the richness and variety of nature.

FLUCTUATIONS in the number of families, and hence in the level of diversity, of well-skeletonized invertebrates living on the world's continental shelves during the past 570 million years are plotted by geologic epoch in this graph. Time proceeds upward.



PHYLOGENETIC MODEL of the evolution of coelomate, or hollow-bodied, marine organisms is based on inferred adaptive pathways. The late Precambrian lineages were chiefly worms, which gave rise to epifaunal (bottom-dwelling), skeletonized phyla during the Cambrian period. The organisms depicted in the drawings at top are modern descendants of the major Cambrian lineages.



TRIANGLE IS PERCEIVED as being transparent and on top of the black and white concentric circles even though no elements of the illustration actually are physically transparent. The triangle is a mosaic composed of individual light gray and dark gray sections.

The Perception of Transparency

Certain mosaics of opaque colors and shapes give rise to the impression of transparency. A simple theoretical model predicts the conditions under which perceptual transparency will occur

by Fabio Metelli

What do we mean when we say that something is transparent? Actually the term has two meanings. If we are referring to the fact that light can pass through a thing or a medium, then the meaning of "transparent" we intend to convey is physical; if, on the other hand, we mean to say that we can see through something, then the meaning we intend to convey is perceptual. The distinction would not be very important if physical and perceptual transparency were always found together. Such, however, is not the case. Air is physically transparent, but normally we do not speak of "seeing through" it. Nor do we always perceive plate glass doors, since we occasionally run into them. It seems useful, therefore, to give a more precise definition of the perception of transparency: One perceives transparency when one sees not only surfaces behind a transparent medium but also the transparent medium or object itself. According to this definition, air and plate glass are not perceptually transparent unless there is fog in the air or there are marks or reflections on the glass.

The fact that physical transparency is not always accompanied by perceptual transparency can be demonstrated. Take a square of colored transparent plastic and glue it onto a larger square of black cardboard. Provided that the layer of glue is spread evenly, the plastic no longer is perceived as being transparent; it appears to be opaque. Changing the color of the cardboard, say from black to white, does not alter the effect [*see top illustration on next page*].

There also are instances where physical transparency is absent and perceptual transparency is present. Wolfgang Metzger of Münster has shown that mosaics of opaque cardboard can give rise to a perception of transparency even though there are no elements in the mosaic that are physically transparent [*see second illustration from top on next page*]. These two examples make it clear that physical transparency is neither a necessary nor a sufficient condition for the perception of transparency. Physical transparency cannot explain perceptual transparency.

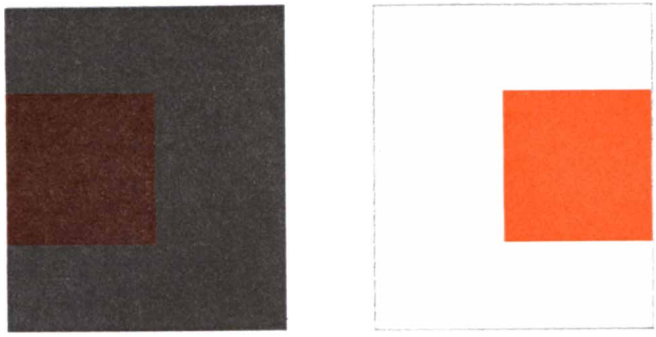
What causes perceptual transparency? As with other visual phenomena, the causes must be sought in the pattern of stimulation and in the processes of the nervous system resulting from retinal stimulation. Light reaches the retina only after having passed through several transparent mediums (air and the transparent mediums of the eye). The input to the retina, however, does not contain specific information about the characteristics of the transparent layers through which the light has traveled and been filtered. The perception of transparency is thus not the result of filtration; it is a new fact originating in the nervous system as a result of the distribution of the light stimuli acting on the retinal cells.

Perceptual transparency depends on the spatial and intensity relations of light reflected from a relatively wide field and not on light reflected only from a local area. This can be demonstrated by juxtaposing two sets of squares that do not appear to have any transparent areas [*see third illustration from top on next page*]. The juxtaposition produces a change from apparent opacity to transparency even though the light reflected from each region has not changed.

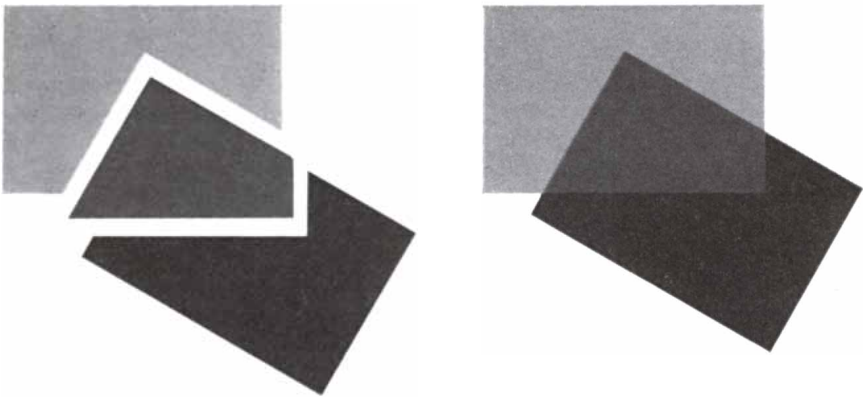
The conditions under which transparency is perceived have been studied by several eminent investigators, beginning in the 19th century with Hermann von Helmholtz and his contemporary Ewald Hering. They were at odds on almost all points. In his treatise on physiological optics Helmholtz described the perception of transparency as "seeing through" and studied it with a simple device in which images of two strips of paper of different colors were perceived one behind the other. The colors were superposed by reflection and transparency. Similar dual images can be found on windows under certain conditions, for example in the evening when one looks outside and sees both the reflection of the illuminated room and the external landscape.

Hering denied the possibility of seeing one color behind another. He argued that when light reflected by two different colors reaches the same retinal region, an intermediate or fusion color will be perceived. He supported his argument with new observations. When an observer concentrated only on the region where the two color images were superposed, just one color, the fusion color, was perceived.

In 1923 the German psychologist W. Fuchs was able to solve the Helmholtz-Hering controversy. He showed that both colors are perceived only when the transparent object and the object seen through it are perceived as independent objects. If the region of superposition of the two objects is isolated (even if it is just by the attitude of the observer), then only the fusion color is perceived. In the following years important findings were made by the Gestalt psychologist Kurt Koffka and some of his students at Smith College. B. Tudor-Hart showed that transparency on a totally homogeneous ground is not possible (for example the transparent plastic on a black cardboard). In 1955 Gaetano Kanizsa of the University of Trieste pointed out that whereas investigators had been concentrating only on the region of superposition of two figures, the conditions for perceiving transparency also applied

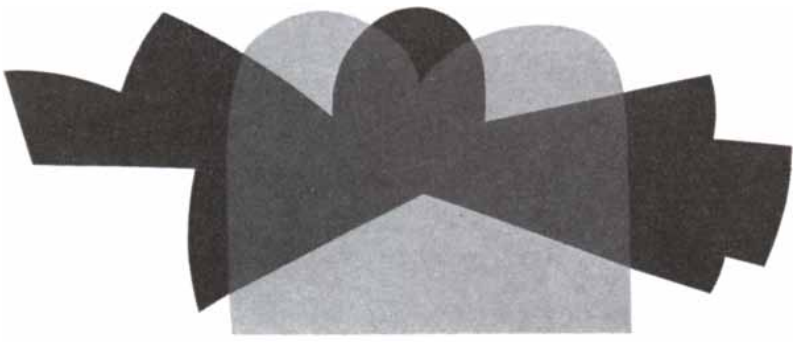


COLORED SQUARES OF TRANSPARENT PLASTIC glued onto a black cardboard (*left*) or a white one (*right*) no longer appear to be transparent. This demonstrates that perceptual transparency is not possible when the underlying field is homogeneous.



MOSAIC METHOD for constructing a figure with perceptual transparency out of opaque pieces is depicted. There is a strong impression of transparency in the central region where the two rectangles overlap. The method was originally developed by Wolfgang Metzger.

CHANGE FROM OPACITY TO TRANSPARENCY is obtained when the two figures depicted in the illustration at the top of the page are juxtaposed in the manner shown here.



TRANSPARENCY EFFECT is much more evident on an opaque figure than on the background, but conditions required to perceive transparency are the same in both instances.

to the regions in which the background could be seen through the transparent surface. The fact that this point had been neglected indicates that transparency on a figure is much more evident than transparency on the background [*see bottom illustration at left*].

The early investigators worked with filters or transparent objects, but after it became clear that physical transparency is not essential for the perception of transparency the use of physically transparent objects was generally abandoned. A number of investigators worked with the episcotister: a wheel with sectors cut out. The wheel generates a strong impression of transparency when it is rotated at high speed [*see top illustration on opposite page*]. This technique enables the experimenter to independently vary the size of the missing sectors (which affects the degree of transparency) and the color of the remaining sectors (which determines the color of the transparent layer).

In my own work I have used the mosaic method developed by Metzger because it offers a means of independently varying the color, the size and the shape of each region of a configuration. With this method it is easy to demonstrate that transparency depends on form as well as on color [*see middle illustration on opposite page*].

There are three main figural conditions for perceiving transparency in overlapping figures: figural unity of the transparent layer, continuity of the boundary line and adequate stratification. Let us examine each condition in turn.

When the unity of the central region of a transparent shape is broken, the perception of transparency is lost [*see bottom illustration on opposite page*]. On the other hand, modification of the shape that does not break up figural unity will not cause transparency to be lost. Figural unity of the transparent layer alone, however, is not sufficient to give rise to the perception of transparency. The boundary that divides the figure into two regions (one light and one dark) must be perceived as belonging to the opaque regions. A break in the continuity of the boundary line where it intersects the transparent layer can destroy the transparency effect. Abrupt changes in the boundary at points other than this intersection do not hinder the perception of transparency [*see top illustration on page 94*].

We have defined the perception of transparency as seeing surfaces behind a transparent medium or object. This means that the layer having the condi-

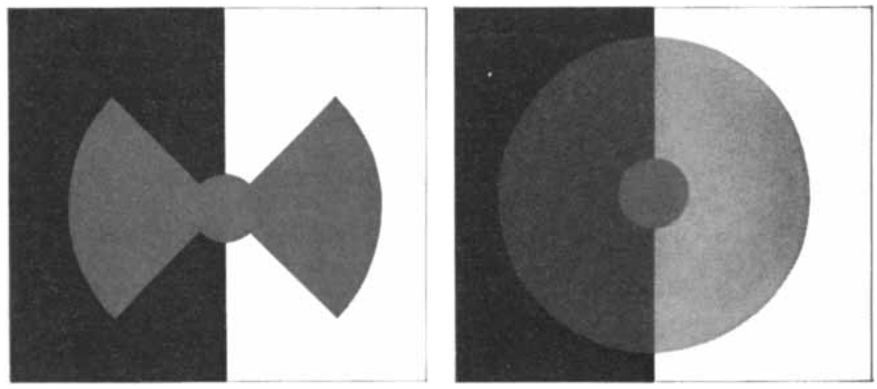
tions necessary to become transparent must be located on or above the surface of the opaque object. It is not sufficient, however, for one surface to be perceived as being on top of another in order to obtain the effect of transparency. It is possible to perceive different strata in figures where no transparency is seen [*see bottom illustration on next page*]. In order to create adequate stratification for transparency the underlying regions must appear to meet under the whole of the transparent layer.

Let us take as a model a figure in which the underlying region is composed of two squares, one black and one white. On these are superposed two smaller squares, one light gray (over the white) and the other dark gray (over the black) [*see illustration on page 96*]. When all the figural and color conditions for transparency are met, then the gray regions appear to be a single transparent surface. (Unbalanced transparency is possible, but here we shall for the most part discuss cases where the transparent layer appears to be uniform.)

How is it that two shades of gray give rise to the same shade of gray in the transparent layer that is perceived? This phenomenon has been described as a case of perceptual scission, or color-splitting. The original gray is called the stimulus color. With the perception of transparency the stimulus color splits into two different colors, which are called the scission colors. One of the scission colors goes to the transparent layer and the other to the surface of the figure below. In 1933 Grace Moore Heider of Smith College formulated the hypothesis (and gave an experimental demonstration) that there is a simple relation between the stimulus and the scission colors: when a pair of scission colors are mixed, they re-create the stimulus color.

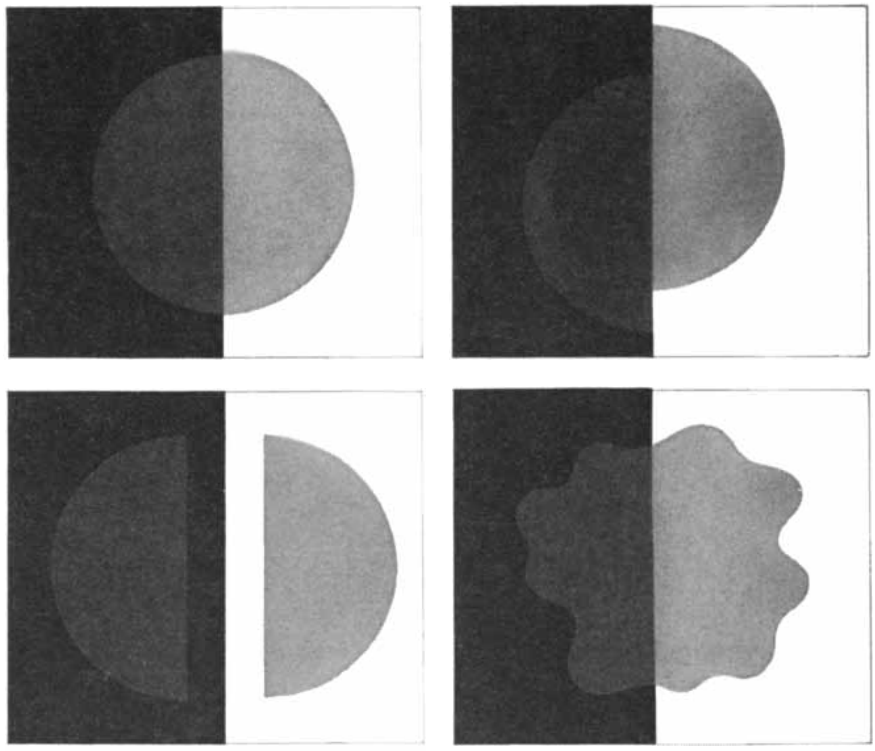
The process of color scission works in a direction opposite to that of color fusion. The law of color fusion, also known as Talbot's law (although it actually goes back to Isaac Newton), enables us to predict what color will be perceived when two colors are mixed. The same law, as Heider demonstrated, can be used to describe the color scission that gives rise to transparency. Since measuring chromatic colors such as yellow, red and blue is relatively complex, we shall limit our discussion to the achromatic colors (white, gray and black), which can be measured in a simple way. The achromatic colors vary only in one dimension: lightness. They can be defined by their albedo, or coefficient of reflectance: the percentage of light they reflect.

Every surface absorbs and reflects

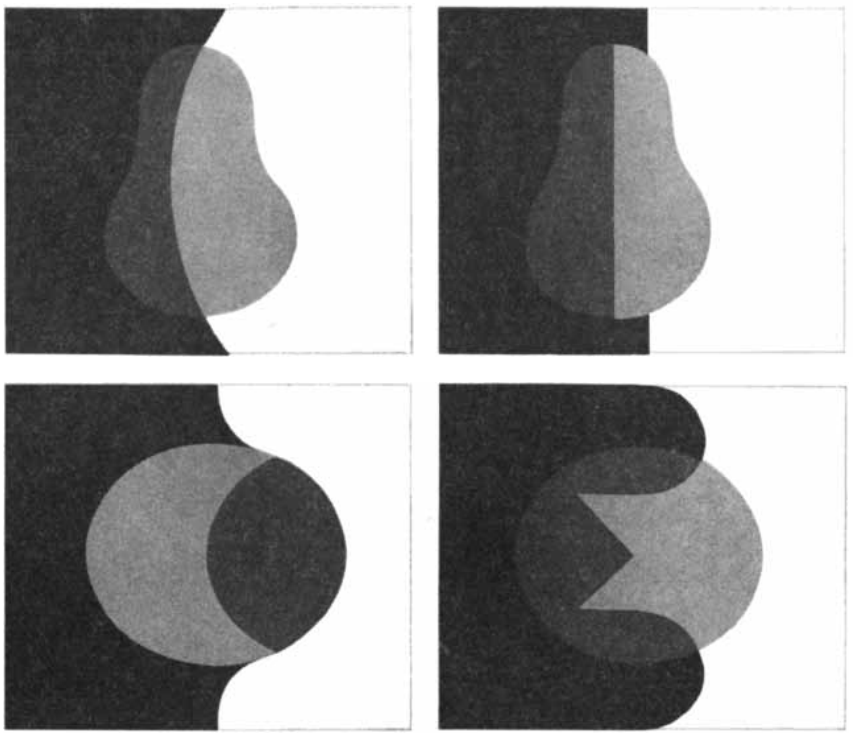


EPISCOTISTER is a wheel with cutout sectors (*left*). When the wheel is rapidly rotated with a suitable background behind it, a strong impression of transparency is created (*right*).

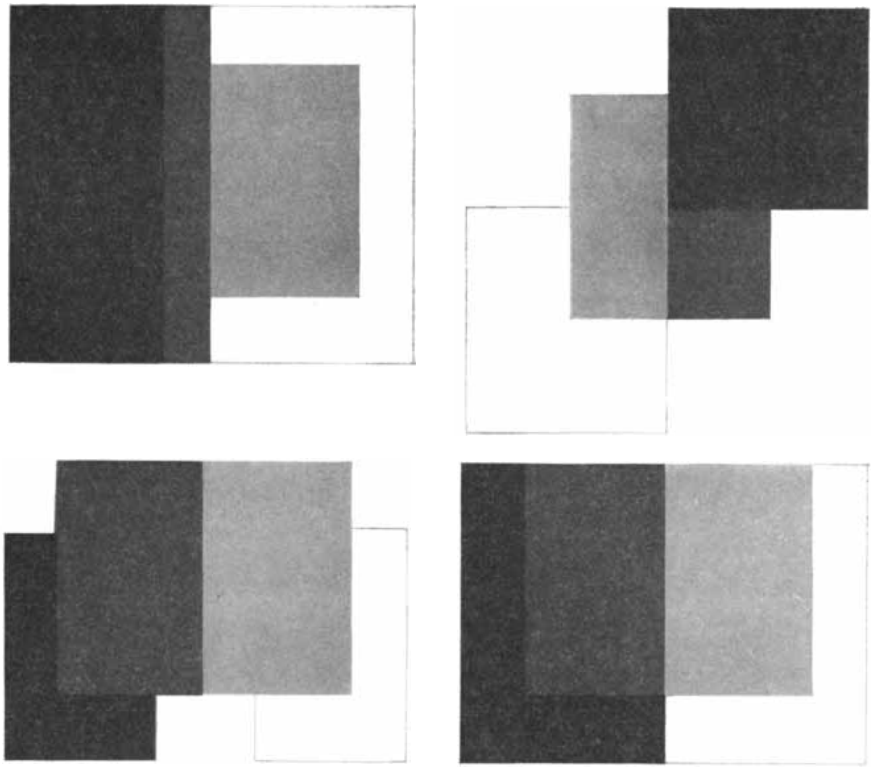
PERCEIVED TRANSPARENCY of the gray circle (*figure at left*) can be abolished either by an abrupt change of form (*middle*) or by an alteration in the color relations (*right*).



FIGURAL UNITY OF THE TRANSPARENT LAYER is a necessary condition for perceiving transparency (*top left*). When the unity of the shape is broken, the transparency effect is lost. Changes in the shape, however, do not destroy transparency (*bottom right*).



BOUNDARY LINE must appear to belong to the underlying opaque regions and must be visible through the transparent layer for transparency to be perceived (*top left*). Sudden change of the boundary line at the points of intersection causes transparency to be lost (*top right and bottom left*), but in other locations, even in the region that appears to be transparent, it can make abrupt changes without affecting the transparency (*bottom right*).



STRATIFICATION OF SURFACES is another necessary condition for the perception of transparency. If the light gray and the dark gray regions of a figure are perceived as being two different strata, figural unity is lost and transparency is not perceived (*top left*). Another example of inadequate stratification is when the gray regions appear to have an opaque layer above them (*top right*). The underlying opaque regions must meet under the whole of the gray regions in order for transparency effect to occur (*bottom left and right*).

part of the light falling on it. An ideal white that reflects 100 percent of the light falling on it would have a reflectance of 100; an ideal black that absorbs 100 percent of the light falling on it would have a reflectance of zero. These limits are never reached; a piece of white cardboard typically has a reflectance of about 80 and a piece of black cardboard a reflectance of about 4. Grays have a reflectance ranging from 4 to 80.

A device for studying color fusion is the color wheel. Two or more colors are placed on the wheel, which is then rotated rapidly. The fusion color perceived depends on two factors: the component colors and the proportions in which they are mixed [*see illustration on opposite page*]. With achromatic colors the resulting fusion color can readily be predicted, but with color scission there is a great variety of ways in which the stimulus color can split. How can we determine how much of the stimulus color will go to the transparent layer and how much to the opaque layer?

Let us consider first the transparent layer. By way of example imagine what happens when you add a dye to a glass of water. As more dye is added the water becomes less transparent and objects seen through the water become less visible. It is therefore plausible that in the scission process the greater the proportion is of color going to the transparent layer, the less its perceived transparency will be.

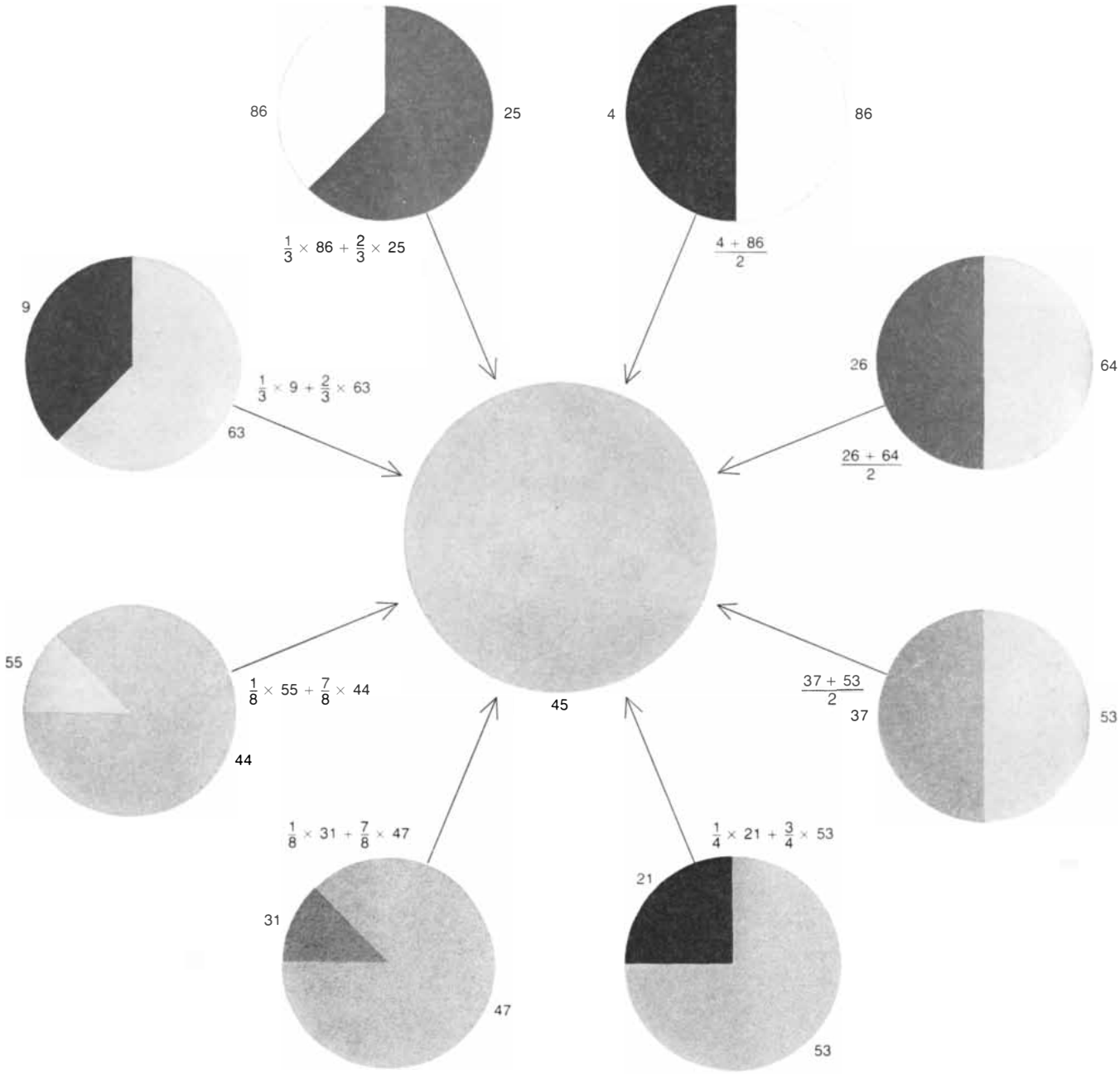
Now let us consider the opaque surface. Suppose that as you view it through a glass of water it is painted with a dye. Obviously the visibility of the opaque surface will increase as more dye is put on it.

The limiting case in the scission process is when all the color goes to one layer. If all the color goes to the transparent layer, it becomes opaque. If all the color goes to the underlying surface, then the transparent layer becomes invisible. Transparency is perceived only when there is a distribution of the stimulus color to both the transparent layer and the opaque layer. Moreover, transparency varies directly with the proportion of color going to the opaque layer. As more color goes to the opaque layer, less goes to the transparent one and the more transparent it appears. The proportion of color going to the opaque layer, which is described by an algebraic formula, can therefore be regarded as an index of transparency.

With achromatic colors it is possible to derive a second algebraic formula that

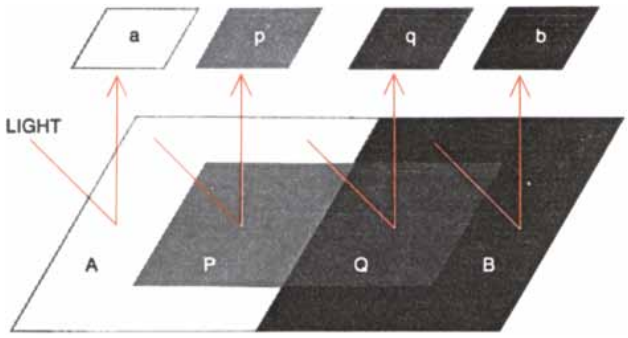
states a relation between the reflectances of the surfaces involved and the color of the transparent layer [*see illustration on next page*]. If the reflectances of the four surfaces in the figure are known, then the index of transparency can be calculated and the relative lightness of the transparent layer can be predicted. Such predictions are possible when (in most cases, as it happens) the transparent layer is perceived to be uniform in color as well as in the degree of transparency; in other words, the transparent layer is a perceptual unit, not divided by the boundary belonging to the opaque layer below.

The validity of the theoretical algebraic formulas can be tested by taking our model figure and altering the color (black, gray and white) of individual regions. When the reflectance values of the gray squares are very different, the calculated coefficient of transparency is large and therefore transparency should be readily perceived. When the gray regions are similar, the coefficient is very small and transparency usually is not perceived. Some necessary color conditions of transparency can be deduced from the theoretical formulas. Transparency is possible only when the darker gray square is on the darker underlying surface and the lighter gray square is on the lighter underlying surface. If these conditions are not met, transparency cannot be perceived. Finally, the difference of reflectance of the colors in the transparent layer must always be less than the difference of reflectance of the

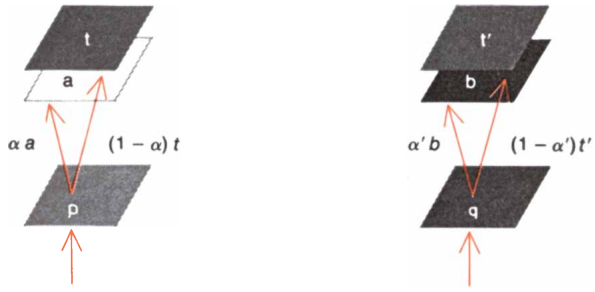


COLOR FUSION is produced when a wheel with sectors of different colors is rapidly rotated. With achromatic colors (black, gray and white) the fusion color can be calculated. For example, if the disk has two sectors of equal size, then the fusion color perceived will be the simple average of the reflectance of each sector. If the sectors are of unequal size, the fusion color is the weighted average. The reflectance figures given here are only representative. The same shade of gray (*center*) can be produced by a variety of color mixtures. Since color scission is the reverse of color fusion, it is apparent that a particular gray can split in a great variety of ways.

1

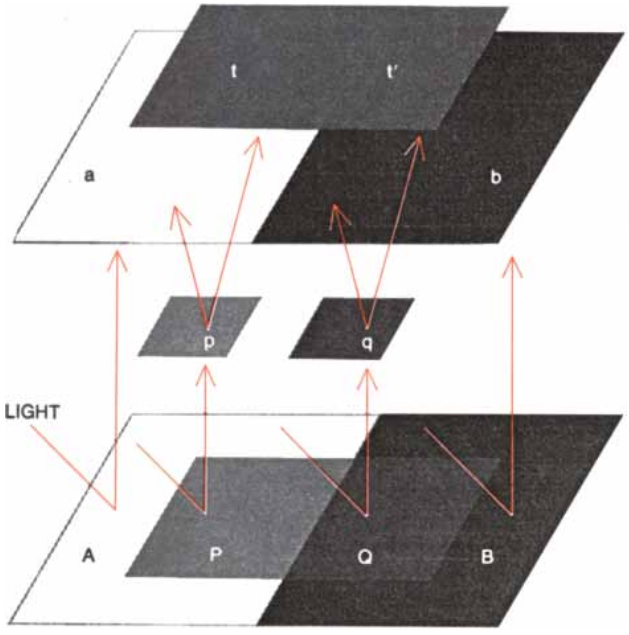


2



$p = \alpha a + (1 - \alpha) t$ OR $\alpha = \frac{p - t}{a - t}$ $\quad q = \alpha' b + (1 - \alpha') t'$ OR $\alpha' = \frac{q - t'}{b - t'}$

3



IF $t = t'$ AND $\alpha = \alpha'$ THEN $\alpha = \frac{p - t}{a - t} = \frac{q - t}{b - t} = \frac{p - q}{a - b}$, $t = \frac{aq - bp}{(a + q) - (b + p)}$

4

LEFT HALF $\alpha = \frac{p - t}{a - t}$					RIGHT HALF $\alpha = \frac{q - t}{b - t}$				
a	>	p	>	t	AND	b	>	q	>	t
a	>	p	>	t	AND	t	>	q	>	b
t	>	p	>	a	AND	t	>	q	>	b
t	>	p	>	a	AND	b	>	q	>	t

colors in the underlying layers [*see top illustration on page 98*].

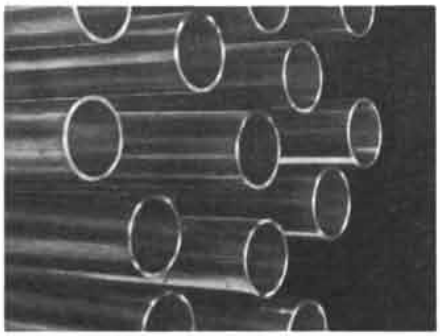
Another powerful factor in perceiving transparency (in addition to the proportion of color going to the opaque and the transparent layer) is the color of the transparent layer itself. All other conditions being equal, the darker the transparent layer, the greater its perceived transparency.

The conditions for the perception of transparency that are deduced from the algebraic formula also enable us to predict the degree of lightness of the transparent layer when the colors of the stimulus regions are varied. With our model figure it is not always easy to judge the color of the transparent layer. With a checkerboard pattern, however, such estimations are easier to make and predictions about the transparent layer are visually confirmed [*see bottom illustration on page 98*].

The color conditions for perceptual transparency discussed here are theoretically derived without any empirical correction or adaptation. They state relations for "pure" achromatic conditions. Figural conditions, as has been noted, play a role and cannot be entirely excluded, but they can be held constant. It must be stressed that the inferences drawn from the theory should be considered as describing some (but not all) necessary conditions for the perception of transparency. In other words, certain instances are described where the perception of transparency is possible and instances where it is impossible. Of

THEORY OF COLOR SCISSION explains transparency as a case of perceptual color-splitting. The achromatic colors can be defined simply by the percentage of light they reflect (*1*). When transparency is perceived, the areas *P* and *Q* split and appear to consist of two surfaces, equal in form and size but different in color. Assuming that this color scission follows the same law as color fusion, then the proportion of the stimulus color going to each of the perceived surfaces can be described by an algebraic formula (*2*). The symbols α and α' stand for the proportion of color (which can vary from zero to one) going to the opaque layers *a* and *b* respectively. The remainder of the color goes to the transparent layers *t* and *t'*. If $\alpha = \alpha'$ and $t = t'$ (*3*), then the algebraic equations can be solved for α (transparency) and for *t* (color of transparent layer). From the formulas certain predictions about the perceived lightness of the transparent layer can be made from the relations of the colors of the *A*, *P*, *Q* and *B* regions (*4*). The symbol > here means "lighter than."

WESTERN ELECTRIC REPORTS



The tolerances of waveguide tubes are ten times more precise than any previously produced by commercial suppliers. In operation, they will have capacity for 230,000 simultaneous voice circuits.

Researcher inspects wear on a "super-mouse." An electromechanical gauge, it quickly measures waveguide geometry by traveling up and down each tube.

Finished waveguide is packed in specially designed crates.

First the medium. Then the message.

A communications transmission system that would harness electromagnetic waves only a few millimeters long has been an exciting prospect since the early 1930's. Its bandwidth would be broader than all of the presently used lower frequencies combined, giving it dramatically increased capacity and lower costs per message.

Solid state devices capable of generating the required high frequencies were developed at Bell Labs. But the system would work satisfactorily only within the controlled environment of a nearly geometrically perfect tube, or waveguide. It would require the development of the first new physical transmission medium in a third of a century—millimeter waveguide.

Increases in long distance calling have made millimeter waveguide an idea whose time has come. In 1969, working with Bell Labs, Western Electric's Engineering Research Center established a branch laboratory at Forsgate, N. J. to solve the problems integral to producing an economical waveguide medium.

Some examples: How to make 30-foot lengths of steel tubing to microtolerances never before consistently attained. How to plate precisely .0005 of an inch of copper all along the insides of the lengthy tubes. How to couple the tubes to each other with their centers precisely aligned for the length of the system.

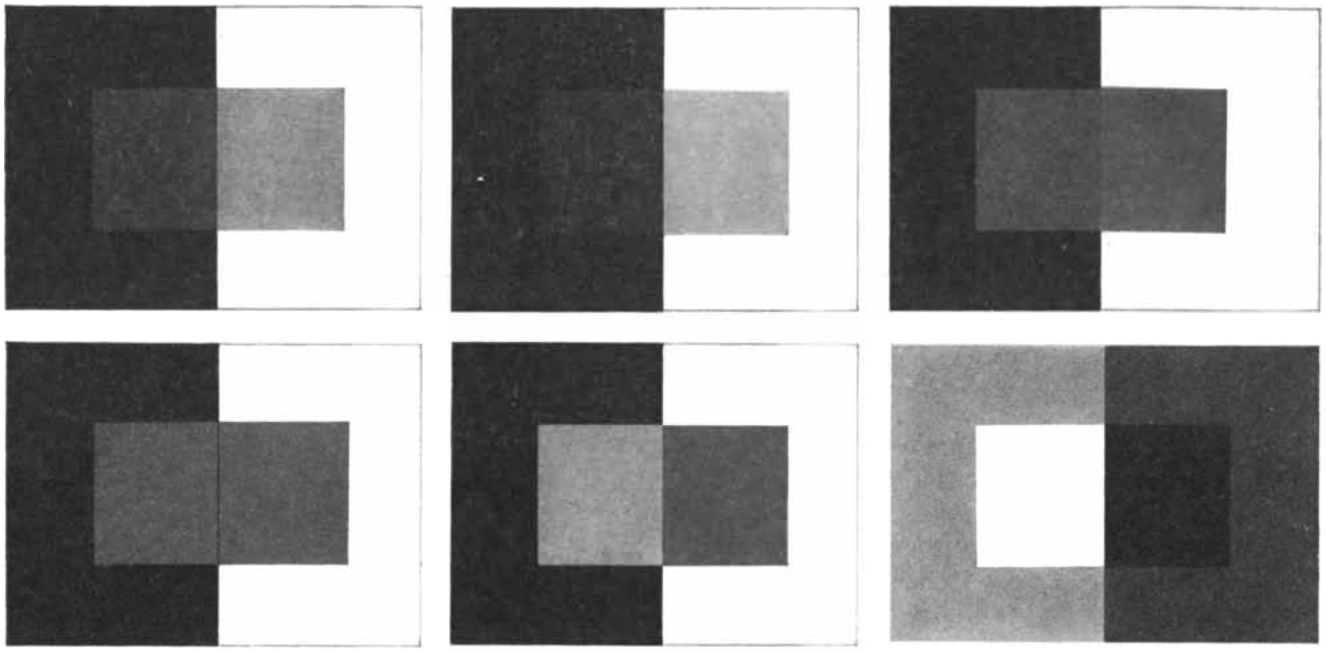
Often the answers lay in the development of new manufacturing techniques. Western Electric engineers developed computer-designed rollers for the straightening equipment to be used by steel tubing suppliers. They innovated a plating process in which the tubes are rotated horizontally as the copperplating solution is circulated through them. They designed flanges which are welded to the tubes with electron beams, and a centering device that helps in machining the flanges precisely.

And they so far outperformed previous expectations that the waveguide's size could be increased for lower transmission losses—and fewer repeater stations will be necessary.

This year and next, extensive field evaluation of the system will be conducted in a 20-mile corridor between New York and Philadelphia.

Western Electric

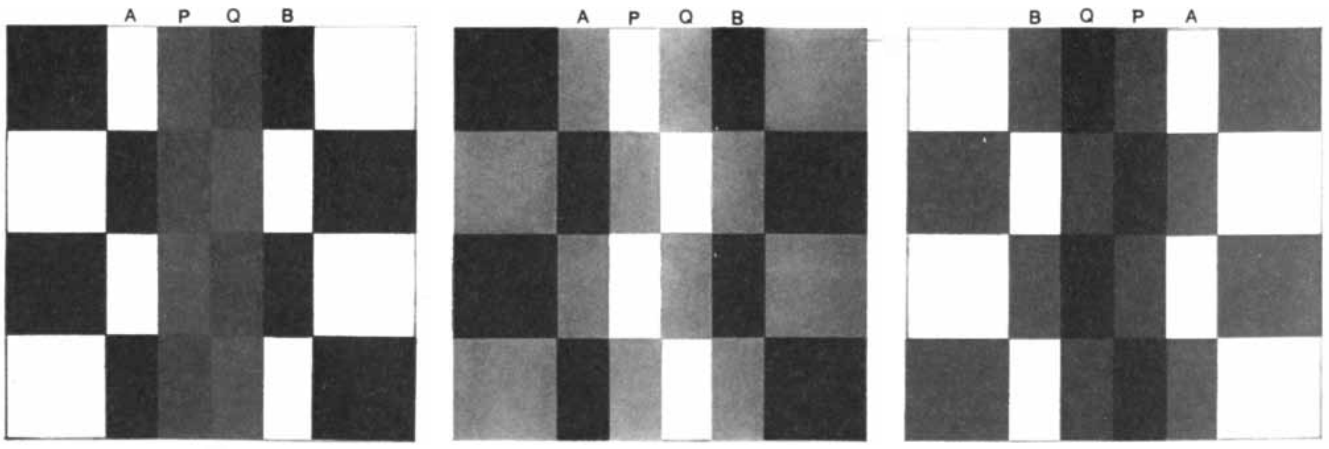
We make things that bring people closer.



COLOR CONDITIONS necessary for the perception of transparency are demonstrated. In the model figure transparency is readily perceived (*top left*). According to the author's theoretical formula, the degree of perceived transparency increases when the difference between the dark and light gray regions is increased (*top middle*). When the gray regions are similar, perceived transparency is low (*top right*). When the grays are identical, no transparency is perceived (*bottom left*). Transparency is impossible when the darker gray is over the lighter background (*bottom middle*). If the difference between light and dark colors of the background is less than the difference between the colors of the central region, then the central region is not perceived as transparent (*bottom right*).

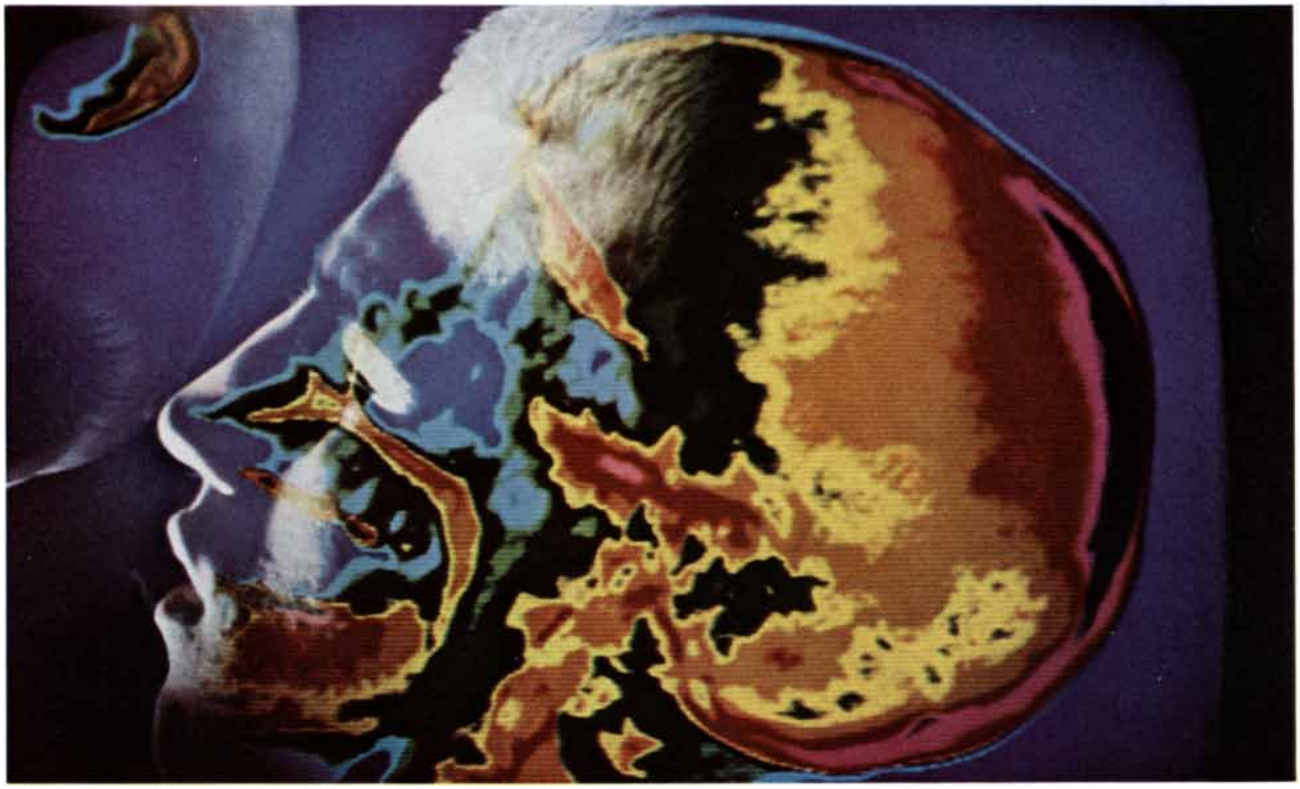
course, not everyone will perceive transparency when it is theoretically possible. On the other hand, when it is predicted that it is impossible for transparency to exist in a figure, no one should be able to perceive it.

There is an important limitation to the index of transparency that has been discussed: it measures the degree of transparency only if the lightness of the transparent layer is held constant. It is possible to develop a new formula in which the color of the transparent layer is variable, but this can only be done empirically, and it would not give rise to the interesting deductions possible with a theoretical formula. We have dealt here primarily with instances of balanced transparency, that is, instances where the perceived transparent layer is uniform in degree of transparency and color. There are instances of unbalanced transparency, where the perceived transparent layer varies in degree of transparency. A special case is that of partial transparency, where one part of the upper layer is perceived as being transparent and the other as being opaque. Unbalanced transparency and partial transparency, of course, require different formulas for their theoretical description. Other factors such as motion and three-dimensionality are often involved in the phenomenon of transparency. It appears, however, that the main conditions for the perception of transparency are to be found in the figural and chromatic conditions that have been described here.



PERCEIVED LIGHTNESS OF TRANSPARENT LAYER depends on the relation of the colors in the figure and can be deduced from the author's theoretical formulas given in the illustration on page 96. According to his theory, if region *A* is lighter than *P* and region *Q* is lighter than *B*, then the perceived transparent layer appears to be darker than *P* and lighter than *Q* (*left*). If region *P* is lighter than *A* and region *Q* is lighter than *B*, then the transparent layer appears to be lighter than any of the colors (*middle*). If region *A* is lighter than *P* and region *B* is lighter than *Q*, then the transparent layer appears to be darker than any of the colors (*right*).

Why jump into a specialized field until you've studied all the possibilities?



If you're a new physician, you might not have decided on a field of specialization yet. That's easy to understand. But while you're deciding, you can be getting valuable experience and expert guidance towards that decision as a physician in the U.S. Air Force. Air Force physicians are exposed to many fields of medicine. They practice with highly professional people in modern facilities. And all Air Force physicians have seemingly unlimited opportunities for furthering their education and professional growth. Air Force graduate medical education programs are available. And because Air Force physicians participate in Group Practice (a collection of specialists in different fields complementing each other by expanding total capability) they all expand their individual ability and horizons with each new patient.

As an Air Force physician (and officer) you can apply for worldwide work assignments. The salary is excellent. And you get a 30-day paid vacation each and every year.

If you're still deciding on a medical specialty...or a future...look into the Air Force before you make a final decision. We think you'll like what you see. For more detailed information about Air Force opportunities in medicine please fill out and mail the coupon.

C-SA-44

Air Force Opportunities
P.O. Box AF
Peoria, IL 61614

Please send me more information on the Air Force Physician Program. I understand there is no obligation.

Name_____ Sex ☐ M ☐ F
(Please Print)

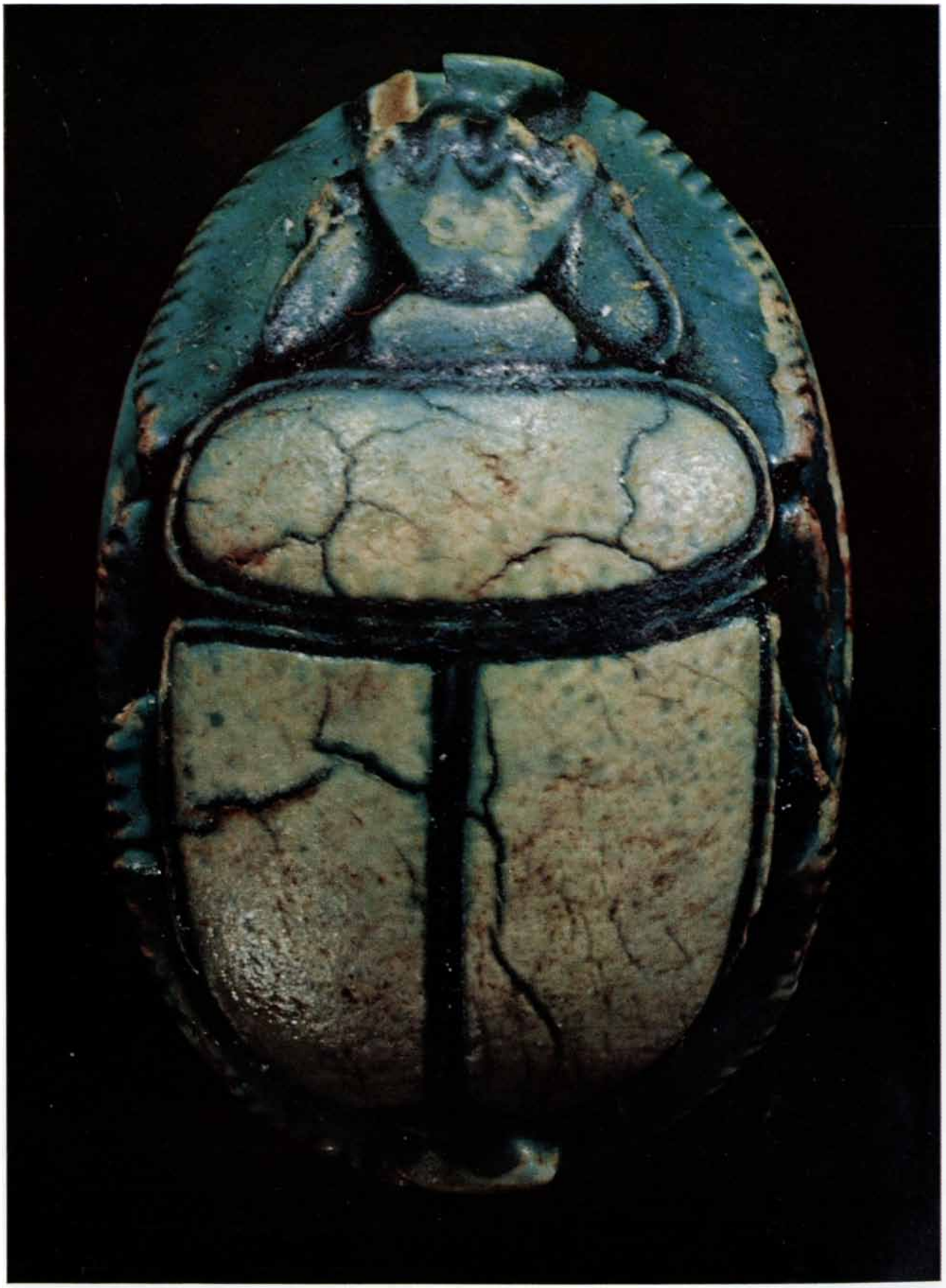
Address_____

City_____

State_____ Zip_____ Phone_____

Soc. Sec. #_____ Date of Birth_____

Health Care at its best.
Air Force.



SACRED SCARAB OF ANCIENT EGYPT is a representation of the dung beetle *Scarabaeus sacer*, a species that populates the perimeter of the Mediterranean. This scarab, made of blue-glazed steatite and about an inch long, is from the Hyksos period of about 1650 to 1580 B.C. It is in the Carnarvon Collection (gift of Edward S. Harkness, 1926) of New York's Metropolitan Museum of Art.

The Biological Control of Dung

When native Australian dung beetles cannot cope with the large, moist dung pads of cattle, dung covers pastureland and breeds insect pests. Now foreign beetles are being imported to help out

by D. F. Waterhouse

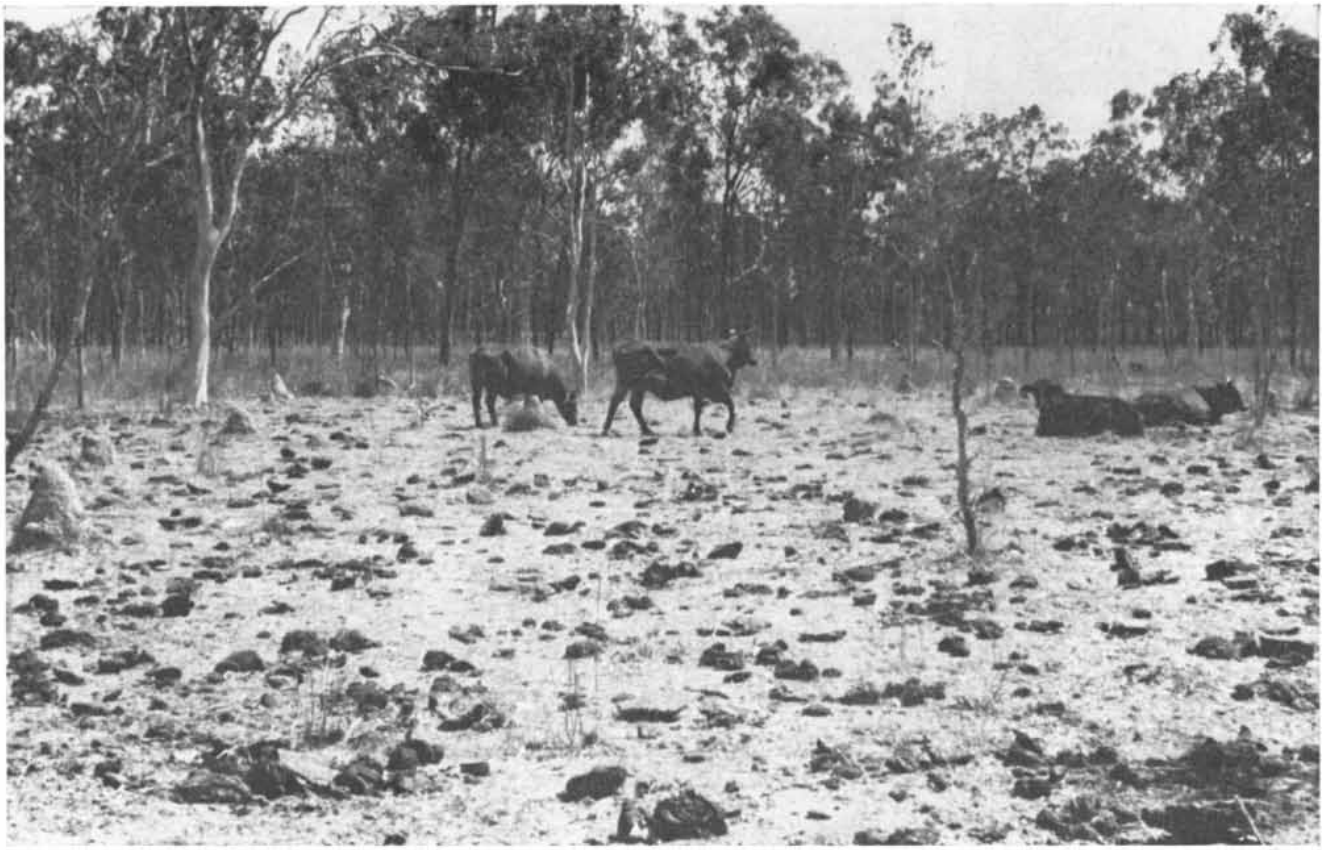
When the first English colonists arrived in Australia in January, 1788, they brought ashore with them five cows, two bulls, seven horses and 44 sheep, together with some uninvited fleas, lice and rats. Unfortunately certain important fellow travelers were missing from the company: bovine dung beetles. The omission created an ecological imbalance because dung beetles break down dung. For lack of dung beetles dung pads persist, obliterating pasturage and breeding insect pests. The results have become increasingly significant in the course of two centuries, as the number of bovines in Australia has increased from seven animals to 30 million. In the past 10 years our group in the Commonwealth Scientific and Industrial Research Organization (CSIRO) has undertaken to redress the imbalance between dung and beetles. In the process we have learned something about the nice interspecies adaptations that develop over millenniums, the trouble that can come when man upsets them and some ways man may be able to undo the damage.

The dung beetle and its works are well known to history. The scarab (*Scarabaeus sacer*), a beetle that populates the perimeter of the Mediterranean, was sacred to the early Egyptians. The ball of dung it forms and rolls along the ground was likened by them to the sun, and the beetle itself to the invisible power that daily propelled the sun across the sky. They went on to represent the sun god, their most important deity, as a scarab in both art and hieroglyphics. Later Aristophanes had Trygaeus, the hero of his comedy *The Peace*, mount to heaven on the back of a dung beetle. As for more scientific and practical references, there was useful investigation and writing concerning dung beetles a century ago, notably by the French naturalist Jean Henri Fabre. Yet it was only in 1960 that G. F. Bornemissza of the CSIRO Division of Entomology pointed out what lack of the proper dung beetles meant in Australia.

Before the arrival of European colonists the largest herbivorous animals in Australia had been marsupials such as kangaroos, which produce comparatively dry, fibrous dung pellets no larger than a golf ball and generally less than half that size. Marsupial dung never accumulated excessively because it provided food and nesting material for a well-adapted group of native insects: some 250 kinds of scarab beetles of the subfamily Scarabaeinae, commonly known as "coprids." As Bornemissza recognized, the situation was completely different with regard to the larger, moister dung pads of imported cattle. These pads are unattractive to most native Australian beetles and so they are seldom more than partially utilized, and that only during limited periods of the year. Most cow pads soon dry out to a hard cake on the ground where they have fallen, often remaining substantially unchanged for months or even years until they are finally disintegrated by weathering, rotting, trampling by stock or attack by termites.

As long as the cattle were few and ranged over large areas, their dung pads were comparatively few and scattered and caused no problems. As their numbers increased, however, it became commonplace to see pastures littered with pads that ranged in age up to several years. These objects make their presence felt in a variety of ways. First of all, the area they cover is significant. On the average 12 dung pads are dropped by a single adult bovine every day. If they are not disposed of, the pads produced by each animal will blanket between 5 and 10 percent of an acre in a year. Moreover, at the periphery of each dung pad there develops a zone of tall, rank herbage that cattle seldom eat and avoid for a year or more unless they are ravenous. The effective area of pasture is thereby reduced by each bovine by about 20 percent of an acre per year. A simple calculation indicates that the 30 million cattle in Australia, producing some 300 million or more dung pads a day, may be putting out of service as much as six million acres of pastureland each year. At least part of this effect may be cumulative, extending into subsequent years and constituting a truly enormous loss to the dairy and beef industries.

The situation is spectacularly different in areas of the world such as Africa, where many large herbivores evolved and still survive. Except when temperatures fall too low (below about 15 degrees Celsius) or when it is too dry for much insect activity, the dung in such regions is disintegrated by an almost bewildering variety of beetles. In Africa upward of 2,000 species of coprid beetles are known to utilize the dung of the many and diverse species of herbivorous animals. Some beetles are relatively undiscriminating and are attracted to dung from quite a range of herbivores. Others are specialists and avoid dung that does not have the size, special texture, composition, moisture content and other features characteristic of a particular herbivore species. Some beetles are adapted or even restricted to open pastures, whereas others prefer lightly timbered savanna woodland. Some coprids move from pad to pad only by night, whereas others are active during the day. There are also special adaptations to temperature, moisture, soil characteristics (sand, loam or clay), seasonal variations in the length of the day and other factors. These various adaptations, which have



COW DUNG ACCUMULATES in a pasture near Townsville, in the state of Queensland in northeastern Australia. Bovine dung beetles would break up the pads for food and nesting material; in their absence pads dry out and remain on the ground for many months.

DUNG PADS in another pasture, in New South Wales, stimulate the rank growth of undesirable herbage that cattle will not eat. This further reduces the amount of pasturage. Moreover, the dung pads serve as a breeding ground for the larvae of insect pests.

evolved over the millenniums, enable each species to occupy a particular ecological niche where it survives in spite of all competitors.

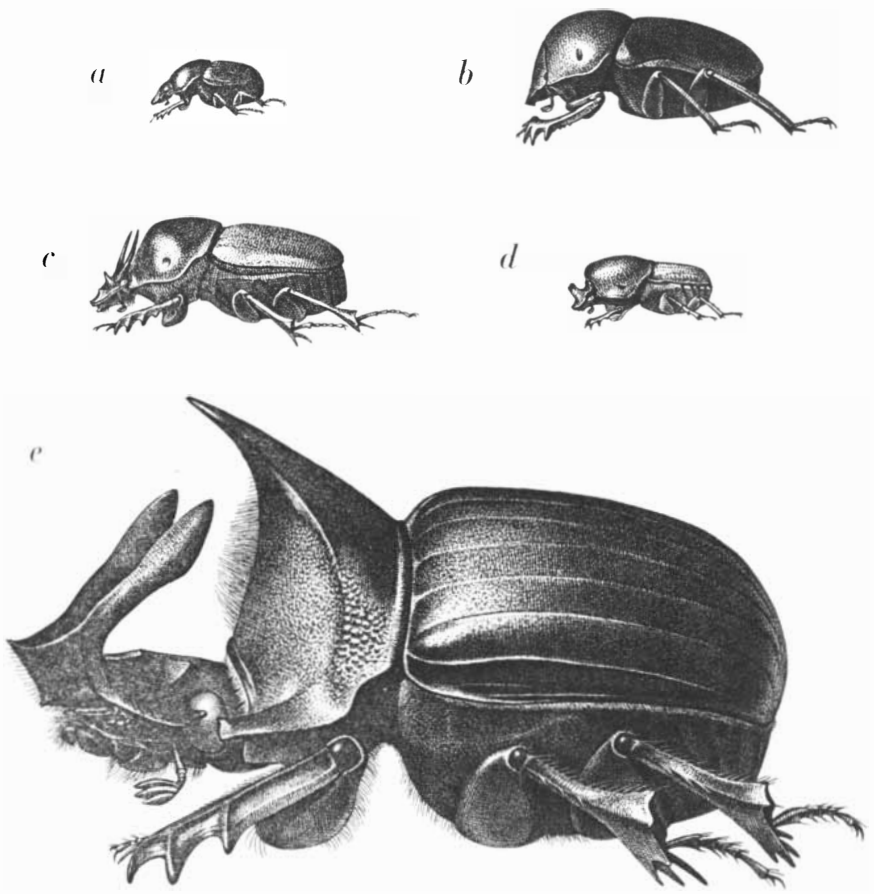
Dung beetles are powerfully attracted to fresh dung. In Africa the beetles are alerted as soon as a nearby buffalo lifts its tail to defecate and passes a little odorous fecal gas. By the time the steaming dung pad has hit the ground beetles are orienting to it. Minutes or even seconds later, before the buffalo has moved very far, many beetles have homed in on the dung and are already burrowing in it. Within a day, sometimes within only a few hours, nothing may remain of the pad except a few dry wisps of plant fiber on the surface of the fresh soil excavated by beetles tunneling under it. The numbers of beetles involved are huge. More than 7,000 have been counted in a single mass of fresh elephant dung in Kruger National Park in South Africa.

Different species of beetles utilize dung in different ways [*see illustration on next page*]. Most species excavate tunnels in the soil under or directly adjacent to the dung pad. They carry the dung down into the tunnels and fashion it into balls in which the females lay their eggs. In selecting the dung many species remove all irregularities from it, leaving material such as seeds behind on the surface. Seeds thus removed fall on ground that has been loosened by the tunneling, which makes it easier for them to become established when they germinate.

Beetles of another group carve a mass out of the dung and move it some distance from the pad before burying it. Some species simply butt the mass over the ground but others knead pieces of dung into smoothly rounded balls and roll them for many yards before they bury them. (Sometimes when the male is exerting prodigious energy pushing or pulling the ball, the female can be seen balancing on top of it as it rolls along.) Egg laying subsequently takes place in the soil chamber where the ball has been buried.

When feeding, adult beetles busily squeeze pellets of moist dung between their mouthparts and suck in the expressed juice. Their digestive tract is usually found to contain a dark fluid mixed with very fine particulate material, and so it appears that dung fluid and the soluble nutrients and microorganisms it contains provide their major food. It is supplemented to some extent by solids derived from dung particles, which are rubbed into a fine paste between specially adapted areas on the mandibles. The eggs of parasitic worms that were passed in the dung are seldom if ever found in the digestive tract of adult dung beetles, so that in spite of flying from pad to pad most species apparently play no part in dispersing the worms. On the contrary, the adult beetles are probably responsible for destroying many worm eggs as they grind their food to a paste.

In contrast to the adult food, the food of the beetle larvae consists of a paste of fine solids mixed with coarse particles. When a newly hatched larva begins to eat into the surrounding food ball, it enlarges the spherical space around itself, rotating as it consumes its food. Coprid larvae typically have a large hump, formed by the dilation of some of the abdominal segments, that encloses a coiled region of the digestive tract. This unusual adaptation makes it easier for the larva to move and feed within the confined space in which it is developing. The larvae are capable of sealing breaks in the wall of their dung balls with liquid feces. If they are removed from their brood balls, they are unable to survive, and so they cannot attack plants or anything else of value to man.

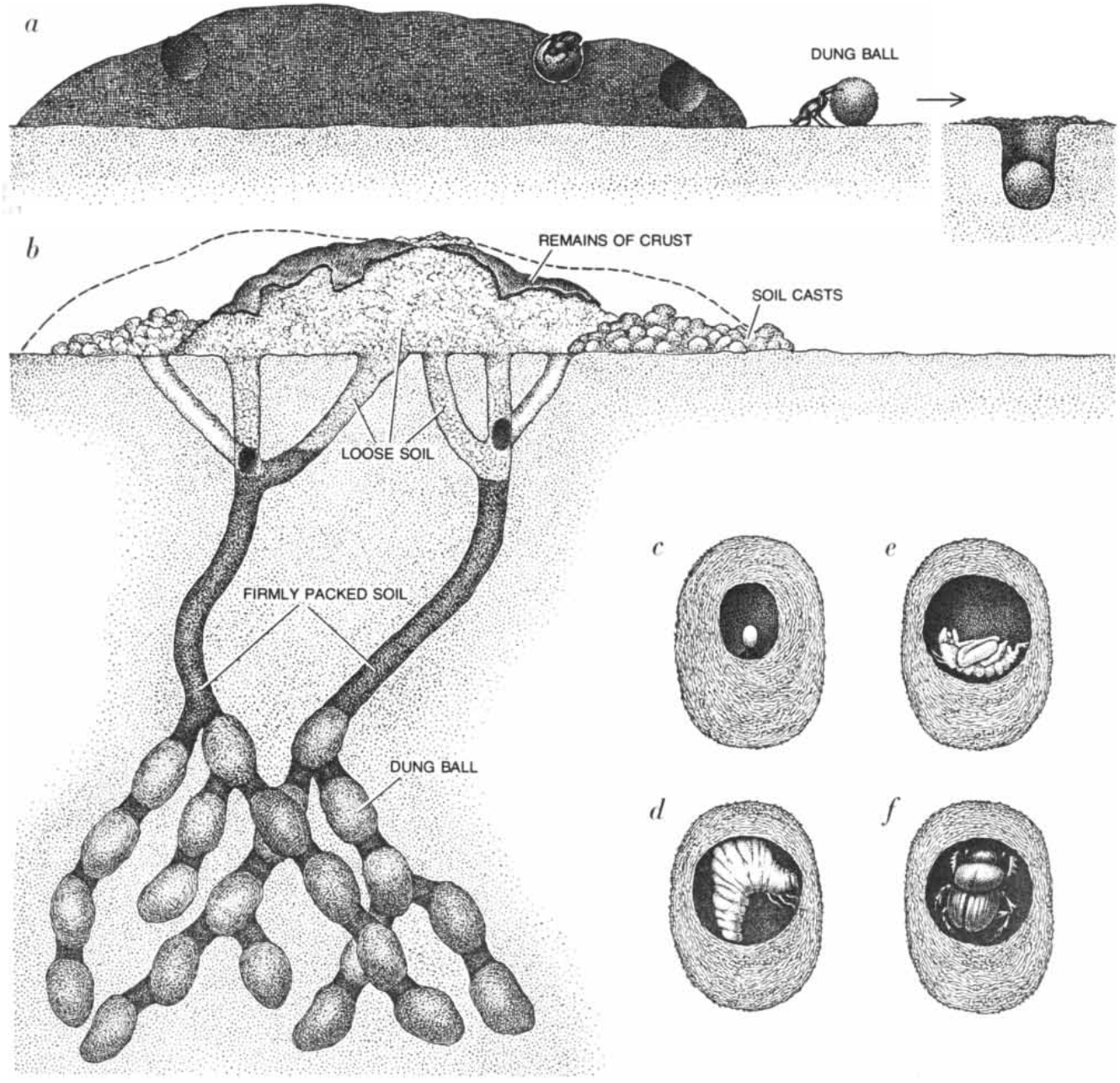


COPRID BEETLES vary widely in size and morphology. Five of them are shown here at the same scale, about twice natural size. *Onthophagus parvus* (*a*) is a native Australian beetle that does not utilize cattle dung. The others are bovine-dung beetles. *Garreta nitens* (*b*), recently introduced, is a ball-rolling species. *Onthophagus gazella* (*c*) and *Euoniticellus intermedius* (*d*) are the two tunneling species that have been most widely established. *Heliocopris gigas* (*e*) is a representative of a giant genus expected to resist attack by toads.

One foreign dung beetle, *Onthophagus depressus* from South Africa, became established (apparently by accident) in Australia before 1900, but the first deliberate attempts to colonize coprids for man's benefit appear to have been made in Hawaii. A species from Mexico was introduced there in 1906 and a second from Germany in 1908, but they failed to become established. In 1923, however, three beetle species were successfully introduced into Hawaii from Mexico to aid in the control of the horn fly (*Haematobia irritans*), and success also attended the later introduction of an Afro-Asian species, *Onthophagus gazella*. Stockmen on the island of Hawaii say that the number of horn flies there has fallen markedly since dung beetles became abundant. In one experiment, when fresh pads were exposed to horn flies but protected against beetles, each pad produced hun-

dreds of flies; pads not protected against beetles produced either a few stunted flies or none at all. The particular dung beetles concerned would colonize only pads dropped in open pastures, however, and not those dropped in the dense growth of mesquite, where the cattle sought shelter from the heat of the day. The flies that originate in the mesquite might yet be dealt with by beetles that are adapted to woodland and scrub.

Australia is plagued with the exceedingly abundant and pestiferous native bush fly *Musca vetustissima* and the buffalo fly *Haematobia irritans exigua,* which was introduced from the Indonesian island of Timor. In experimental situations we have noted 80 to 100 percent reductions in bush-fly production when adult *O. gazella* beetles were allowed access to dung on which flies had deposited eggs. The beetles broke up and buried the cow pads within 30 to 40 hours, and a few surviving maggots matured as stunted flies that were in turn capable of laying very few eggs or none at all. Neither fly eggs nor maggots were ever found in the dung balls; it is evident that they were discarded or destroyed by the beetles during the elaborate process of converting lumps of dung into brood balls. Almost complete control was also obtained in midsummer under natural field conditions near Pretoria in South Africa. It is relevant to note that *Musca vetustissima,* or at least a fly so similar that it produces fertile hybrids with its Australian relative, is found in South Africa. Whereas in Australia the



DUNG IS DISINTEGRATED in different ways by different dung beetles. One group, represented by such species as *Garreta nitens,* cuts bits of dung out of the pad, forms them into spherical balls and rolls the balls away to be buried in a shallow pit (*a*). Most species instead form their nests in a network of tunnels excavated below or adjacent to the pad (*b*). Working together, a male and female beetle dig a tunnel and carry dung down into it. The female forms a ball (ovoid, in the case of *Onthophagus gazella,* shown here), lays an egg in it and closes the ball (*c*). As many as 40 balls are formed. The tunnels are backfilled with firmly packed soil; loose soil fills the upper parts of the tunnels and is left on the surface along with some remains of the dung pad. When the larva hatches (*d*), it feeds on the dung. After passing through the pupal stage (*e*) young adult (*f*) emerges and makes its way to surface.

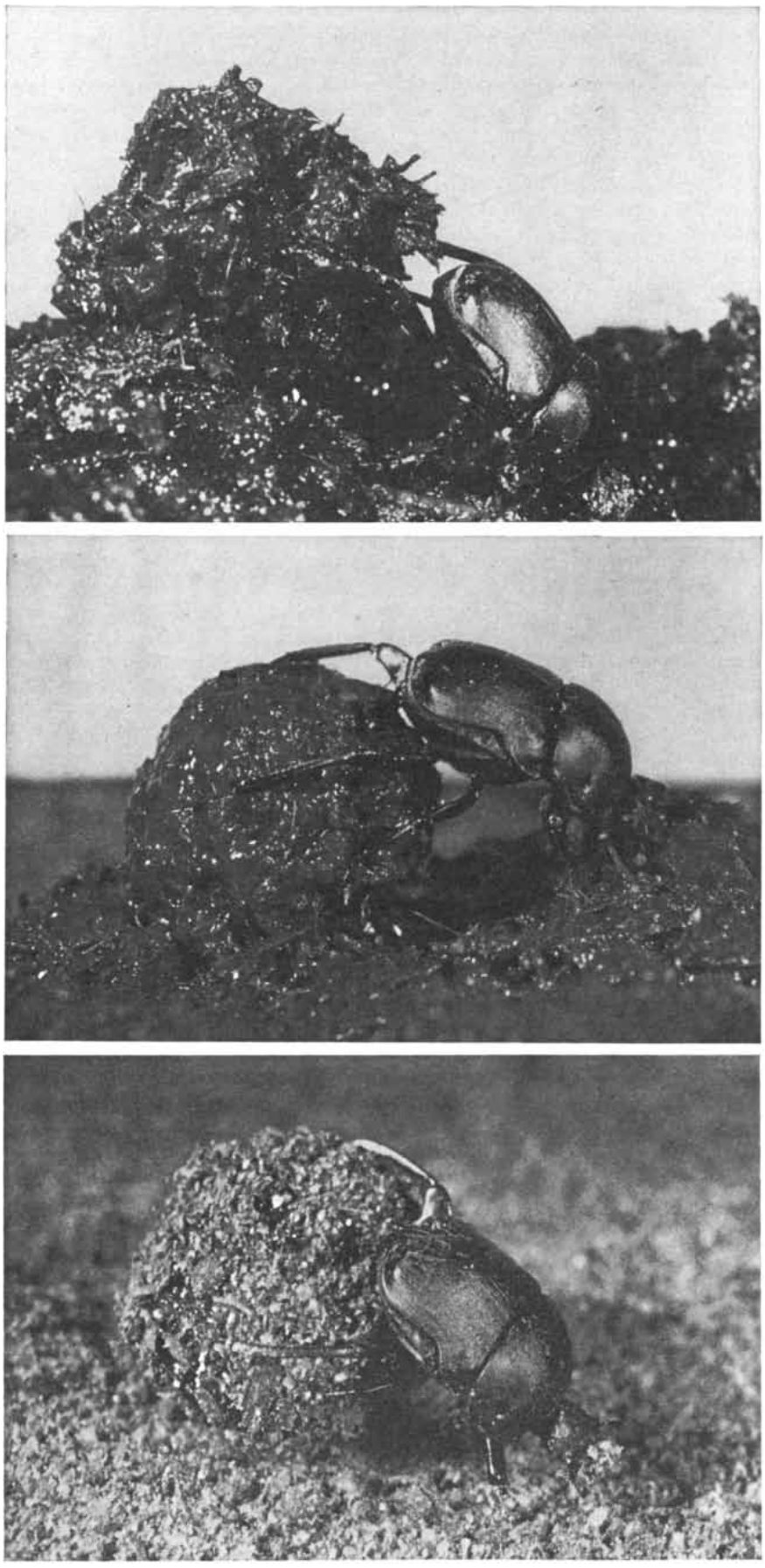
abundant bush flies can make outdoor life in summer a misery for humans and domestic animals alike, in South Africa the flies are scarce and seldom a nuisance.

Flies are not the only pests affected. In both South Africa and Australia experiments on the transmission of common helminths, or intestinal worms, of cattle have shown that the activities of the beetles greatly reduce the number of infective larvae that reach the pasture from a dung pad. Pads attacked by beetles produced from 48 to 93 percent fewer worm larvae than intact pads.

Dramatic results have also been obtained in experiments designed to measure the effect of dung beetle activity on plant yield [*see bottom illustration on page 107*]. The coprids increase fertility primarily by dispersing the fertilizing dung through the soil. Other beneficial effects include an improvement in the permeability of soils to water. For example, five times as much water was required to produce waterlogging of a loamy soil worked by beetles as was required for undisturbed soil. There is also an improvement in soil structure, humus content and degree of aeration.

The decision was made in 1963 to begin establishing in Australia a range of beetles adequate to dispose of bovine dung pads in the most important locations and under the most prevalent conditions where cattle are raised. Two important questions had to be considered before the project could safely be undertaken. One was whether the candidate dung beetles themselves were likely to produce any adverse effects. What we knew about the feeding habits of both adults and larvae made it most unlikely that they would attack anything other than dung. Furthermore, there were no records indicating that the extremely abundant coprid fauna of Africa was causing any problems there.

The second problem was whether it would be possible to introduce a wide range of dung beetles without their being accompanied by any of the serious diseases of cattle, such as foot-and-mouth disease and rinderpest, that occur in Africa and some other places but not in Australia. Examination of the surfaces of adult beetles revealed that they do commonly carry many mites, that nematode worms shelter under the wing covers and that fungi and bacteria abound both on the cuticle, or tough outer surface, and in the excreta. The results of careful treatment with specific pesticides still did not give complete confidence that all fellow travelers could invariably



GARRETA NITENS is shown at work in these photographs made by John Green of the author's laboratory. First the beetle cuts a portion of dung out of a fresh cow pad (*top*). It shapes the dung into a ball (*middle*) and rolls ball away from pad to be buried (*bottom*).

be eliminated, and so an entirely different approach was adopted.

We found that we could transplant beetle eggs from the cavities in dung balls where they had been laid into "artificial" balls made of Australian dung, with a cavity simulating the one made by the female parent. The beetle eggs are removed from brood balls and carefully washed in a detergent solution to remove all adhering material. The clean eggs are then immersed for three minutes in a 3 percent solution of formalin, drained and thoroughly rinsed in sterile distilled water. This removes all reasonable doubts about security. Nevertheless, in our laboratory at Pretoria additional steps are taken to reduce the risk still further. The surface-sterilized eggs are transferred into moist, sterile peat moss in containers that have been sent in sealed packages from Australia. When the resealed packages are received in quarantine in Australia, artificial dung balls are prepared of a size, moisture content and consistency appropriate to the particular beetle species, an egg is transferred to each and the ball is sealed and then buried in moistened sandy soil. When one to three months later the resulting adults emerge (still in quarantine), they are allowed to mate and make their own dung balls. The eggs that subsequently hatch are surface-sterilized and are now considered clean enough to remove from the quarantine area so that mass rearing can be initiated. I have described this elaborate procedure in detail partly in the hope that it will discourage any enthusiast from attempting to import beetles without adequate precautions acceptable to quarantine authorities, a venture that might result in serious consequences for entire continents.

The first dung beetles were released in Australia in April, 1967. In the next three summers about 275,000 beetles of four species were liberated, mainly in tropical (northern) Australia. One species, *Onthophagus gazella,* has made spectacular progress [*see illustration below*]. Within two years it multiplied to countless millions and dispersed far and wide, colonizing 250 miles along the northern Queensland coast around Townsville, penetrating 50 miles inland and in the process closing the 50-mile gaps between the five individual release sites. During the first year it demonstrated its ability to make long-distance flights over water by colonizing Magnetic Island, four and a half miles off Townsville, and a year later it reached Palm Island, 18 miles off the coast.

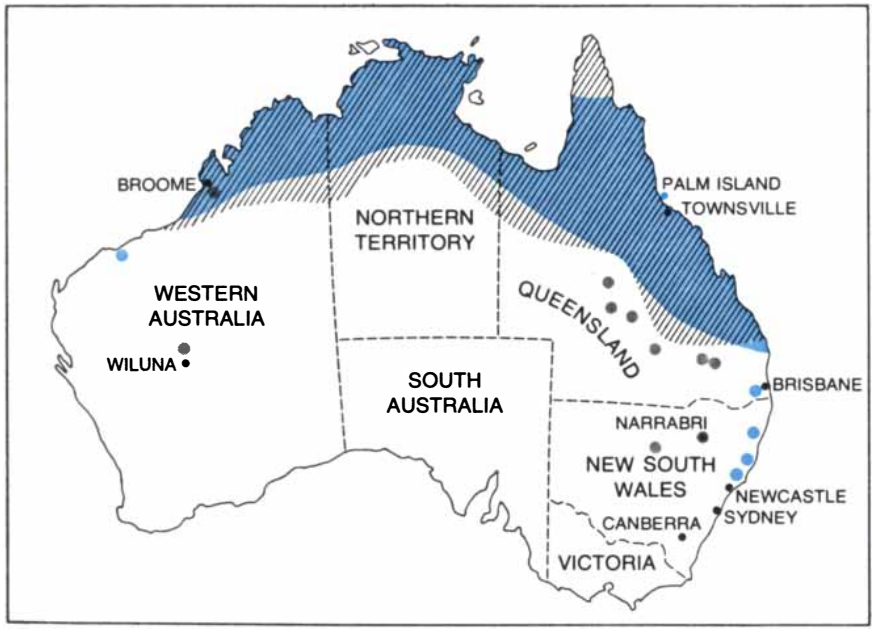
In more recent years *O. gazella* has been distributed at many sites across the northern half of Australia and is reported to have built up into very large numbers in many places. In late 1973 it was found that colonizing had been successful on the east coast almost as far south as Newcastle.

In areas where *O. gazella* has become well established dung disposal is now almost complete for part of the year. Pads may disappear within 48 hours during a considerable portion of the wet summer period, which runs from January to April in the Townsville district, and partial disposal extends over another two months or so preceding and following this period. During the peak period of the beetle's activity the buffalo-fly nuisance was somewhat reduced for the first two seasons. There has not, however, been a prolonged, noticeable reduction in the number of flies in subsequent years. The buffalo fly becomes active early in the spring, before the temperature and soil moisture are high enough for *O. gazella* to begin disposing of pads; in the fall temperatures and moisture decrease enough to inhibit *O. gazella* activity several weeks before the flies cease to breed. Clearly additional species are required to complement *O. gazella* and provide more efficient dung burial over a wider range of seasonal conditions. This long-term need was recognized from the outset, since we had observed in Africa that many species colonize pads simultaneously and that the spectrum of species usually changes markedly with the season.

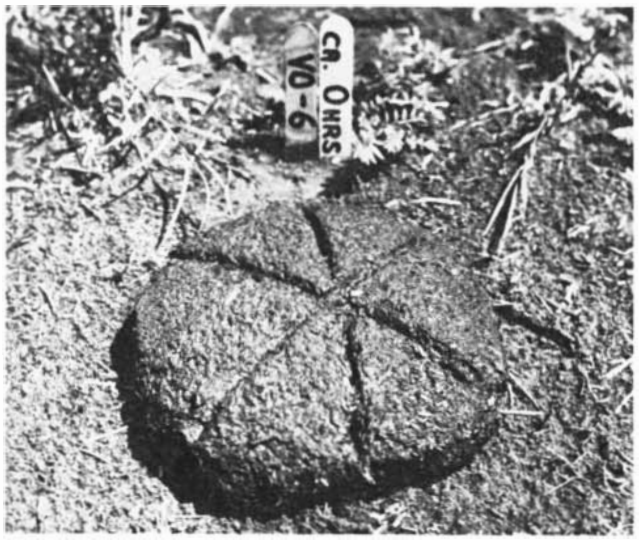
(The only complaint we have received about *O. gazella* has come from a cattle raiser who for years had used dried dung pads to level up the pipes he was using in the irrigation of his pastures. He complained that he now had to carry blocks of wood with him for the purpose!)

Two other African species, *Euoniticellus intermedius* and *E. africanus,* have increased spectacularly in northern New South Wales and southern Queensland little more than a year after their introduction. In irrigated pastures in the Narrabri district practically all dung pads are now eliminated rapidly and completely. *E. intermedius* has also increased to very large numbers in many areas where there is less rainfall, not only in central Queensland but also at places such as Broome and Wiluna in Western Australia. Already the economic returns, particularly in terms of the increased availability of pasture, must amount to hundreds of thousands of dollars a year.

In addition to the beetles I have mentioned, which are tunneling species, a number of ball-rolling species also hold great promise. One group of eight closely related species of the genus *Sisyphus* (aptly named for the miscreant in Greek mythology who was condemned to endlessly rolling a heavy stone uphill) has been studied closely in Zululand in South Africa, where all eight occur together. The fact that they are able to coexist indicates that each is adapted to



IMPORTED DUNG BEETLES were first released in Australia in 1967, primarily in the tropical northern regions afflicted by the buffalo fly (*hatching*). *Onthophagus gazella* has become particularly well established (*color*). *Euoniticellus intermedius,* released in generally cooler and drier regions (*gray*), has also increased to large numbers. Many more species are now being released that are adapted to varying climatic and other conditions.



EFFECT OF BEETLE ATTACK on a 1,000-milliliter dung pad (*left*) was observed near Pretoria in South Africa. After 24 hours less than 10 percent of the dung remained (*right*), mixed with loose soil excavated from tunnels in which dung had been buried.

EFFECT ON SOIL FERTILITY was demonstrated with Japanese millet. Six days before seed was sown a measured amount of dung and 20 pairs of dung beetles were put in the pot at left. The next pot had the same amount of dung but no beetles. Dung was placed in the third pot from left but removed before sowing. The control pot (*right*) received neither dung nor beetles. The plant with both dung and beetles took up much more nitrogen, phosphorus and sulfur than the others and its total yield was significantly greater.

utilizing either a different portion of the dung pad or pads found in particular situations in that environment. Since *Sisyphus* and other ball-rolling species obtain their dung preferentially from crevices in the pad or from around its base, they are likely to be of particular value in fly control, since these are the locations where pest flies most frequently lay their eggs.

A number of other countries where cattle are raised now but where there were no large native herbivores are also likely to benefit by the introduction of dung beetles. These include Papua New Guinea, New Zealand, various islands in the Pacific and the Atlantic, and perhaps North America, where both the horn fly and another dung-breeding cattle nuisance, the face fly, are prevalent. North America has a less diverse and less abundant dung-beetle fauna than might have been expected for the home of the bison and a region where many other large herbivores evolved. One hypothesis is that during the last ice age the herbivores that roamed the prairies were driven southward into regions where grasslands gave way to desert or jungle, and that many species became extinct along with their associated dung beetles. The precursor of the bison, according to this theory, later repopulated North America from Asia by means of a partially frozen land bridge across the Bering Strait—but left behind its associated beetles. Whether this sequence of events or some quite different one is responsible for the scarcity of North American coprids, the situation is now subject to change, and disease-free stocks of several dung-beetle species have already been supplied from Australia to the U.S.

Some attention has also been paid in Australia to another group of beetles that inhabit dung: the histerids. They do not eat dung, but both the adults and the larvae attack and voraciously consume fly maggots and pupae in the dung. Unfortunately the histerid species so far examined are not expert hunters, and only about 30 to 50 percent of the maggots in the cow pads are destroyed. Even this degree of destruction would be of value, however, and so five species have been introduced. Two of them, *Hister chinensis* and *H. nomas,* are known to be established.

The foregoing account may give the impression of a project that is complete except for the transfer of an adequate range of beetles to Australia. The ramifications of biological-control operations, however, are complex. Many new aspects are already known to require investigation and doubtless many more will emerge as the work proceeds.

One interesting question is whether or not it has been sensible to exclude from the consignments the mites that abound on the beetles in their native habitats. Carrion in the north of England is consumed competitively by blowfly maggots and carrion beetles of the genus *Necrophorus.* The beetles commonly provide transport for up to 30 mites, which appear to do the beetles no harm but which are known to attack and eat blowfly eggs and small maggots. When *Necrophorus* arrives at carrion, the mites immediately leave the beetles and move rapidly over the carcass in search of their blowfly food. If *Necrophorus* is deprived of its mites, it is unable to compete effectively with maggots for carrion. If the dung-burying activities of the introduced coprids do not reduce fly breeding

GIANT HELIOCOPRIS beetles are so large and strong that they cannot be retained in the closed hand. These are a male and a female *Heliocopris dilloni.* Such species are to be introduced from Africa into Australia in an effort to counter the effects of the toad *Bufo marinus,* which has begun to prey on the smaller imported beetles such as *Onthophagus gazella.*

to acceptable levels, it may be desirable to consider introducing some of the hundreds of species of predaceous mites that the beetles ordinarily carry with them.

A second problem requiring investigation relates to an apparent change in the behavior of the giant toad *Bufo marinus*, which was itself introduced via Hawaii in the 1930's to aid in the control of beetles that were damaging the roots of sugarcane in Queensland. (Opinion has been divided ever since on whether or not the toads provide any measure of control, although they appear to have reduced the abundance of many interesting and harmless native beetles. Certainly toads are often to be seen in summer sitting in a circle under streetlamps in coastal Queensland, waiting to consume insects attracted to the light.) Although these toads were not known to do so before, they are now observed to seek out the nearest fresh dung pad when they become active in the evening. With a flashlight one can spot them on or near a pad, swallowing *Onthophagus gazella* beetles as they land, and dissections show that a toad can consume as many as 80 beetles a night.

In order to deal with *B. marinus*, dayflying coprids are being sought, since such beetles should largely escape attack by the toads. Another possible countermeasure is the introduction of certain giant, heavily armored and immensely powerful dung beetles of the genus *Heliocopris*. Some are almost the size of a golf ball and cannot be retained in the closed hand, so powerful are the digging motions of their legs [*see illustration on opposite page*]. They fly principally at dusk or dawn, when toads are active, and it seems likely that if a toad swallows one of these beetles whole (as is the toad's habit), the beetle would be strong enough to break out through the toad's body wall. (This has been observed to happen with a small Australian frog that swallowed the native *Onthophagus cuniculus*.) Even apart from their toad-proof potential, the 50-odd species of *Heliocopris* beetles in Africa are worth serious consideration. A single pair are capable of burying a dung pad overnight, fashioning from it brood balls as big as croquet balls.

I shall close with a sobering and yet somehow inspirational indication of the formidable disposal task that faces our growing force of dung-devouring immigrant beetles. During the 30 minutes you may have taken to read this article more than six million cattle dung pads have been deposited on the surface of Australia!

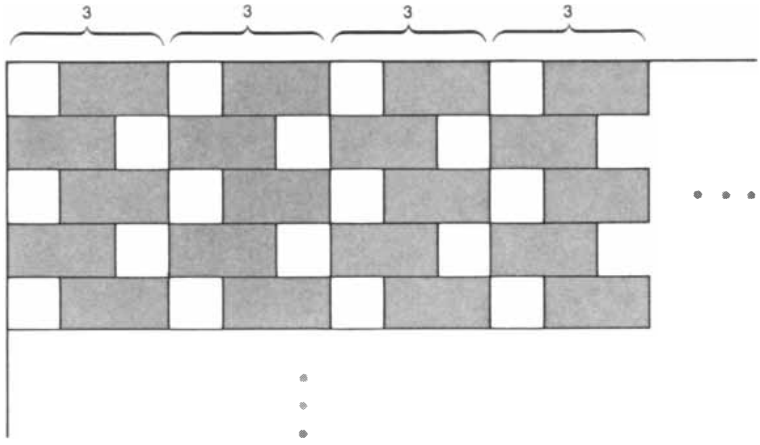
MATHEMATICAL GAMES

Nine challenging problems, some rational and some not

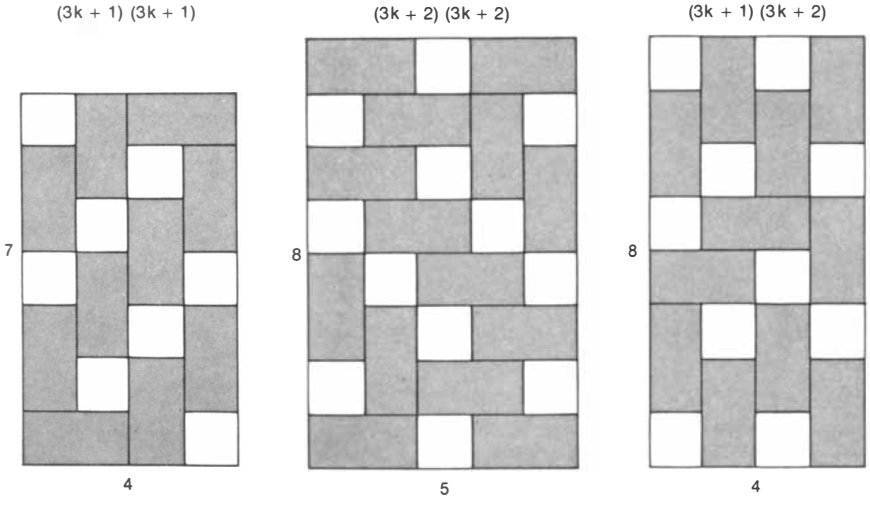
by Martin Gardner

The answers to the following set of miscellaneous short problems will be given next month. As always when such a collection is presented here, letters offering corrections, generalizations, variations, improved solutions or any other comments on the problems are welcome. It is not possible to reply to all such letters, but they will be carefully read and matters of special interest will be reported on in a later article.

1. The Gunport Problem. Last February's article introduced a game called cram. Two players alternately place a domino on a rectangular checkerboard so that the piece covers two squares. The last to play is the winner, or (in reverse play) the loser. Len Gordon and Michael Beeler have called my attention to the fact that determining the longest possible cram game on a given m-by-n board is equivalent to a pleasant and still unsolved combinatorial problem. It was introduced and named by Bill Sands in his article "The Gunport Problem" (*Mathematics Magazine,* Volume 44, September, 1971, pages 193–196).



Pattern for maximizing "gunports" when one side of rectangle is 3k



Maximum gunports on rectangles having no sides of form 3k

The problem is simply stated: What is the maximum number of 1-by-1 "holes" that can be obtained by arranging dominoes on an m-by-n field? It is assumed that m and n are each greater than 1.

Sands was able to prove that the number of holes cannot exceed the number of dominoes. He also showed that if either side of the field is a multiple of 3, a repeated pattern provides a simple way of achieving the maximum number of holes [*see upper illustration on this page*]. In other words, if one side of the field is of the form $3k$, the maximum number of holes is $mn/3$. Otherwise the maximum number of holes must be less than this.

Murray Pearce of Bismarck, N.D., writing in the November 1973 issue of the London monthly *Games & Puzzles,* page 24, conjectured that if neither side is $3k$ but both are equal modulo 3 (that is, both are either $3k + 1$ or $3k + 2$), the maximum number of holes is $(mn - 4)/3$, and if one side is $3k + 1$ and the other $3k + 2$, the maximum is $(mn - 2)/3$. Examples of how the predicted maximum can be obtained for the three types of field that have no side equal to $3k$ are shown in the lower illustration on this page. (Readers interested in subscribing to the London magazine, now in its third year, can obtain a sample copy by sending $1, or a year's subscription by sending $9, to *Games & Puzzles,* Box 1176, Palo Alto, Calif. 94302.)

Pearce's formulas set a maximum of 26 holes for the 8-by-10 field. Sands confessed he was unable to do better than 24 holes, using 28 dominoes. Can the reader find a 26-hole solution, using 27 dominoes?

I shall, of course, welcome any pattern that exceeds the maximum predicted by Pearce's formulas or any proof that Pearce's two formulas are correct.

2. Figures Never Lie. An old burlesque routine involves two simpleminded men who divide 28 by 13 to get 7, then verify this result by multiplying 13 by 7 to get 28 and finally double-check it by adding 13 seven times to get 28. This is how Irvin S. Cobb told the story in his anthology of 366 jokes (one for leap year), *A Laugh a Day Keeps the Doctor Away* (1923):

"Three patricians of the coal yards fared forth on mercy bent, each in his great black chariot. Their overlord, the yard superintendent, had bade them deliver to seven families a total of twenty-eight tons of coal equally divided.

"Well out of the yards, each with his first load, Kelly and Burke and Shea paused to discuss the problem of equal

distribution—how much coal should each family get?

" 'Tis this way,' argued Burke. 'Tis but a bit of mathematics. If there are 7 families an' 28 tons o' coal ye divide 28 by 7, which is done as follows: Seven into 8 is 1, 7 into 21 is 3, which makes 13.' He triumphantly exhibited his figures made with a stubby pencil on a bit of grimy paper:

```
7/28/13
   7
  --
  21
  21
  --
  00
```

"The figures were impressive but Shea was not wholly convinced. 'There's a easy way o' provin' that,' he declared. 'Ye add 13 seven times,' and he made his column of figures according to his own formula. Then, starting from the bottom of the 3 column, he reached the top with a total of 21 and climbed down the column of 1's, thus: '3,6,9,12,15,18,21,22, 23,24,25,26,27,28.' 'Burke is right,' he announced with finality.

"This was Shea's exhibit:

```
  13
  13
  13
  13
  13
  13
  13
  --
  28
```

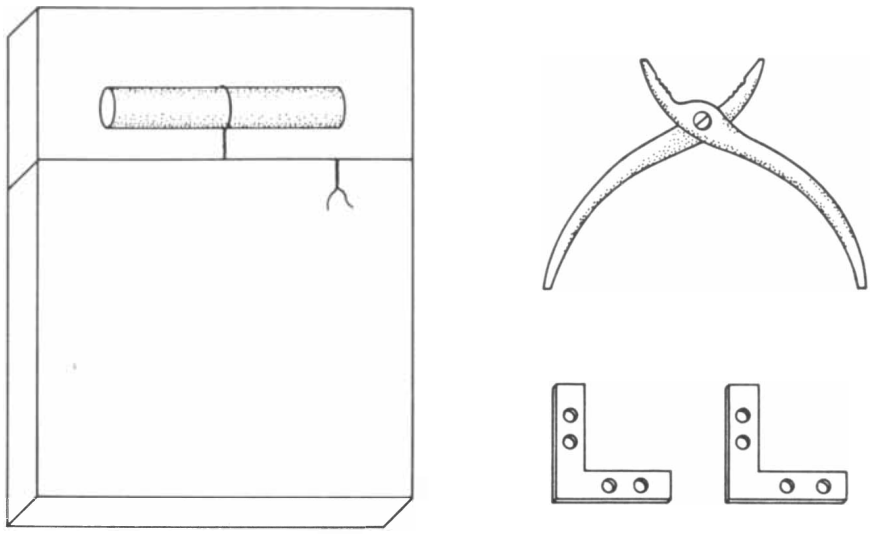
" 'There is still some doubt in me mind,' said Kelly. 'Let me demonstrate in me own way. If ye multiply the 13 by 7 and get 28, then 13 is right.' He produced a bit of stubby pencil and a sheet of paper. 'Tis done in this way,' he said. 'Seven times 3 is 21; 7 times 1 is 7, which makes 28. 'Tis thus shown that 13 is the right figure and ye're both right. Would ye see the figures?'

"Kelly's feat in mathematics was displayed as follows:

```
  13
   7
  --
  21
   7
  --
  28
```

" 'There is no more argyment,' the three agreed, so they delivered thirteen tons of coal to each family."

The comedian Flournoy Miller made effective use of the routine, and he published his version of it in his book, *Shufflin' Along*. A few years ago Flip Wilson did the bit on his television show and was sued by Miller's daughter for unauthorized use of the material. The case was apparently settled out of court.



The stand problem

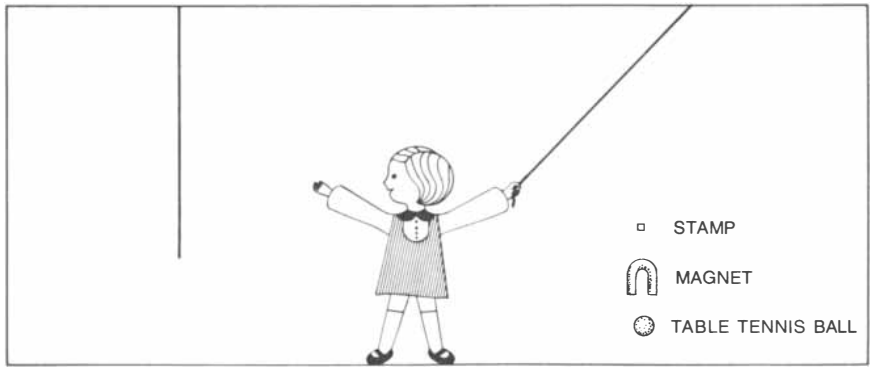
"Is there something special about the numbers 7, 13 and 28?" asked the late William R. Ransom, a mathematician at Tufts University. The answer is no. There are just 22 triplets of numbers—one number is a single digit, the other two are two digits each—that can be substituted for 7, 13 and 28 without changing a single word in the routine. Readers are asked to list the 22 triplets.

3. Functional Fixedness. Past experience sometimes has a negative effect on creative thinking. When this involves a difficulty in seeing how a familiar object can be used in an unorthodox way, psychologists call it a manifestation of "functional fixedness." Here are two problems, familiar to psychologists, that illustrate the concept:

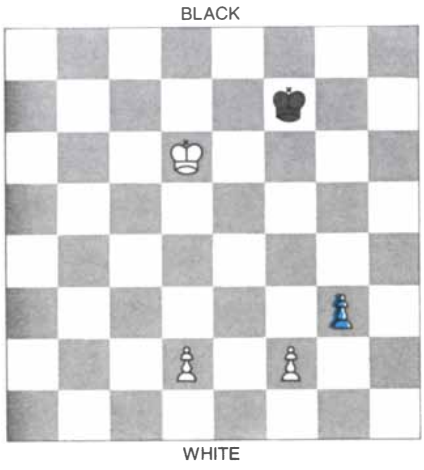
You are seated at a bare table and given six objects: a board, pliers open to maximum extent, two small metal angle irons with screw holes, a peg and a length of wire that has been used to bind the peg firmly to the board [*see illustration above*]. How can you arrange these objects so that the board becomes a horizontal stand several inches above the tabletop and firm enough to support a vase of flowers?

You are in a bare room. Two strings hang from the ceiling [*see illustration below*]. Your problem is to tie the ends together. When you grasp one end of the string, however, the other dangles many feet beyond your reach. You are not allowed to use anything you are wearing or have on your person (such as your stockings for the purpose of swinging them to catch a string), but you may use any or all of three objects on the floor: a table-tennis ball, a small horseshoe magnet and a postage stamp.

4. Monochromatic Chess. Many readers reported on how much pleasure they had derived from mathematician Raymond Smullyan's brilliant "retrograde" chess problem in last May's article. Smullyan, who is now preparing a book of his unorthodox, unpublished chess



The string problem



Smullyan's monochromatic problem

puzzles, has given permission to describe another of his problems. It is from a group based on what Smullyan calls "monochromatic chess."

The illustration above shows the position of an end game with only five men on the board: the black and white kings, two white pawns and one pawn of unknown color (*shown in color*). During the course of the game no piece has moved from a square of one color to a square of another color. Is the unknown pawn black or white?

5. The Two Bookcases. Robert Abes of the Courant Institute of Mathematical Sciences at New York University originated this problem and sent it to me last year. A room nine by 12 feet contains two bookcases that hold a collection of rare erotica. Bookcase *AB* is 8½ feet long and bookcase *CD* is 4½ feet long. The bookcases are positioned so that each is centered along its wall and one inch from the wall.

The owner's young nephews are coming for a visit. He wishes to protect them and the books from each other by turning both bookcases around to face the wall. Each bookcase must end up in its starting position but with its ends reversed [*see illustration below*]. The bookcases are so heavy that the only way to move them is to keep one end on the floor as a pivot while the other end is swung in a circular arc. The bookcases are narrow from front to back, and for purposes of the problem we idealize them to straight line segments. The ends of the bookcases cannot pass through walls in mid-swing, or through each other. What is the minimum number of swings required to reverse the two bookcases?

6. Irrational Probabilities. It is very easy to use a penny as a randomizer for deciding between two alternatives with probabilities expressed by rational fractions. Suppose you wish to decide between *A* and *B* with a probability of 3/7 for *A* and 4/7 for *B*. The number of equally likely ways a penny can fall when flipped n times is 2^n, so that three flips of the coin give eight possible triplets: *HHH*, *HHT*, *HTH* and so on. Eliminate one triplet, then pick any three of the remaining seven and designate them triplets that decide for *A*. The other four triplets decide for *B*. Flip the penny three times. If the result is the eliminated triplet, ignore it and flip three more times. Eventually you will flip one of the seven triplets. The chance that this will be a triplet in the set of three is clearly 3/7, with 4/7 as the probability that it will be in the set of four.

The procedure is easily extended to a decision between n alternatives, each with a rational probability. Suppose there are three alternatives with the probabilities $A = 1/3$, $B = 1/2$ and $C = 1/6$. Use the above procedure to decide between 1/3 and 2/3 (the sum of 1/2 and 1/6). If the decision is for *A*, you are finished. Otherwise you must continue by deciding between *B* and *C*. To do so divide 1/2 (*B*'s fraction) by 2/3 to obtain 3/4, and divide 1/6 (*C*'s fraction) by 2/3 to obtain 1/4. The penny is used as before to decide between $B = 3/4$ and $C = 1/4$. The procedure obviously generalizes to n alternatives, provided that the probabilities are rational fractions.

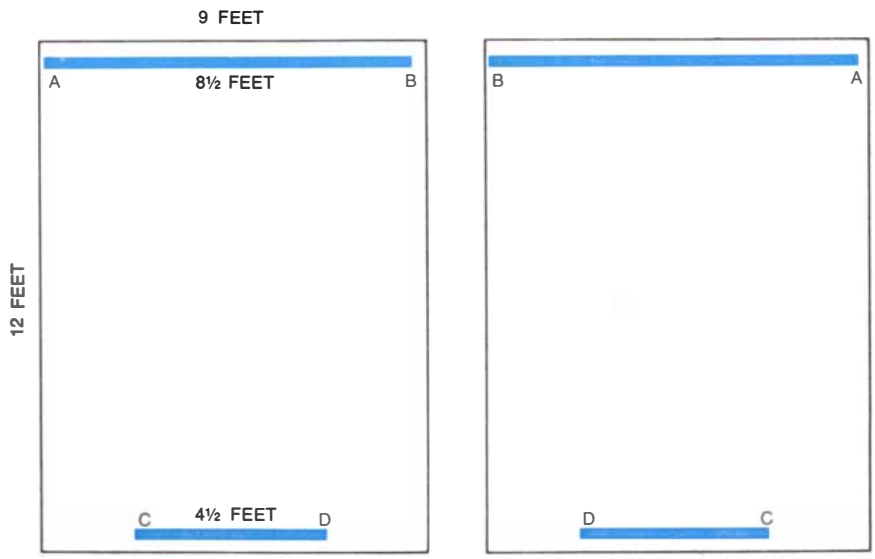
Moreover, the coin need not be a fair one. Suppose it is biased and falls heads with a probability of $1/\pi$. The probability of heads followed by tails remains equal to the probability of tails followed by heads, and so you simply flip doublets, ignoring *HH* and *TT*. Let *HT* count for heads and *TH* count for tails. With this new definition of heads and tails, each equally likely, the biased coin clearly can be used for deciding between n alternatives, each with rational probabilities.

Suppose, now, you wish to decide between n alternatives, each with an *irrational* probability. For example: *A* is the fractional part of the square root of 2, *B* is the fractional part of π and *C* is 1 minus the sum of *A* and *B*. If you can decide between two irrational probabilities using a fair coin, you can do it with a biased coin by redefining heads and tails as explained, and if you can decide between two irrational alternatives, you can decide between any number of irrational alternatives by the method given for n rational alternatives.

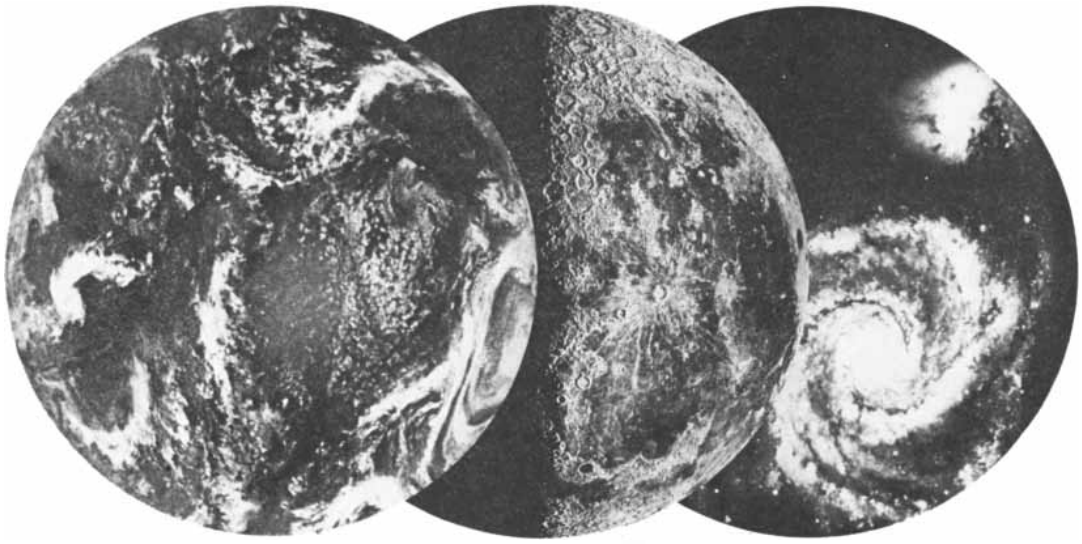
But how can a coin be used to decide between two irrational probabilities? Let us focus the problem with a precise example. $A = .1415926535\ldots$, the fractional part of π. $B = .8584073464\ldots$, which is 1 minus *A*. You wish to decide between *A* and *B* by flipping a fair coin. A delightful procedure for doing this, which applies to all irrational fractions, was recently devised by Persi Diaconis, a statistician at Harvard University, and will be disclosed next month. (Hint: The method makes use of binary notation.)

7. Who's Behind the Mad Hatter? *Word Ways*, a quarterly journal of recreational linguistics, is now in its seventh year of publication. Readers who enjoy puzzles involving the English language can subscribe by sending $7 to the editor, A. Ross Eckler, Spring Valley Road, Morristown, N.J. 07960. The following Carrollian problem, devised by John F. Collins of Santa Monica, Calif., appeared in the August 1968 issue:

"The March Hare and the Mad Hatter were sipping their eggnog and watching



The bookcases problem



Flexible Contemporary Pertinent

SCIENTIFIC AMERICAN Offprints

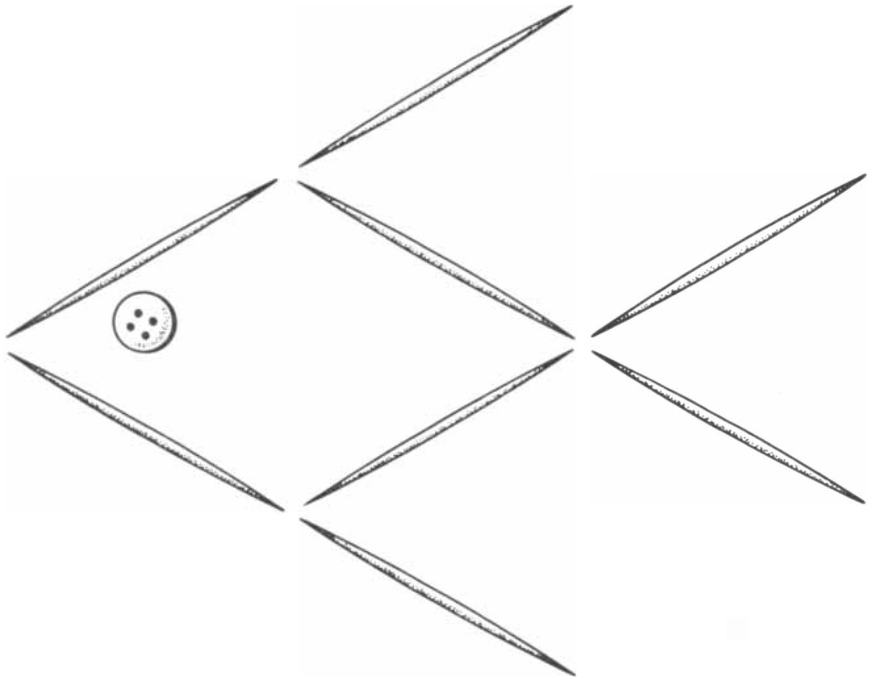
More than 800 articles are now available in the SCIENTIFIC AMERICAN Offprint series. New titles are being added constantly to this important collection of reprints from SCIENTIFIC AMERICAN. Each Offprint costs 25¢. This sampling of titles will show you how much pertinence and educational flexibility a quarter can buy.

125. Brazier: The Analysis of Brain Waves
144. Cole: The Ecosphere
1023. Kolff: An Artificial Heart Inside the Body
1068. Pirie: Orthodox and Unorthodox Methods of Meeting World Food Needs
1129. Tietze & Lewit: Abortion
1170. Berns & Rounds: Cell Surgery by Laser
1197. Singer: Human Energy Production as a Process in the Biosphere
1199. Bodmer & Cavalli-Sforza: Intelligence and Race
309. Gardner: Can Time Go Backward?
311. Alfven: Antimatter and Cosmology
319. Weisner & York: National Security and the Nuclear-Test Ban
480. Sachar: Behavioral Science and Criminal Law
483. Barron, Jarvik & Bunnel: The Hallucinogenic Drugs
498. Levine: Sex Differences in the Brain
506. Calhoun: Population Density and Social Pathology
524. Grinspoon: Marihuana
611. Leontief & Hoffenberg: The Economic Effects of Disarmament
618. Haagen-Smit: The Control of Air Pollution
626. Taeuber: Residential Segregation
631. Lewis: The Culture of Poverty
633. Comer: The Social Power of the Negro
650. Alexander: The Cost of World Armaments
656. Converse & Schuman: "Silent Majorities" and the Vietnam War
885. Wenk: The Physical Resources of the Ocean
889. Hess, Kovach, Gast & Simmons: The Exploration of the Moon

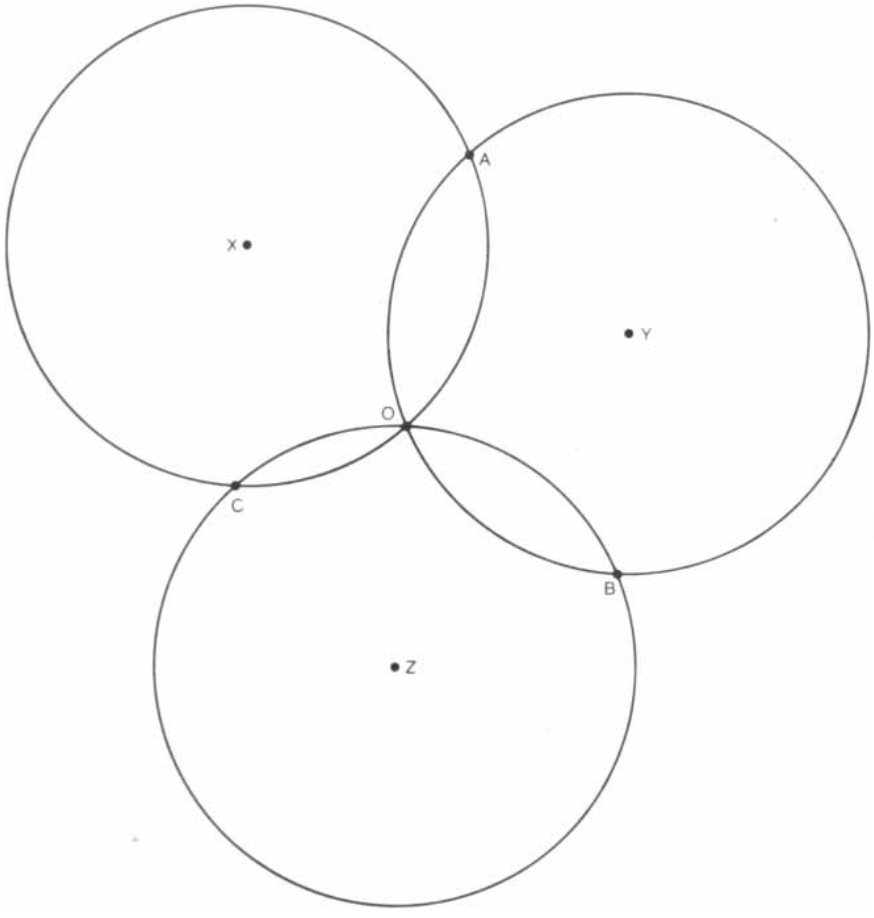
Students' sets will be collated at no additional cost. Upon request, instructors will receive a complimentary copy of each Offprint ordered in quantity for class use.

For a FREE descriptive catalogue, write to

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



The toothpick puzzle



Intersecting-circle theorem

the crowd when Alice happened to glance in the Hare's direction and ask, 'Why are you giving me such an angry look?'

" 'I'm not *giving* it to you, I'm giving it *back*,' replied the Hare.

" 'I didn't look crossly at you.'

" 'Well, *somebody* did,' the Hare said, turning to glare at the Hatter.

"Just then, someone came up from behind and put his hands over the Hatter's eyes.

" 'Guess who!', said the newcomer in a thin, flat voice.

"The Hatter froze for a moment and declared, rather coldly, 'I have no use for practical jokers.'

" 'Ha! Neither have I,' retorted the stranger, still keeping his hands over the Hatter's eyes.

"At that, the Hatter seemed to accept the challenge of the game and started asking a series of questions in a manner that mingled hope with care.

"Question: 'Ahem. Would you, by chance, be in a black suit this evening?'

"Answer: 'I would, but not by chance, by design.'

"Q. 'I presume you're a member of all the posh clubs?'

"A. 'Afraid not. Never even been invited.'

"Q. 'Surely you're better than average?'

"A. 'Yes, indeed!'

"Q. 'Not spotted, I hope?'

"A. 'Knock wood.'

"Q. 'Married?'

"A. 'No, happy.' "

Who is behind the Mad Hatter?

8. Reverse the Fish. This charming little brainteaser for children is well known in Japan but not in this country. I found it in a recent Japanese puzzle book by Kobon Fujimura. Arrange eight toothpicks and one button as shown in the top illustration at the left. Now see if you can change the position of just three toothpicks and the button so that the fish looks exactly the same as before except it is now swimming in the opposite direction.

9. The Intersecting Circles. This is one of those elegant theorems in old-fashioned plane geometry that seem at first to be exceedingly difficult to establish but that yield readily to the right insight. Three circles of unit radius, with centers at *X*, *Y*, *Z*, intersect at a common point, *O* [*see bottom illustration at left*]. The problem is to prove that the other three intersection points, *A*, *B*, *C*, lie on a circle that also has a unit radius. The problem comes by way of Frank R. Bernhart, a mathematician at the University of Waterloo in Canada.

What would you do with a Statistics and Number-Crunching Computer that starts at $6,800*, has 16K Hardwired Basic Language and 28 Major Peripherals?

Plenty!

The new Wang System 2200 is a System. It gives you the raw power and the peripherals you must have for a wide range of problem solving. For under $7,000 you get a CPU with 16K bytes of BASIC language instructions hardwired into the electronics . . . plus a 4K operating memory. You also get a big 16 lines (of 64 characters each) CRT display, a console mag tape drive and your choice of either alpha or BASIC Keyword keyboards.

Some Words About Language: The hardwired MOS ROM language in your System 2200 finally ends your dollar tradeoffs . . . economy systems that are costly to program or very expensive systems that are relatively easy to program. Many, if not most, of your people already know BASIC. They'll be solving problems the day your system is delivered (and, we can deliver in about two weeks). Most of your budget will go into problem solving; not system support.

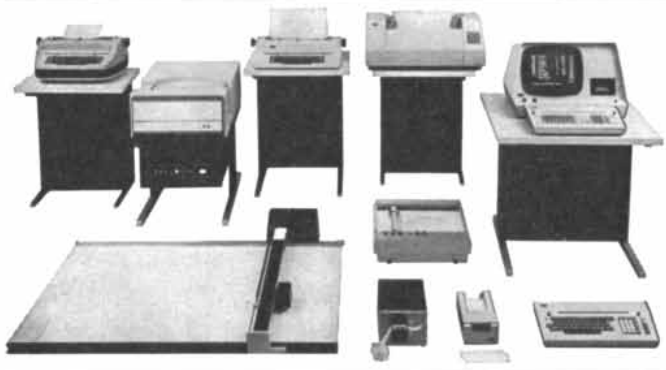
Try To Out-Grow It: Main memory is field expandable in 4K increments (at $1,500 per 4K). Up to 32K. You can choose from three kinds (and 7 price ranges) of printers . . . one even has a stepping motor for very precise 4-quadrant incremental plotting. Speaking of plots, we have a new, very large flatbed (31" × 42") for only $7,500 or a smaller one if you plot small. Both print alphanumerics and plot under full program control. Been appalled lately by disk prices? Starting at just $4,000, we offer you our new "floppy" disk in single, double and triple disk configurations (.25, .50 and .75 MB's). For big disk power, you can have 1, 2 or 5 megabyte fixed/removable disk systems. All peripherals, including punched or mark sense hopper card readers, paper tape readers and on-line BCD or ASCII controllers are easily added-on in the field so your System 2200 will grow with your needs.

The Wise Terminal: If you are now or may soon be getting into terminals, we have several new products that will instantly upgrade your System 2200 for telecommunications with any other System 2200 or a mainframe computer. And, you still have a powerful stand-alone system. Another approach, of course, is to justify it as a powerful terminal and get a "free" stand-alone computer. Wise?

We Do A Lot For You: System 2200 is backed by over 250 factory-trained Wang Service Technicians in 105 U.S. cities. Naturally, we guarantee or warranty everything you buy from us. If you want, there are free programming/operating schools here in Tewksbury, Massachusetts, almost every week. We have a growing program library on a wide range of statistics and math/science applications. Our user group (with the unlikely name of "SWAP") could help you cut programming costs even further. We do a lot for you.

* All prices U.S. List. If you're the entrepreneur type, we've just announced a new 7-module Basic Accounting System software package for the business end of your business like payroll, invoicing, inventory, receivables and some really fancy management reports.

WANG



Even if you call the Wang System 2200 a small system . . . you have to admit it's a big idea.

☐ Please send literature

☐ Please have your representative phone me:_____

Name_____ Title_____

Company_____

Address_____

City_____

State_____ Zip_____

Wang Laboratories, Inc.,
836 North St., Tewksbury, Ma. 01876, Tel. (617) 851-4111

THE AMATEUR SCIENTIST

Curious bubbles in which a gas encloses a liquid instead of the other way around

Conducted by C. L. Stong

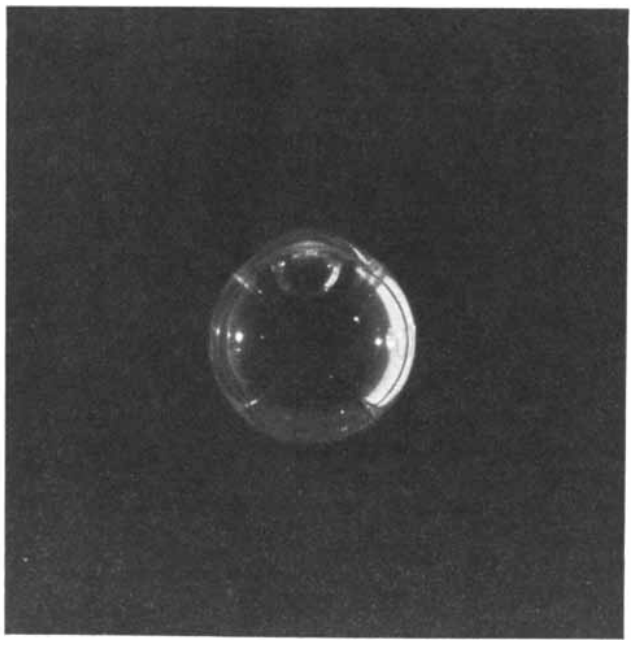
A film of liquid enclosing a gas is called a bubble. It is not generally known that one can also have a film of gas enclosing a liquid. Such objects are called globules, boules or antibubbles, depending on their size and shape, the surrounding medium and the whim of the experimenter who is describing them. Like bubbles, these objects make fascinating playthings.

Gerard Schol, a physics teacher of Drachen in the Netherlands, has done a series of experiments with drops of water that float on water. A film of air separates the drops from the supporting surface. The drops Schol launches on a collision course frequently coalesce on impact to form floating puddles that he calls globules. Additional experiments led him to conclude that the integrity of a globule can be maintained at least in part by forces of electrical repulsion arising from the polarized nature of the water molecule [see "The Amateur Scientist," SCIENTIFIC AMERICAN, August, 1973]. Kenneth C. D. Hickman, who heads the distillation research laboratory of the Rochester Institute of Technology, has made much larger bodies of water that float in water and still larger hemispherical masses of isopropyl alcohol that float in isopropyl alcohol. He refers to these remarkable objects as boules, a term he has borrowed from the manufacturers of synthetic gems.

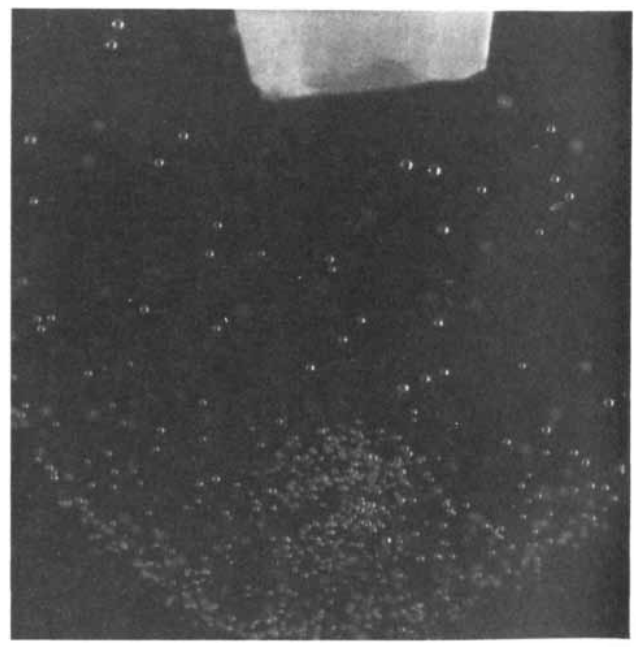
Hickman suggests that forces other than electrical repulsion may contribute to the support of Schol's globules, because experiments demonstrate that the integrity of a boule is maintained by a concave film of vapor flowing radially. To make boules Hickman designed an apparatus that superheats the supporting liquid, that is, raises the temperature of the liquid slightly above the point at which it would normally boil. At this temperature the surface liberates vapor at an abnormally high rate.

Hickman reduces the tendency of the superheated liquid to boil by continuously purifying it and draining flotsam from its surface by means of an associated retort. Liquid at the surface in the experimental flask drains through a tube into the boiler of the retort. Vapor that rises from boiling fluid in the retort enters the cool tube at the top, where it condenses. The condensate trickles down the cool tube and encounters a glass rod [*see illustration on page 118*]. The now purified liquid adheres to the rod, flows down it and drips off its rounded end. The height of the rod can be adjusted experimentally to let drops fall gently on the clean surface of the superheated liquid. Typically the flasks are made of borosilicate glass with a capacity of one liter (1.057 quarts).

Drops of condensate will merge to form a boule that floats in the superheated liquid if appropriate conditions have been established. The supporting liquid must be heated at least .15 degree Celsius above the temperature at which it would ordinarily boil. The superheated liquid must be purged of ab-



A water antibubble in water (with a smaller antibubble inside it)



Air bubbles left when antibubble is pricked by squeeze bottle **(top)**

sorbed gases. Its surface must be free of contaminating substances. The condensate must not be dropped to the supporting surface through a distance of more than about one centimeter (.4 inch). The boule and the supporting liquid must be kept electrically neutral, preferably interconnected by a thin wire bent into the shape of an inverted *U*. The upper parts of the vessel in which boules are made should be warmed above the temperature at which the vapor condenses. Finally, all parts of the apparatus that come in contact with the liquid must be scrupulously clean.

The temperature of the supporting liquid and the vapor above it can be measured by a pair of thermocouples. The probe that senses the temperature of the supporting liquid should be placed approximately one millimeter (.04 inch) below the surface and beyond the edge of the area within which boules are likely to form. The remaining probe can be put anywhere in the vapor above the liquid.

Boiling tends to purge a liquid of absorbed gas. A useful accessory for inducing a superheated liquid to boil is a glass rod that carries at its bottom an inverted test tube three or four millimeters in diameter. Bubbles that form at the open end when the tube is pushed below the surface initiate boiling.

The experimental vessel in which boules are made can be superheated by air from a short metal chimney enclosing a torus of heating wire. The temperature of the torus can be adjusted by means of a variable transformer. The application of concentrated heat at the center of the experimental vessel should be avoided. Heated air will rise by convection and warm the upper part of the vessel to prevent condensation if the unit is placed in a protective housing. The housing can be fitted with an access door and observation windows of any clear plastic that can be heated without damage to 100 degrees C.

Hickman washes the assembled apparatus with household detergent. He avoids the use of abrasives and brushes that might mar the polished glass surfaces. All parts are rinsed with concentrated hydrochloric acid, lightly rinsed with ammonia and finally rinsed for five minutes in running water. The apparatus is best used right away, but it can be dried and kept in a dust-free environment. This cleaning procedure minimizes the presence of particles and substances that could function as nuclei to initiate boiling.

Boules have been made of many liquids, including tap water. To set up an experiment, fill the experimental vessel to the level of the drain tube. Add about 200 milliliters to the retort. Apply power to the heating torus of the experimental vessel and to the mantle of the retort. The liquid in the freshly filled experimental vessel will doubtless boil spontaneously when its temperature reaches the boiling point. (Water normally boils at a temperature of 100 degrees C. under an atmospheric pressure of 760 torr, or 14.7 pounds per square inch.)

Let the liquid in the retort boil continuously, but gradually lower the temperature of the torus until boiling stops in the experimental vessel. Then raise the temperature again. The liquid becomes increasingly pure as the retort continues to operate. Ultimately liquid in the experimental vessel can be superheated by .15 degree C. without spontaneous boiling.

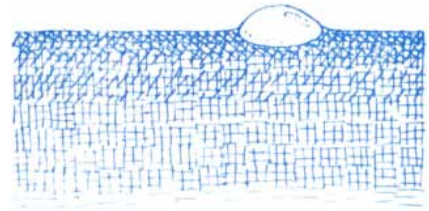
At this stage induce boiling by thrusting the inverted test tube below the surface. Lower the temperature until the induced boiling stops. This procedure purges the liquid of absorbed gas. Repeat it three times. Thereafter the surface will remain undisturbed at a temperature of .15 degree C. above the boiling point.

Drops of distillate that fall from the drip rod through a distance of not more than one centimeter will float and accelerate across the superheated surface. Many of them will merge with the supporting liquid after moving a few centimeters. Others will proceed to the edge, rebound from the meniscus and continue to accelerate. The boatlike film of vapor that carries the drops is thinnest at its forward edge, where a bow wave forms, and thickest astern, where a turbulent wake resembling the exhaust of a rocket marks the escape of vapor.

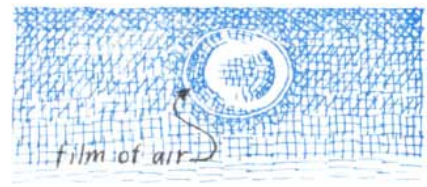
Lower the drip rod toward the surface until it is no more than three-fourths of the diameter of a drop above the surface. A drop will form, its bottom resting on the superheated liquid and its top anchored by surface tension to the rod. A metal wire *must* electrically interconnect the drop and the supporting liquid.

The drop will slowly expand into a constantly pulsing boule. The motion is caused by the radial flow of vapor from the center of the concave film below the boule. The vapor escapes at the circular interface between the boule and the supporting liquid.

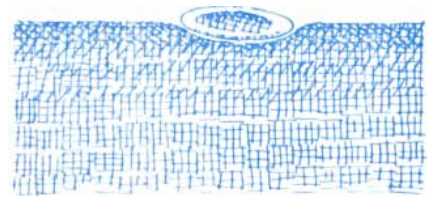
The radial flow also imparts two internal motions to the liquid of the boule: a ring vortex and a downstreaming vortex or whirlpool that is encircled by the ring vortex. The whirlpool discloses its



Air bubble



Antibubble



"Globule"

presence by a shallow depression that forms at the center of the upper surface of the boule [*see top illustration on page 119*]. The ring vortex can be observed by adding a minute crystal of potassium permanganate to the condensate at the point where it flows from the drip rod. The complex flow of the supporting liquid can be investigated with dye of a contrasting color.

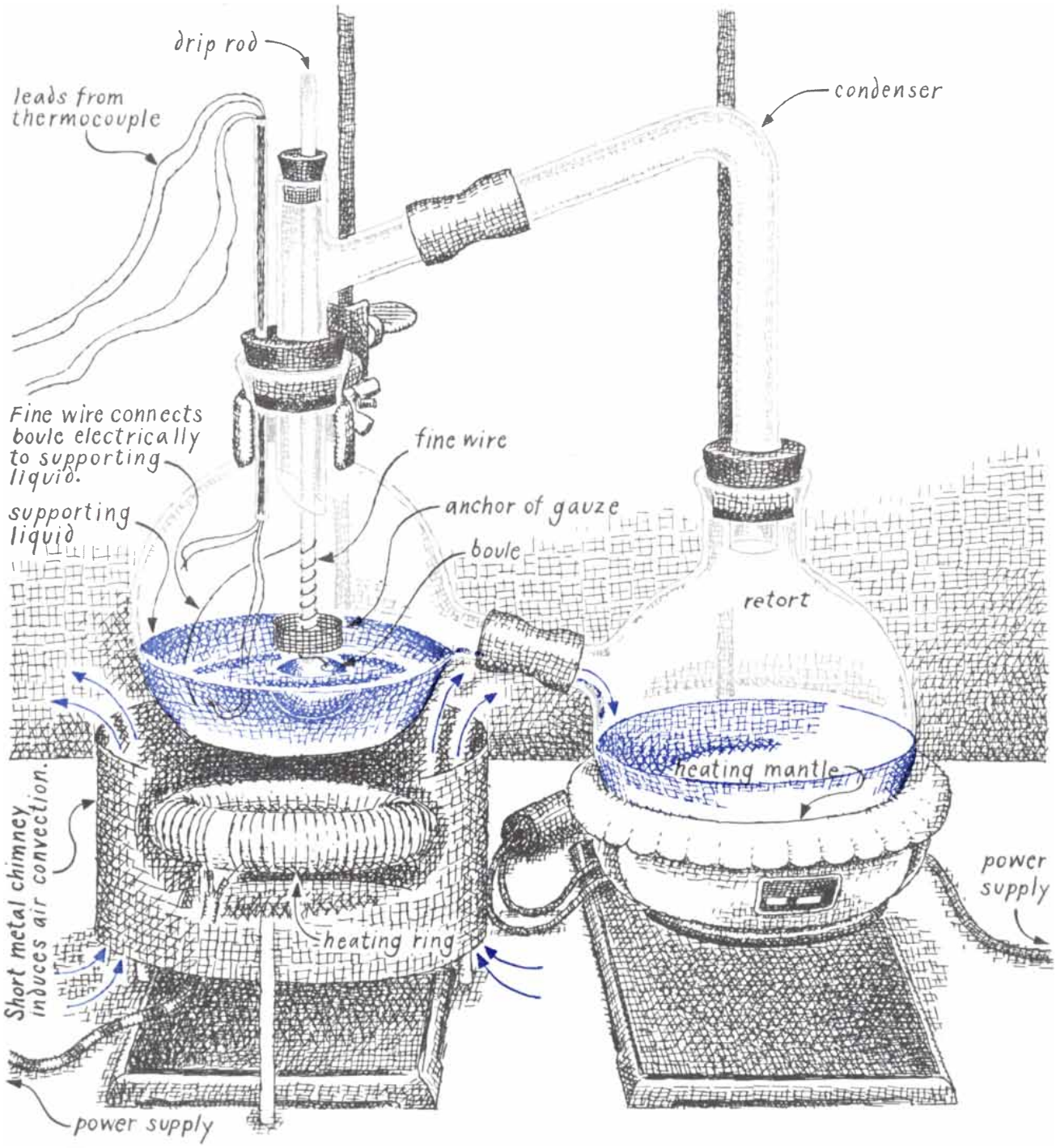
When a boule has grown to a critical size, it either merges with the supporting liquid or overcomes the restraining force of the rod, whereupon it breaks away, accelerates to the wall and merges with the supporting liquid. The critical size depends partly on the purity of the liquid but mainly on the degree of superheating. Another limit is fixed by the tendency of the boule to wobble around its point of suspension. The tendency can be decreased by attaching to the rod a ring (or concentric rings) of wire gauze. Boules that grow beyond the restraining force of the rod can be anchored by lowering the rod so that a meniscus forms between the bottom edge of the gauze and the upper surface of the boule.

The ultimate size to which a boule can be grown, according to Hickman, depends significantly on the patience of the experimenter. Hickman has grown a boule of water to a volume of 65 milliliters (about two fluid ounces) and one of

isopropyl alcohol to 250 milliliters (about eight ounces). The rate of growth increases significantly with the temperature of the superheated liquid. The velocity at which the liquids and the vapor move increases similarly. The resting surfaces begin to ripple when the temperature of the supporting liquid is increased above .15 degree C. When it is superheated to five degrees or more, violent wave motion usually destroys the boule.

Although boules appear to have no immediate practical application, they can function as a remarkably sensitive indicator of surface contamination. Hickman and his colleague Peter Harris have demonstrated that boules grow more readily and to larger sizes in a vessel of borosilicate glass that has been well leached through use than they do in an identical vessel of new glass. When liquid at the surface of both the old and the new vessel drains into the retort through a weir, boules form in both vessels. After the vessels have been equipped for subsurface drainage, however, the relative rate of production and the lifetime of the boules decrease significantly in the new vessel. The difference can be explained only by supposing the surface of liquid in the new vessel is contaminated by flotsam dissolved from the flux in the borosilicate glass.

Fully as interesting as Hickman's boules are objects that consist of a sphere of water enclosed by a spherical film of air submerged in water. Last



Modification of Kenneth C. D. Hickman's apparatus for making boules

summer J. E. Connett of the department of mathematics at Northern Illinois University described them as follows in a letter.

"It is possible with a small amount of care and very little equipment to produce a 'bubble' in water that consists of a drop of water surrounded by a thin spherical shell of air. I have succeeded in making such bubbles up to about a half-inch in diameter. They appear to be only slightly less dense than the surrounding water, because they rise sluggishly toward the surface and float just below it. A friend who has observed some of my experiments calls them antibubbles. Often the antibubbles break below the surface, leaving a trail of tiny air bubbles.

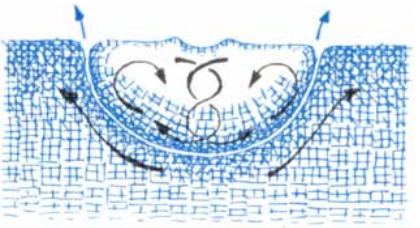
"Here is a fairly workable method for producing antibubbles. Obtain a medicine dropper or a pipette with a fairly large nozzle. Add to a glass of tap water a few drops of liquid synthetic detergent (not soap). Liquid Lux and Swan appear to work best. Partly fill the dropper from the glass and then hold it vertically about one centimeter above the surface of the solution in the glass. Squeeze the rubber bulb to release liquid but do not squeeze too vigorously. Several trials may be necessary before the experimenter acquires the knack of forming antibubbles. Make sure there are no air bubbles in the tube of the medicine dropper.

"Usually small antibubbles break within half a minute. I have no complete explanation for these remarkable objects, although I think it is easier to see how they might persist after they have been created than how they are created in the first place. I would guess that the role of surface tension is quite different in the antibubble from what it is in an ordinary soap bubble floating in air. I suspect that the only effect of surface tension in the antibubble is to cause the enveloped water to assume a spherical shape.

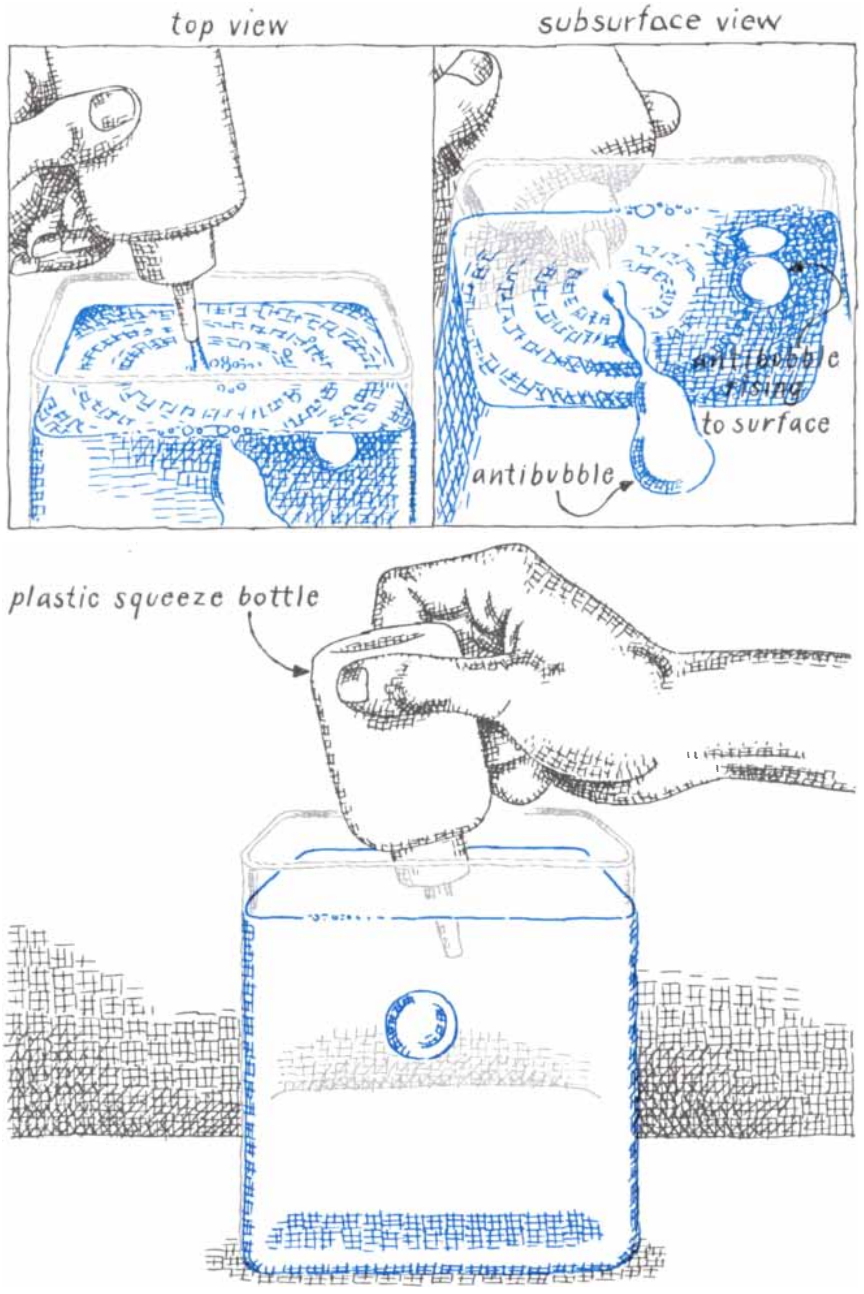
"The thin film of air around the central liquid appears initially to have uniform thickness, but in time the film becomes thinnest at the bottom and thickest at the top. I suspect that if it were possible to create an antibubble in detergent-free water, the object would persist just as long as one that is made in detergent solution. I have no idea how long antibubbles might last under ideal conditions, how large they can be made or whether they can be made in other mediums. I should not be astonished to learn that they are a well-known phenomenon, perhaps a forgotten curiosity of the 19th century, although I have found no mention of them in standard references or in C. V. Boys's classic lectures on soap bubbles."

Recently Connett wrote again. "After poking around in our library," he said, "I learned that I am indeed not the first discoverer of antibubbles. A short letter by W. Hughes and A. R. Hughes, published in *Nature* in 1932, describes how they formed antibubbles with soap in a trough of rainwater. Their antibubbles measured only one millimeter to four millimeters (.04 to .16 inch) in diameter. They concluded that the drops were enclosed by a film of soap.

"During recent months I have succeeded in making antibubbles in cold water, salt solution, sugar solution at room temperature and flat beer. A few drops of liquid detergent must be added to each of these liquids. By means of a plastic squeeze bottle with a nozzle about three millimeters (.12 inch) in diameter I have made antibubbles with a diameter of more than an inch and a half. They are less stable than the small-



Side view of currents in a boule



J. E. Connett's apparatus for making antibubbles

er antibubbles but are far more striking in appearance.

"It is possible to keep an antibubble submerged by directing a jet of solution downward at it with the nozzle of the squeeze bottle. It is even possible with sufficient practice to inject a smaller antibubble into a previously formed large one. The experimenter learns by trial and error just how abruptly and forcefully to squeeze the bottle.

"As I have mentioned, antibubbles usually break soon after they rise to the surface. To prevent them from rising I added salt to the detergent solution from which the antibubbles were formed and injected the resulting antibubbles of higher density into a detergent solution of plain tap water. Much to my dismay the antibubbles broke when they sank to the bottom.

"After more experimentation I managed to surmount this problem by coating the bottom of the container with a layer of honey roughly one centimeter thick in which density increased with depth from about 1 to 1.4. To form the layer I poured a few ounces of honey slowly into the container of plain detergent solution. The honey in the slender entering stream spread across the bottom of the container and diffused upward within a few minutes to form the desired layer. The depth of the vessel was approximately 18 centimeters (seven inches).

"When antibubbles were made in this container of liquid, I observed an unexpected effect. The antibubbles fell rapidly to the layer of honey, bounced upward and (after several oscillations) came to rest just above the honey solution. Some broke spontaneously. Others that appeared to have about the same density as the surrounding solution hung for a time and then rose slowly to the top. Usually they made the trip to the surface in two or three minutes, but I observed one that required 7½ minutes to ascend.

"A rising antibubble accelerates as it approaches the surface. Doubtless the cause is in part the diminishing hydrostatic head of pressure that allows the film of air to expand proportionately. I have repeated the experiment many times with a variety of solutions. The results are invariably the same: some antibubbles break spontaneously on the bottom and others become buoyant. The behavior does not seem to depend on the relative temperature of the antibubble and the solution.

"For a time I suspected that the change in density of the enclosed solution might arise from the diffusion of water molecules through the thin barrier of air—a sort of osmotic effect across a membrane of air. Molecules of water that evaporate from the solution might condense on the enveloped sphere until both solutions approach the same concentration. I discarded this hypothesis, however, when I observed that the gaseous film surrounding an antibubble clearly absorbs air at a higher rate when the experiment is made with a solution of fresh tap water. The mysterious added buoyancy is contributed not by the dilution of the enveloped solution but by air released from the tap water that joins the thickening film of the antibubble.

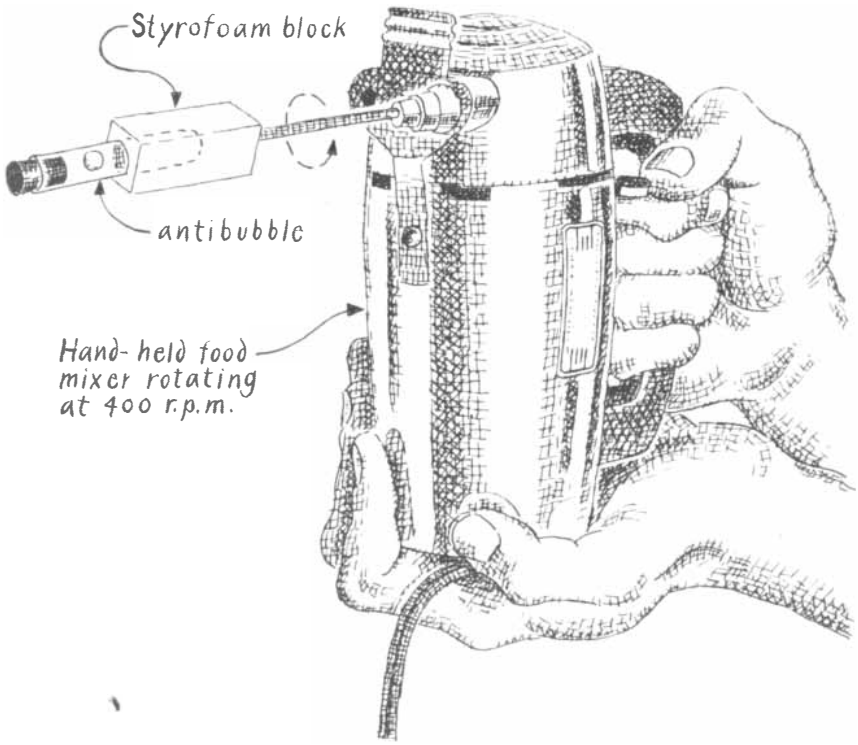
"In recent experiments I have enjoyed the cooperation of Ewing Lusk, an enthusiastic investigator and gifted amateur photographer, who succeeded in photographing many antibubbles as well as the minute air bubbles that remain when an antibubble breaks. We concluded that an antibubble could survive indefinitely if a technique were devised for counteracting the effect of gravity. I wondered what would happen to an antibubble trapped at the center of a spinning vessel of solution. Lusk and I decided to try the experiment.

"We made an antibubble in a conventional container. Then we manipulated a solution-filled test tube over it, let the antibubble rise inside and corked the tube. Next the test tube was inserted into a snugly fitted Styrofoam holder, which was attached to the shaft of an electric food mixer. We turned the mixer on. The antibubble moved to the center of the test tube and stayed there.

"On our first attempt the antibubble lasted until the food mixer began to overheat. We had to shut the mixer off after about 15 minutes. The antibubble then drifted to the wall and promptly broke. An antibubble of the same size in an ordinary container would not last more than two or three minutes. We have repeated the experiment a number of times with the same result.

"The most frantic step in this experiment is confining the antibubble in the test tube. We made the antibubble in a deep dishpan that contained a fresh detergent solution of tap water. The test tube was immersed and filled completely. One of us made the antibubbles. The other manipulated the inverted tube vertically to a position where an antibubble would rise into it. The tube was promptly corked under water and inverted to prevent the antibubble from rising to the stopper and breaking.

"Next, we began madly twirling the tube with our fingers. With continued axial twirling and a vertical reversal or two we made the liquid spin at a rate sufficient to hold the bubble near the axis. We then quickly pushed the test tube into the Styrofoam holder and switched on the mixer. Our test tube was 15 centimeters (six inches) long and 2½ centimeters (one inch) wide. The mixer turned at 400 revolutions per minute,



Apparatus for spinning an antibubble

but we believe a speed of 100 r.p.m. would have been enough to keep the antibubble from striking the wall. We think an antibubble that survives until the mixer is turned on will last as long as the machine runs.

"I determine the approximate thickness of the air film surrounding an antibubble with nothing more than a beaker of clear glass. The antibubble is made to rise into the inverted, water-filled beaker and to break. I estimate the diameter of the rising bubble; after it breaks I estimate the diameter of the resulting air bubble. Next I calculate the surface area of the antibubble, which is equal to 12.57 times the square of its radius, and the volume of the air bubble, which is equal to 4.19 times the cube of its radius. Finally I divide the volume of the air bubble by the surface area of the antibubble to find the approximate thickness of the film of air.

"The estimate may err on the high side because most antibubbles have a thickish bulb of air at the top. Even so the estimates seem reasonable. For example, the area of an antibubble one centimeter in diameter would be $12.57 \times .5 \times .5 = 3.14$ centimeters. The approximate volume of an accompanying air bubble .152 centimeter in diameter would be $4.19 \times .076 \times .076 \times .076 = .0018$ centimeter. The thickness of the air film would be approximately $.0018/3.14 = .0005$ centimeter, which is equivalent to .0002 inch.

"Hickman mentioned in one of his articles a baffling phenomenon that has also puzzled Lusk and me, namely that it is sometimes inexplicably difficult to make floating globules and antibubbles. As Hickman put it: 'They are very capricious: they might refuse to form on a cold dry day or at the approach or retreat of a tall person or a short person.' One evening Lusk and I were trying without success to make antibubbles form in a jar in his living room. When we carried the jar into the kitchen, they formed easily. We repeated the experiment several times with the same result. Then we opened the window in the living room to get some fresh air. Thereafter antibubbles formed readily in the living room.

"Hickman reports that he has twice had samples of water and of carbon tetrachloride that for 10 days refused to form boules. Then, for reasons still unknown to him, they lost this property and never regained it. Experiences such as these convince me that anyone who takes up this hobby will encounter enough puzzles to keep him delightfully busy."

Buried cities · polar bears · penguins · dragon lizards · drifting continents – sooner or later all the natural world reveals its secrets to members of the

NATURAL SCIENCE BOOK CLUB

Just about every fascinating facet of the natural world comes under the scrutiny of the NATURAL SCIENCE BOOK CLUB in books sure to delight the eye, quicken the pulse, expand the intellect, transport the mind to times past and times ahead. You're only a postage stamp away from enthralling adventure at welcome savings!

(retail prices shown)

50281. GIANT REPTILES. *Sherman A. Minton, Jr., and Madge Rutherford Minton.* A storehouse of information of the giant reptiles of the world — the crocodiles and alligators, the giant tortoises of the Galapagos and Aldabra islands, the sea turtles, the dragon lizards of Komodo, and more. Illustrated. **$9.95**

47340. EXPLORING THE OCEAN WORLD: Revised and Expanded Edition. *Edited by C. P. Idyll.* This richly illustrated volume is the non-professional's ideal introduction to oceanography. Leading scientists explain continental drift, marine ecology, exotic underwater life, man's future in the sea, much more. **$14.95**

55561. THE INSECT SOCIETIES. *Edward O. Wilson.* "For anyone, layman or specialist, interested in a single, concise, lucid, and authoritative account of the most significant facts and theories about insect societies, *The Insect Societies* is the best available and will be for many years." —*Natural History. Counts as 2 of your 3 books.* **$20.00**

Natural Science Book Club
Riverside, New Jersey 08075 4-79A

Please accept my application for membership and send me the three volumes indicated billing me only 99c each. I agree to purchase at least three additional books during the first year I am a member at special members' prices. 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 99c EACH. Indicate by number the 3 books you want.
Some expensive books count as 2 selections.

Name_____

Address_____

City_____ State_____ Zip_____

(Offer good in Continental U.S. and Canada only. Prices slightly higher in Canada.)

Take any 3 books
(values as high as $50.00)
for only 99¢ each

If you will join now, and accept only 3 more books within the next 12 months — all at stunning member discount prices.

63820. A NATURAL HISTORY OF GIRAFFES. *Text by Dorcas MacClintock. Illustrations by Ugo Mochi.* If you loved artist Ugo Mochi's sensitive silhouettes for *Hoofed Mammals of the World,* you'll love this perceptive and informative study of the members of the giraffe family. **$5.95**

75060. SAFARI SOUTH AMERICA. *Christina Wood.* The dramatic, and sometimes dangerous incidents surrounding a jungle rescue mission to save several thousand wild animals trapped by the rising waters of a hydro-electric dam. Fascinating observations of jaguars, capybaras, boa constrictors, and Saki monkeys, among others. **$7.95**

68720. PHYSICAL FITNESS & DYNAMIC HEALTH. *Thomas K. Cureton, Jr.* One of the world's truly famed authorities takes a hard look at the problem of maintaining good health in an age of convenience. Sets forth practical and original views on what physical fitness really means. Recommended for men and women of all ages. **$7.95**

88120. YELLOWSTONE: A Century of the Wilderness Idea. *Ann and Myron Sutton.* A lavish celebration — in evocative words and breathtaking photographs (72 of them in color) — of deer, bison and bears, awesome canyons, the spectacular lakes and forests of this great century-old national park. *Counts as 2 of your 3 books.* **$25.00**

35751. BEGINNER'S GUIDE TO ARCHAEOLOGY. *Louis A. Brennan.* The techniques and background information that will enable anyone to find, survey, test, excavate, preserve, analyze and record artifacts and other evidences of prehistoric life. Includes state by state lists of existing sites. **$9.95**

69621. POWER PLAY: Oil in the Middle East. *Leonard Mosley.* Behind-the-scenes look at the explosive oil situation in the Middle East. Offers clues to the political and environmental crises we find ourselves in today as well as being a crackling good history. **$10.00**

36141. BEYOND STONEHENGE. *Gerald S. Hawkins.* The internationally famous astronomer and author assesses the astronomical significance of ancient sites such as the temples of the Nile, Mayan ruins, the enormous linear drawings in the Peruvian desert, and the huge geometrical earthworks of the U.S. **$10.00**

61210. MATHEMATICAL THOUGHT FROM ANCIENT TO MODERN TIMES. *Morris Kline.* Magnificent survey of the "queen of sciences" will delight and enlighten anyone whose curiosity responds to the excitement of ideas. *Counts as 2 of your 3 choices.* **$35.00**

69711. THE POLAR WORLDS. *Richard Perry.* A fascinating parallel between the environments of the Arctic and Antarctic and the native species which inhabit them; in Antarctica, the penguins, seals, and whales; and in the Arctic, the walruses, caribou, and polar bears. **$7.95**

48700. FOOD IN HISTORY. *Reay Tannahill.* Pioneering work drawing on archaeology, biology, economics, medicine, and other disciplines to show how diet and eating habits have shaped man's development. Lavishly illustrated. **$15.00**

54100. HUMAN SEXUALITY: Second Edition. *James Leslie McCary.* A best seller in its original trail-breaking version, this revised and updated edition of the best marriage manual available provides sound and sensible information on the physiological, psychological, and sociological aspects of sex. **$14.50**

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 three to four 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 book at all, simply indicate your decision on the reply form always enclosed with the **News** and mail it so 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. • After completing your trial membership, you will be entitled to take advantage of our fabulous bonus plan which offers savings up to 70% and more.

BOOKS

The vicissitudes of building stones and the nature of biological clocks

by Philip Morrison

STONE: PROPERTIES, DURABILITY IN MAN'S ENVIRONMENT, by E. M. Winkler. Springer-Verlag ($33.70). The traveler who gazes admiringly on the Parthenon is pretty sure to take pleasure in the color of those marbles, a subtle honey-pink in the bright Athenian sun. The Pentelic quarries furnish no such warm marble; sections restored with stone newly taken from the ancient quarries are a cool white, almost blued by the contrast. Traces of ferrous carbonate are found in the rock. This easily dissolves in pore water over the years, particularly if the water is acidified by dissolved carbon dioxide. Diffusing through the water outward to the surface, the compound oxidizes to a thin tarnish of yellow limonite once the pore water evaporates. A similar process may account for the slow darkening of light-colored stones and rock in arid deserts, the "desert varnish" in which the Nazcans long ago marked out giant figures across the high Peruvian desert to delight their gods.

Men build with stone, which is the name for worked rock. Our cities are indeed desertlike, and in them rock weathers more severely than in green and humid climates. Cities increase temperature contrast even though they lessen direct sunshine; they reduce wind speed and very much enhance the frequency of fog. In cities stones become hotter, receive more moisture, dry less thoroughly and gain salts from polluted air and from neighboring weathered stone. Fog droplets hang suspended in the still air, "able to scavenge the atmosphere of suspended pollutants and slowly react with stone ... more destructive ... than many deleterious ingredients." Most cities treat their stones much worse than the Acropolis does.

Professor Winkler is a Notre Dame geologist with "lithic interests that border on genuine personal affection for commercial stone," as the foreword by a Stanford colleague explains. Winkler has assembled a thorough and remarkably accessible treatise on the hows and whys of the properties of stone and of their changes, based on full technical analysis and experiment but presented so that a wide class of readers—those who work stone and design for it no less than those who make its study their profession—might profit. A figure sculptured in porous sandstone, from a castle in the Rhein-Ruhr, emphasizes his concerns. In that industrial valley, the time-lapse photographs show, exposure from A.D. 1702 up to 1908 induced only light damage, stains and uneven weathering, but by 1969 all features are lost and what remains is only a ruin of a human figure. Acid rains, their carbon dioxide and sulfate ion arising mainly from automobile engines and coal fires, wreak most of the damage. A well-documented case leads to a graph for carbonate rocks: a linear deepening of the damaged layer by solution at a rate of one millimeter in 50 years in a rural setting and somewhat more than twice as fast in the suburbs. Industrial sites are clearly more extreme.

Such rates fit the measurements of marble gravestones as well as theoretical calculations based on the solubility of the rock, but individualities of texture, structure and impurity are complex and influential. Indeed, the theme of the volume is the inherent complexity of rock and therefore of its appearance and behavior. The pasts of beach and desert, of volcano, reef and fault are frozen in the inhomogeneities of rock composition and structure. The work of water and ice, of bacteria and of corrosive ions depends on all the properties of rock, particularly on the soluble impurities and on the hidden network of pores and channels through which water can travel. Add to that list the diversity of weather change and industrial siting. The sulfate content of rural English air is only a seventieth of the concentration measured in the terrible killing London fog of 1952, and downtown Los Angeles has registered about 100 times the nitrate concentration of average air.

The earth is not static; most rocks were stressed in the geologic past. The fractures this stress induced determine the size of stone it is economic to recover from any quarry. Once the rock is removed the open pit tends to attain a new equilibrium, its burden having shifted. "A vertical cut by saw or torch can close while cutting is still in progress" or only after centuries. Granite blocks should be "aged" before use, both to case-harden them and to relieve stress. A granite can increase in compressive strength by a third in only six month's aging through recovery from the natural prestressing.

Is there a remedy for diseased stone? The chapter on stone conservation is brief and not very cheerful. It is easy to seal the surface: "liquid hot wax [has] been applied for centuries." Moisture comes along many complex internal routes, however, by capillary flow from the ground or from higher drainage levels. Cleopatra's Needle in London was promptly sealed and has been protected; its twin in New York, on the other hand, was treated tardily, 10 years late, and the humidity had already invaded the stone. By now its south face is heading toward complete destruction, although the east face remains fully legible. There is no hope for a panacea. Stone conservation is like the practice of medicine, a healing art based on the sympathetic scientific understanding of the importance to the individual case of a bucketful of poorly known variables. "Unwise treatment can do more harm than good"; Hippocrates also bound the physician at least not to do harm.

The volume is expensive and handsome; its graphs and pictures, some in color, are instructive and convincing. The entire work is a model for the exposition of a complex body of science in a form that is effective at many levels.

THE PHYSIOLOGICAL CLOCK: CIRCADIAN RHYTHMS AND BIOLOGICAL CHRONOMETRY, by Erwin Bünning. Springer-Verlag, New York ($7.80). The subtle and pervasive clock rhythm that beats in many forms of life became credible for most biologists, a demonstrable part of the science of biology, with the first edi-

tion of this survey by Professor Bünning, which appeared in German in 1958. Today the physiological clock is the subject of about 600 papers a year. The succinct, modest, yet thorough review of the whole matter in the hands of its master investigator has been a popular book indeed, its clear language, rich graphical data (no photographs, however) and complete references fully justifying the wide sale. Only six years after the second English edition this new paperbound version carries the narrative along to 1973.

It is appropriate to note mainly the changes the past six years have brought, indicating the content of the entire book only by inference. Where is the clock? We still do not know. The isolated eye of a marine invertebrate shows a circadian rhythm of nerve activity after days of unbroken darkness, and yet the rhythms within the body of the creature are retained even after the eyestalks have been removed. Most rhythms seem merely to exhibit the time; they are "hands" for the clock but are not part of the works. For example, brainless silkworm moths emerge at random by day or night. Implant the brain "loose in the abdomen," however, and the moths come out periodically. Surely some hormonal signals from the brain, established by the day-night cycles of its past, set this particular "dial" reading. Yet it remains true that no single gland can be shown to hold the master clock.

The setting of the clock, entraining its near-24-hour rhythm to the steadier spin of the earth in a definite phase, is widely found to be mainly the result of day-night changes in light intensity. (Temperature compensation is well known to be excellent for these clocks of life, even without quartz or invar.) Measurements by Bünning's own group at Tübingen have lately demonstrated the fascinating fact that the dim natural-daylight intensity reached about half an hour before sunrise and again half an hour after sunset gives the "best reference values for a physiological process." At those hours the rate of change of light intensity is at a maximum, so that the random disturbing effects of cloud cover are minimized, amounting to errors of only a couple of minutes a day. A few days' averaging might well do better than that. And so the basis for the accurate clock—not yet a chronometer—used in orientation and navigation by many forms of life from water skimmers to carrier pigeons begins to be understandable.

Theorists are plainly attracted to this body of facts, so naturally cast into mathematical form. The study of transients, of entrainment, of phase shifts, of the damping of the master clock and of various "dials" is a rich field for the model-builder. A recent feedback model (developed by H. G. Karlsson and A. Johnsson) is able to embody many features of the experimental results, although of course it has a restricted value since it does not locate the clock circuits either in space or in their chemical nature.

Most biochemical oscillations are no part of the clock. It is classical that enzyme inhibitors such as cyanides do not stop the clock, even though they may efface this or that "dial." Only when respiration, say, drops to nearly a lethal level does the clock respond. Alcohol and heavy water do slow the clock; adding heavy water to drinking water can slow down the activity rhythm of mammals by more than an hour a day. Perhaps some general fluid phenomenon such as membrane diffusion holds the master clock. The various organelles of the cell show rhythms: not only mitosis but volume changes, shape changes and many another effect, even in unicellular organisms. And yet the nucleus cannot be the sole clock vault either; enucleated cells of algae can photosynthesize rhythmically for a month under constant conditions. Perhaps the mitochondria, the "powerhouse" organelles that house the complex ATP factories of most cells, hold the main clock too. They are absent in bacteria, and so is the circadian rhythm; in fungi they are complex and the rhythms are unusually diverse; in the higher cells the mitochondria are more regular and universal and so is the rhythm. An antibiotic that affects the mitochondria strongly shifts the phase of the cycle.

There are other great cycles. The "virtuoso isopod" *Excirolana,* a small shore scavenger from the coast of California, shows a 24.9-hour free-running period in swimming activity. For two months this rhythm displayed an amplitude modulation parallel to the lunar cycle of high and low tides characteristic of its home waters! It is "capable not only of maintaining tempo, but of repeating entire ornamented phrases of an environmental score" (J. T. Enright). Small signals can convey the time of the outer world—brief flashes of red light, or the clock in a room given an artificial daylight cycle (experiments were done using fast- and slow-running clocks), or the faint sounds of the daily bustle. Humans plainly have an entrained internal clock, which to our discomfort we can disturb by heedlessly flying across time zones. "That period of twenty-four hours... is particularly distinguished in the physical oeconomy of man.... It is, as it were, the unity of our natural chronology." So wrote C. W. Hufeland almost 200 years ago.

THE TOP: UNIVERSAL TOY, ENDURING PASTIME, by D. W. Gould. Clarkson N. Potter, Inc./Publisher ($7.50). Their wood top weighs 12 pounds. Winding the half-inch rope around it is a major operation. The spinner pulls the rope taut with all his might as he winds. Launched like a discus from shoulder height, the top spins for a second on the hard ground. Then the other member of the two-man team scoops the top up swiftly and with a spadelike wood blade deftly transfers it to a metal saucer—where a spin can last 100 minutes! The sport is decorous, amateur, but serious; grown men are the intent players (the transfer player is as important as the spinner) and they enjoy a variety of contests besides that of simple duration. This dynamical sport delights the skillful villagers of the coastal states of Malaysia, north and east of the capital at Kuala Lumpur.

The spinning top as an amusement for children or adults is widely diffused in space and time. The author of this original and well-informed book is an engineer and a learned collector of tops. He has studied the classical and the ethnographic literature, he has consulted patents and travelers' tales and he has worked with the device itself, spinning, measuring, timing, puzzling. Painted Greek vases display many tops, and Gould makes out a good case that one museum holds an antique ceramic yo-yo, itself decorated in vase-painting style. The yo-yo, a recuperative top that both spins and translates, appeared as a craze in Paris in about 1790; the children of the emigration carried it all over Europe as noble families fled Madame Guillotine. The cyclical return of the fad some 20 years ago was surely only one peak of many.

Tops come in a bewildering variety of forms, with many ingenious accessories and embellishments. In 1896 *Scientific American* published a wonderfully crowded engraved plate (apparently cribbed from an article in *La Nature* of the previous year) that displayed 51 different current tops from the hands of the fertile inventors of the 19th century. The Japanese have a similar tradition: seven pages in the book show tops from Japan and demonstrate their specific qualities; one is formed and decorated like an eggplant, for example. The toy there was less associated with children than it was in postclassical Europe; it

belonged rather to the showman, the juggler, the itinerant peddler gathering an amused crowd. The marvelous bamboo provided ready-made cylindrical forms, strong, hard and naturally weighted near the perimeter. Only lathe-turned wood or metal could rival bamboo; Japanese tops were inexpensive and superior. The ironclad top, weighted at the rim by an iron band, is a familiar Japanese device.

Gould divides tops into six classes: the twirler, spun between thumb and fingers or even between palms; the supported cord-spun top; the top cast free of the cord to the ground, called a peg top; the whip top, found worldwide save perhaps in the Pacific islands and widespread already in antiquity, with specimens shown here from the Egyptian Middle Kingdom; the buzzer, a disk with two strings through it, which spins and often hums or buzzes as the twisted cords are alternately tensed and relaxed; finally, the yo-yo.

It is not clear how closely the toy top is related to the other less playful applications of rotation: the fire drill, the spindle whorl, the potter's wheel. The author's plausible opinion is plain. The top as a plaything is found more widely than any other rotational artifact. It is not likely that it spread in many forms from a common source whereas other essential arts did not. Either it was invented many times afresh or it goes back as a cultural holding in some form to the remotest past, diffusing like speech itself. Top styles and popularity wax and wane, but tops spin the spinning world around, now as they always have.

The theory of the spinning top's wonderful defiance of gravity is given here in wholly qualitative form. The role of tabletop friction is correctly brought out but the author's version of the representation of rotation by a vector—without using those words—will not convince every mathematically innocent reader. The inverting twirler, the so-called Tippe Top, gets a fair account here, even a radiograph (although one misses that wonderful photograph of Niels Bohr demonstrating the device to the intent, stooping, corpulent Wolfgang Pauli, the great theorist of electron spin). Gould reports to a candid world his direct measurements of the rotational speed of various tops. It is impressive to learn that the familiar cheap plastic twirler turns, if well spun, at about 3,300 r.p.m., and that a spinning button may get up to 4,000. (Adepts with stroboscopes should report their best performances.)

The engineering top, whether the heavy turbine wheel or the elegant gyro of inertial-guidance devices, is only touched on here. The whirling dervish, the Eskimo hunter spun on an ice cake and the human centrifuge of our amusement parks are all adequately treated. The entertaining and substantial volume is fit celebration of the author's view: "*Play* brought the top into existence." One may guess that the workaday rotator—for hole-making, pots, thread, machinery or bullets—was a spin-off from the older source.

A Handbook of Integer Sequences, by N. J. A. Sloane. Academic Press ($10). Field guides are uncommon in mathematics. Here is one, identifying with care some 2,300 sequences of positive integers, each with an implied—and named—rule for continuation, world without end, unto infinity. A few are interlopers—like a parakeet escapee listed in a bird book, too attractive to omit although alien to the region. (The curious sequence 1, 15, 29, 12, 26, 12... extends through 51 entries: it is the days of the months at fortnightly intervals from January 1, going on into the next year, implicitly arriving at the leap year and various century years with differing rules. One can read it off piecemeal from most calendars.) Many playful and puzzle sequences occur, some even outside the main list. There are bread-and-butter sequences as well. It would have been impossible to omit π, of which 60 decimal digits are given, with a source cited for 100,000. One of the most explosively growing sequences (for which only four entries can be given: 1, 2, 720, 620448401733239439360000) is expressed by the neat formula (N!)!

The arrangement is lexicographical, exactly as in any dictionary except that one orders digits, not letters. And so 1, 2, 0, 0, 1, 1, 2... (a very bland sequence) heads the list, no entry exceeding 4 in two entire lines of the sequence. It is described merely as being "related to the divisors of N." The source is always given, in a compact but precise form; this book is not self-contained but is meant to be used with the resources of a mathematical library. A small library will do: the author points out that "a considerable fraction of the sequences will be found in the following nine great works" and in the runs of four journals, of which one is recreational in emphasis and three are more workaday. The references do not seek to celebrate the discoverer of a sequence but rather to give the source of the most extensive published tables. The last entry of the volume is 1, 744, 196884..., called the "coefficients of a modular invariant." (The author limited himself to sequences that begin 1, *n*, where *n* is between 2 and 999.) Initial 1's and 0's are lightly treated; indeed, an initial 1 has been furnished to every sequence that lacked it, serving as a marker for some subtle need of the automatic typesetter that produced these good-looking pages from magnetic tape.

The sequences arise in many branches of mathematics. Recurrences are frequent members of the list, of course; an introductory section describes the tricks of finding such formulas, given a sequence. The list itself contains such famous examples as the Fibonacci numbers, the Catalan numbers, the Bernoulli numbers, the Bell numbers, all with their methods of generation. The Davenport-Schinzel numbers, however, stand inscrutably: 1, 4, 8, 12, 17, 22, 27, 32; more can be learned about them only by referring to the conference proceedings that are cited! A prolific generator of sequences is the domain of graphs and trees, with their many and often deep kinships with generalized algebra and with natural science. No fewer than 31 listed sequences are chemical in origin, enumerating molecule skeletons that are formed according to some specific rules, whether realized or not. Hydrocarbons and alcohols, boron trees and ammonium compounds are listed here, but what they are in detail remains locked in the references; most of these chemical enumerations come, unexpectedly, from only a few papers in the literature of the 1930's. The more recent results of the physicists' counting of various paths on lattices are also here.

Every reader can aspire to be his or her own Martin Gardner. His kind of combinatoric game is found in plenty: the necklaces, folded strips of stamps, knots, polyominoes, flexagons and switching functions. The theory of numbers provides much grist for the sequence mill: from two lines of ordinary primes to 52 varieties of special primes, other strings sorted by various ways of sieving, more recondite sets and several lists suggested by the Goldbach conjecture. Partitions are rife too, and naturally permutations.

It is not evident how often people wish to identify some integer sequence fleetingly encountered in reading or shyly hidden in their own mathematical woodlands. They surely will seek this book. Many another reader with mathematical bent can profit by it as a starting point into a very wide variety of mathematical literature or merely as a challenge, a fresh scent along the trail of mathematical truth. It does not tell too much; what it does is exhibit a specimen,

list a name and include a pointer to more. Many an amateur will browse in it and most libraries will want it. The author, a mathematician once at Cornell University and now at Bell Laboratories, hopefully says: "It might . . . be useful to have around when the first signals arrive from Betelgeuse." (He gives a few sequences that depend on the English language or on Arabic numerals; they won't help much then.) There is an excellent index, enticing a casual reader with whimsies such as those numbers dubbed polygonal, abundant, perfect, amicable or just plain good.

GRZIMEK'S ANIMAL LIFE ENCYCLOPEDIA, VOLUME 10, MAMMALS I; VOLUME 9, BIRDS III. Bernhard Grzimek, editor-in-chief. Van Nostrand Reinhold Company ($29.95 per volume; complete set $325). Like a visit with the curators at some transcendental zoo, this 13-volume encyclopedia of thick, chunky, handsome volumes offers the reader a lively and well-informed look centered on the most colorful and congenial of the million-plus species of living animal forms. The work is arranged systematically: every big class, such as birds and mammals, gets a brief general essay that concerns itself with the insides of the beasts, their evolution and their development. The great bulk of the work is more zoolike: one sees the outsides only of living animals—no fossil species —and there is only a little attention to biochemical or physiological issues. The books celebrate the life of the animal viewed by devoted, acute observers in the wild or the zoo without instruments or experiments, not in the laboratory or the museum. Family by family the animals are described, the history of man's knowledge of the form is sketched, the range is mapped and the behavior and husbandry of the living animal are outlined, generally by an author who knows and has raised the animal. The compendious set appeared first in German seven years ago and has been rendered into French, Italian, Dutch and now English. So far about half of the set has appeared in America; it is scheduled for completion this year. The nearly 200 contributors are a highly international group of zoologists, chiefly behavior, conservation and zoo people; Grzimek adds more than his share of excellent and personal writing.

Two volumes are chosen here for sampling, the first of the four devoted to mammals and the last of the three given to birds. (The lower animals draw one volume for all the microscopic forms, the sea phyla, the wormy ones and others up

to the spider and the lobster; insects, mollusks and reptiles draw a volume each, and the fishes rate two.) The first mammal volume begins with the egg-layers—the platypus and its kin—and closes with the gorilla. It describes some 600 species of marsupials, insectivores, hedgehogs, lemurs, monkeys and apes. (Chimpanzee and man belong to the next volume, not yet at hand.) These are strange tales well told. Consider the pygmy flying possum of the Australian east coast. The size of a "delicate mouse," these not uncommon feather-tail gliders glide by night from high tree nests to seek insect prey and flower nectar. Or consider the armored shrews of tropical Africa. These fellows look a little like big mice, but their interlocking, faceted and spined vertebrae make them "almost indestructible." A grown man can stand on such a shrew for several minutes with his full weight; released, "the shrew merely shakes itself and runs off unharmed." One gatefold colored painting shows the fauna of Madagascar, as strange as any on this planet. The prosimians and the tenrecs, unlike any other mammals, respond sharply to external temperature. Active in the sunset hours, they are torpid and cool most of the day, their blood heat falling, from a normal value not much below our own, by a chilly 15 degrees Fahrenheit at rest. In a way they hibernate daily; the blood remembers the reptile! The big orang-utan graces the jacket of the volume, full-face and grave; the essay on the species by Barbara Harrisson of Cornell, who befriended these cousins of ours in Borneo, is a special pleasure.

As for the birds, one can report on the marvelous honey-guide family, woodpecker variants with fully parasitic breeding habits that eat the beeswax of the honeycomb. They harbor bacteria that can produce fatty acids from the generally inert wax. They can detect hives by the smell of the wax and they get access to the well-hidden comb by enlisting a partner, man or badger, whom they lead on with a continuous excited call, tree after tree, until the bee tree is reached. Their actions are, of course, "not purposive, nor is there any insight involved." All of this takes two pages out of 650 that include lists, indexes and a four-language dictionary.

These books are a kind of paper zoo. Any library that is zoo-related or wishes to stand as a zoo surrogate needs the set; the individual enthusiast of this or that form or family will find the right volume an authentic and readable source of brief, intimate information about his favorite animal.

INDEX OF ADVERTISERS

APRIL 1974

BIBLIOGRAPHY

Readers interested in further reading on the subjects covered by articles in this issue may find the lists below helpful.

THE DELIVERY OF MEDICAL CARE IN CHINA

PUTTING CHINA'S MEDICINE IN PERSPECTIVE. H. Jack Geiger in *Medical World News,* Vol. 14, No. 20, pages 43–49; May 18, 1972.

THE BAREFOOT DOCTORS OF THE PEOPLE'S REPUBLIC OF CHINA. Victor W. Sidel in *The New England Journal of Medicine,* Vol. 286, No. 24, pages 1292–1300; June 15, 1972.

MEDICINE AND PUBLIC HEALTH IN THE PEOPLE'S REPUBLIC OF CHINA. Edited by Joseph R. Quinn. Department of Health, Education, and Welfare Publication No. (NIH) 73–67. John E. Fogarty International Center, 1973.

SERVE THE PEOPLE: OBSERVATIONS ON MEDICINE IN THE PEOPLE'S REPUBLIC OF CHINA. Victor W. Sidel and Ruth Sidel. Josiah Macy, Jr., Foundation, 1973.

INTEGRATED OPTICS

INTEGRATED OPTICS: AN INTRODUCTION. Stewart E. Miller in *The Bell System Technical Journal,* Vol. 48, No. 7, pages 2059–2069; September, 1969.

LIGHT WAVES IN THIN FILMS AND INTEGRATED OPTICS. P. K. Tien in *Applied Optics,* Vol. 10, No. 11, pages 2395–2413; November, 1971.

INTEGRATED OPTICS AND GUIDED WAVES—A REPORT OF THE TOPICAL MEETING. R. V. Pole, S. E. Miller, J. H. Harris and P. K. Tien in *Applied Optics,* Vol. 11, No. 8, pages 1675–1685; August, 1972.

SWITCHING AND MODULATION OF LIGHT IN MAGNETO-OPTIC WAVEGUIDES OF GARNET FILMS. P. K. Tien, R. J. Martin, R. Wolfe, R. C. Le Craw and S. L. Blank in *Applied Physics Letters,* Vol. 21, No. 8, pages 394–396; October 15, 1972.

THE EMBRYO AS A TRANSPLANT

HUMAN TRANSPLANTATION. Edited by Felix T. Rapaport and Jean Dausset. Grune & Stratton, Inc., 1968.

CHORIOCARCINOMA: THE CLINICAL BIOLOGY OF THE TROPHOBLAST AND ITS TUMORS. K. D. Bagshawe. The Williams & Wilkins Company, 1969.

FRONTIERS OF BIOLOGY, VOL. 20: FOETAL AND NEONATAL IMMUNOLOGY. J. B. Solomon. American Elsevier Publishing Co., Inc., 1971.

THE IMMUNOBIOLOGY OF TRANSPLANTATION. Rupert E. Billingham and Willys K. Silvers. Prentice-Hall, Inc., 1971.

IMMUNOBIOLOGY OF MAMMALIAN REPRODUCTION. Alan E. Beer and R. E. Billingham in *Advances in Immunology,* Vol. 14, pages 1–84; 1971.

WOOD PULP

FUNDAMENTALS OF PAPERMAKING FIBRES. Edited by F. Bolam. Technical Section, British Paper & Board Makers Association, 1957.

THE CHEMISTRY OF WOOD. Edited by Bertie L. Browning. Wiley-Interscience, 1963.

PULPING PROCESSES. Sven A. Rydholm. Wiley-Interscience, 1965.

TEXTBOOK OF WOOD TECHNOLOGY: VOL. I. Alexis J. Panshin and Carl de Zeeuw. McGraw-Hill Publishing Co., 1970.

THE CENTER OF THE GALAXY

COMPOSITION AND ACTIVITY OF THE NUCLEUS OF OUR GALAXY AND COMPARISON WITH M 31. J. H. Oort in *The Nuclei of Galaxies,* edited by D. J. K. O'Connell. American Elsevier Publishing Co., Inc., 1971.

OBSERVATIONS OF NEUTRAL HYDROGEN NEAR THE GALACTIC CENTER, II: THE NUCLEAR DISC. R. H. Sanders and G. T. Wrixon in *Astronomy and Astrophysics,* Vol. 26, No. 3, pages 365–377; August (1), 1973.

INFRARED MAPS OF THE GALACTIC NUCLEUS. G. H. Rieke and F. J. Low in *The Astrophysical Journal,* Vol. 184, No. 2, Part 1, pages 415–425; September 1, 1973.

THE POSSIBLE RELATION OF THE 3-KILOPARSEC ARM TO EXPLOSIONS IN THE GALACTIC NUCLEUS. R. H. Sanders and K. H. Prendergast in *The Astrophysical Journal,* Vol. 188, No. 3, pages 489–500; March 15, 1974.

PLATE TECTONICS AND THE HISTORY OF LIFE IN THE OCEANS

DYNAMICS IN METAZOAN EVOLUTION. R. B. Clark. Oxford University Press, 1964.

GLOBAL TECTONICS AND THE FOSSIL RECORD. James W. Valentine and Eldridge M. Moores in *The Journal of Geology,* Vol. 80, No. 2, pages 167–184; March, 1972.

EVOLUTIONARY PALEOECOLOGY OF THE MARINE BIOSPHERE. J. W. Valentine. Prentice-Hall, Inc., 1973.

A REVOLUTION IN THE EARTH SCIENCES: FROM CONTINENTAL DRIFT TO PLATE TECTONICS. A. A. Hallam. Oxford University Press, 1973.

THE PERCEPTION OF TRANSPARENCY

GESETZE DES SEHENS. Wolfgang Metzger. Verlag von Waldemar Kramer Frankfurt am Main, 1953.

MARGINI QUASI-PERCETTIVI IN CAMPI CON STIMOLAZIONE OMOGENEA. Gaetano Kanizsa in *Rivista di Psicologia,* Vol. 49, No. 1, pages 7–30; January–March, 1955.

ZUR ANALYSE DER PHÄNOMENALEN DURCHSICHTIG KEITSERSCHEINUNGEN. F. Metelli in *Gestalt und Wirklichkeit.* Duncker und Humboldt, Berlin, 1967.

AN ALGEBRAIC DEVELOPMENT OF THE THEORY OF PERCEPTUAL TRANSPARENCY. F. Metelli in *Ergonomics,* Vol. 13, 1970.

THE BIOLOGICAL CONTROL OF DUNG

COULD DUNG EATING INSECTS IMPROVE OUR PASTURES? G. F. Bornemissza in *Journal of the Australian Institute of Agricultural Science,* Vol. 26, No. 1, pages 54–56; March, 1960.

THE NATURAL HISTORY OF DUNG BEETLES OF THE SUBFAMILY SCARABAEINAE (COLEOPTERA, SCARABAEIDAE). G. Halffter and E. G. Matthews in *Folia Entomologica Mexicana,* Vol. 12; 1966.

BIOLOGICAL CONTROL OF PESTS AND WEEDS. D. F. Waterhouse and Frank Wilson in *Science Journal,* Vol. 4, No. 12, pages 31–37; December, 1968.

A NEW TYPE OF BROOD CARE OBSERVED IN THE DUNG BEETLE *ONITICELLUS CINCTUS* (SCARABAEIDAE). G. F. Bornemissza in *Pedobiologia,* Vol. 9, pages 223–225; 1969.

MYCETOPHAGOUS BREEDING IN THE AUSTRALIAN DUNG BEETLE, *ONTHOPHAGUS DUNNINGI.* G. F. Bornemissza in *Pedobiologia,* Vol. 11, pages 133–142; 1971.

THE AMATEUR SCIENTIST

FLOATING DROPS AND LIQUID BOULES—A FURTHER LOOK. Kenneth Hickman, Jer Ru Maa, Andrew Davidhazy and Olivia Mady in *Industrial & Engineering Chemistry,* Vol. 59, No. 10, pages 18–41; October, 1967.

SMIRNOFF SILVER® VODKA. 90.4 PROOF. DIST. FROM GRAIN. STE. PIERRE SMIRNOFF FLS., DIV. OF HEUBLEIN. ©1973, HEUBLEIN, INC., HARTFORD, CONN.

The Silver Martini.
For people who want a silver lining without the cloud.

Smirnoff Silver. Ninety point four proof. Smirnoff leaves you breathless®

Test drive a Camaro with your own two eyes.

Run your eyes over a new Camaro.

And it's easy to imagine yourself out on the road driving one.

The shape itself seems meant for motion. Smooth sides. Front and back flowing into new improved bumper systems.

There's the solid grip of widespread wheels, with steel-belted radials available. A feeling of closeness to the ground. Of hugging the road.

Under that long hood available engines range from an economical Six to the responsive Z28 engine package. The standard V8 now measures 350 cu. in. Power steering is standard on all V8 models.

Test drive a Camaro with your own two eyes and, then, with your own two hands.

Basic Sport Coupe, Luxury-Touring Type LT, or Z28 performance version—it'll be one of the least surprising things you'll do this week.

Because the way Camaro looks is the way Camaro goes.

Chevrolet

CHEVROLET MAKES SENSE FOR AMERICA.