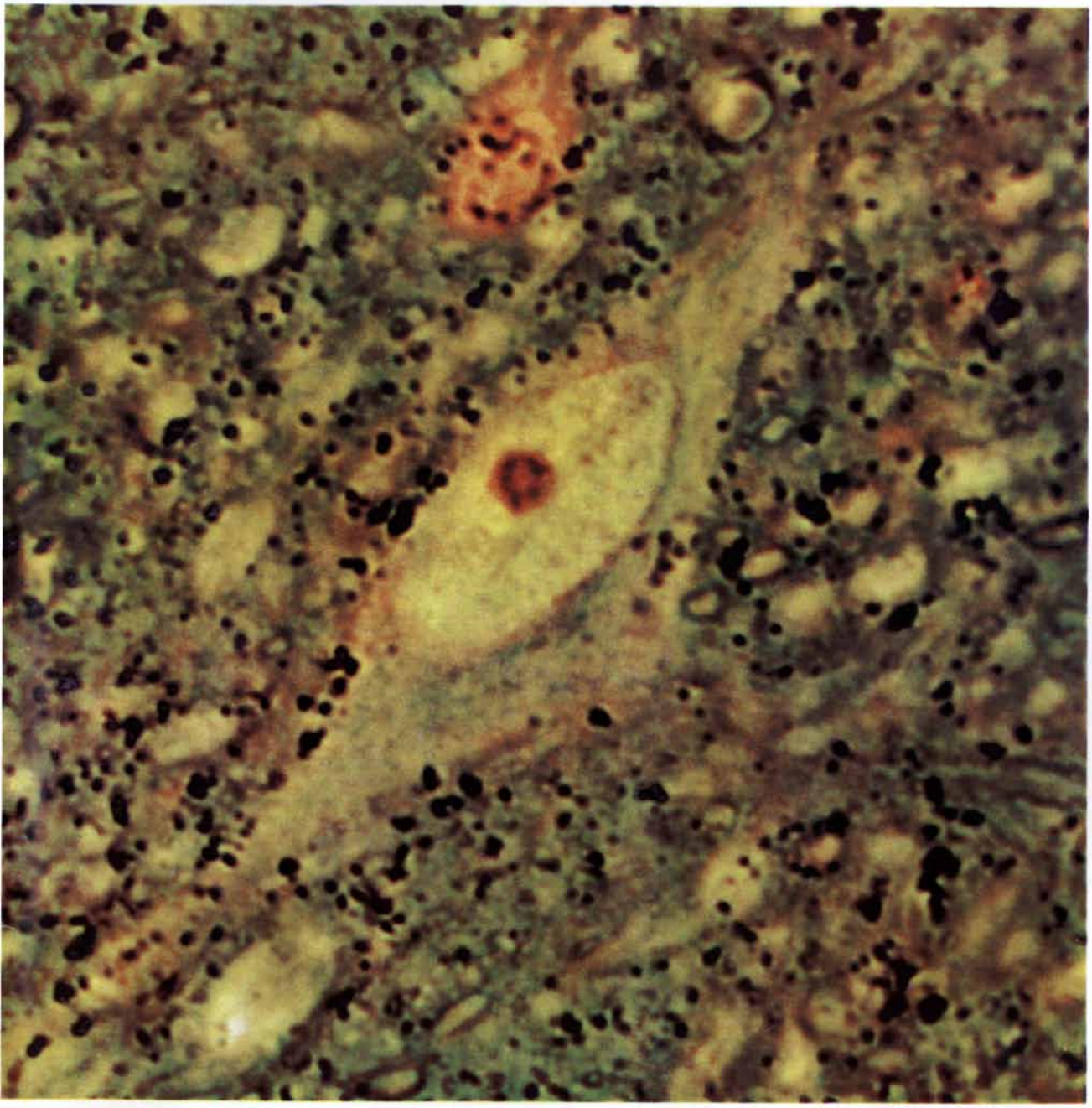
SCIENTIFIC AMERICAN



PATHWAYS IN THE BRAIN

ONE DOLLAR

July 1971

We have a way to find out what's really going on in the other 3/4 of the world.

Why is the weather a favorite topic of conversation? What else can we do besides talk about it?

One thing we know for sure is where most of our weather comes from: the seas around us.

It is too costly to put weathermen in the world's oceans. That was the problem: to find a less expensive way to put information-gathering devices in the oceans where they would provide data for more accurate weather predictions.

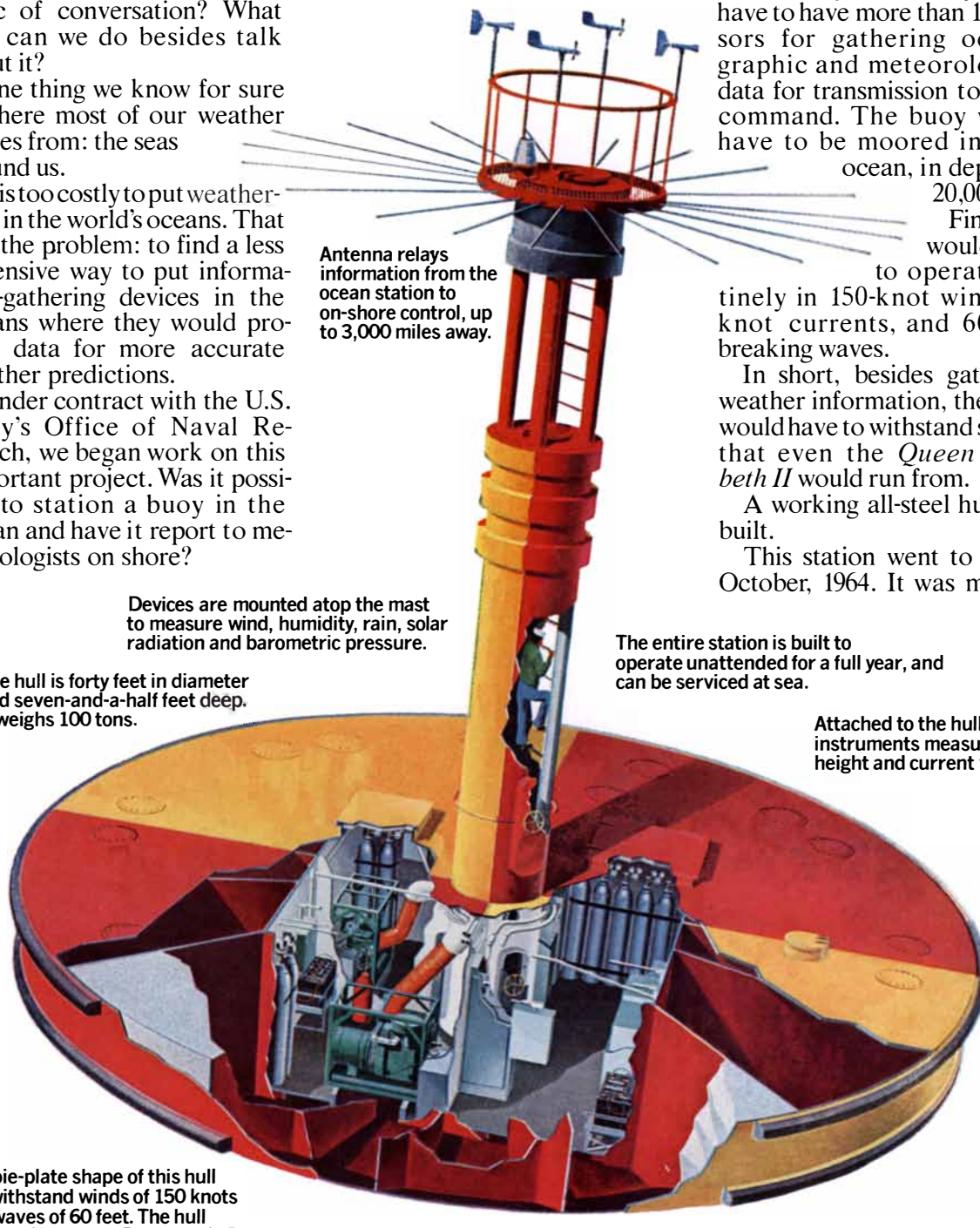
Under contract with the U.S. Navy's Office of Naval Research, we began work on this important project. Was it possible to station a buoy in the ocean and have it report to meteorologists on shore?

To do its job, the buoy would have to have more than 100 sensors for gathering oceanographic and meteorological data for transmission to shore command. The buoy would have to be moored in deep ocean, in depths to 20,000 feet. Finally, it would have to operate routinely in 150-knot winds, 10-knot currents, and 60-foot breaking waves.

In short, besides gathering weather information, the buoy would have to withstand storms that even the *Queen Elizabeth II* would run from.

A working all-steel hull was built.

This station went to sea in October, 1964. It was moored



Antenna relays information from the ocean station to on-shore control, up to 3,000 miles away.

Devices are mounted atop the mast to measure wind, humidity, rain, solar radiation and barometric pressure.

The hull is forty feet in diameter and seven-and-a-half feet deep. It weighs 100 tons.

The entire station is built to operate unattended for a full year, and can be serviced at sea.

Attached to the hull, other instruments measure wave height and current velocity.

The pie-plate shape of this hull will withstand winds of 150 knots and waves of 60 feet. The hull rode out Hurricane Betsy in 1965.

Inside, electronic equipment gathers and stores information from 100 different sensors for transmission to shore.

Below the surface, more devices are placed on the mooring lines to measure water temperature and salinity.

in the middle of the Gulf Stream, off the Florida Coast, in a hurricane lane.

The station carried environmental sensors to see what a hurricane would do to the buoy.

In calm seas, the buoy worked to our fullest expectations. The most physical test of the system was to come.

In September, 1965, the eye of Hurricane Betsy passed within fifty miles of the station.

Waves were 45 feet high. Wind speed averaged 80 miles per hour. Gusts as high as 110 miles an hour.

Through this punishing storm the buoy functioned routinely.

Today, ocean data stations are operated off our east coast by the Commerce Department's National Oceanic and Atmospheric Administration. Next year, additional stations will be in service in the Gulf of Mexico to monitor the hurricane season. In the future, a network of these stations could form a major part of a worldwide weather watch.

Prototype weather station off Florida.

This is the kind of data ocean stations send back:

Meteorological data on wind speed and direction, barometric pressure, air temperature, relative humidity, precipitation and solar radiation.

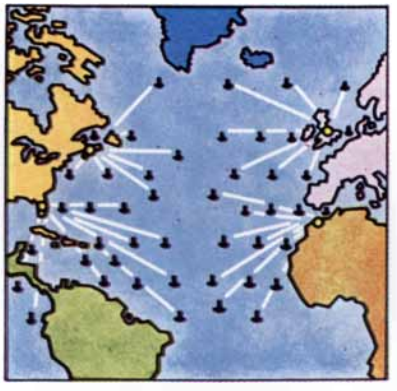
Oceanographic data on ocean current, direction and speed, water temperature, salinity and surface wave profile.

By knowing the exact water temperature a thousand miles away, we'll know better whether a storm will bring snow or rain when it reaches land.

One way or another, this information will benefit everyone. For instance, by helping tell when to plant and harvest, schedule a vacation, even locate fish.

Because of the buoy's capabilities, the U.S. Coast Guard decided to use it as an unmanned navigation aid. So far, our Electro Dynamic Division has outfitted seven of them and the Coast Guard is using them to replace lightships, at a great reduction in operating costs.

Navigation buoys are on station now off the entrances of New York and San Francisco harbors, in Delaware Bay, and off the English Coast, among other places.



Potential ocean station network.

Off our busy harbors, these navigation buoys could form networks of traffic sensors reporting to onshore stations and helping control ships in and out of ports.

Because we could put so many technologies together, we could put together the weather and navigation stations.

They're typical of something else at General Dynamics. Our people develop new technologies for specific needs. Then someone else in the company finds other uses for those technologies.

It's happening in all sectors of our business, in defense and in our growing commercial markets. It happens in marine systems. It also happens in aerospace, shipbuilding, telephone systems, electronics and natural resources.

It explains why we're a company that keeps doing things no one's been able to do before.

Navigation buoy off the English Coast.

GENERAL DYNAMICS

Crystal-clear bottle with more class than glass.

The Marlex® Marketing Team created it.

The packaging industry needed it. A bottle that feels soft and satiny, has sparkling clarity plus the unbreakability, light weight and economy of plastic.

The Marlex Marketing Team made it happen. The first clear polypropylene bottle for commercial use.

Using Marlex polypropylene resins, the team developed our patented Orbet process—a sophisticated fabricating technique for producing crystal-clear bottles with all the best features of plastic. They can be used for practically everything from dishwashing liquids and detergents to syrup and juices.

But it took years of research and development by our Marlex Marketing Team. And years of working with packaging experts. Testing for clarity, impact strength, moisture vapor transmission, environmental stress cracking.

The point is: whatever your product, if you have any reason to work with plastics, you have every reason to work with Phillips, inventors of genuine Marlex polyethylene and polypropylene resins. The resins that are worth more to you.

Get help from the Marlex Marketing Team. Call:

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

PHILLIPS 66

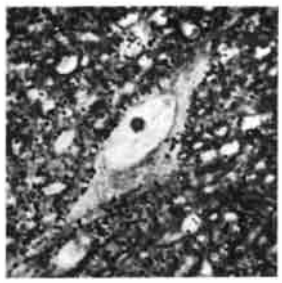
At Phillips 66 it's performance that counts

Established 1845 **SCIENTIFIC AMERICAN** July 1971 Volume 225 Number 1

ARTICLES

DEPARTMENTS

PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $10 PER YEAR. EUROPEAN SUBSCRIPTION COPIES DISTRIBUTED C/O INTERNATIONAL DISTRIBUTORS, VLAGSTRAAT 128, BORGERHOUT 2200, BELGIUM.



THE COVER

The photomicrograph on the cover enlarges a thin section of the brain of a rat 1,800 diameters. The section was embedded in plastic and stained by a method that identifies the fibers of cells that have degenerated because their cell bodies were destroyed by an experimental lesion. In this case the large black silver particles indicate degenerating axons and boutons (terminals of axons) of cells in the olfactory system. Nerve cells and their processes are yellow or yellow-green; the dark red spot is a nucleolus. Through study of such micrographs and then of electron micrographs of the same or similar sections it is possible to trace the routes by which nerve cells in the brain communicate with one another (see "Pathways in the Brain," page 48).

THE ILLUSTRATIONS

Cover photograph by Lennart Heimer, Massachusetts Institute of Technology

Page	Source
18–25	Graphic Presentation Services, Inc.
26–31	Enid Kotschnig
33	Bell Telephone Laboratories
34–40	Gabor Kiss
49–50	Lennart Heimer, Massachusetts Institute of Technology
51	Bunji Tagawa
52	Walle J. H. Nauta, Massachusetts Institute of Technology (*top*); Lennart Heimer, Massachusetts Institute of Technology (*bottom*)
53	Lennart Heimer, Massachusetts Institute of Technology (*top*); Bunji Tagawa (*bottom*)
54	Bunji Tagawa
55	Lennart Heimer, Massachusetts Institute of Technology
56	Lennart Heimer, Massachusetts Institute of Technology (*top and bottom right*); Bunji Tagawa (*bottom left*)
57, 60	Lennart Heimer, Massachusetts Institute of Technology
65	Simone-Antoinette Deyts, Archaeological Museum of Dijon
66–67	Eric Mose
68–73	Simone-Antoinette Deyts, Archaeological Museum of Dijon
74–78	Hale Observatories
79–80	Tom Prentiss
81–82	Hale Observatories
83–85	Tom Prentiss
86–87	Donald Paterson, courtesy Photo Researchers, Inc.
88–93	Jim Egleson
95	Alan D. Iselin
96	Stanford Linear Accelerator Center
97–98	Alan D. Iselin
99	Stanford Linear Accelerator Center
100–104	Alan D. Iselin
106	Jerome Kuhl
107–108	Robin Ingle
110–113	Roger Hayward

People don't buy the MGB just because it's different. But because they are.

They know the MGB's got what it takes to separate the real thing from its imitators.

They know, for instance, that the MGB's 1798 c.c. twin-carb engine is proven time and again in competition. And yet, it turns in the kind of performance that delivers up to 25 miles per gallon.

Other race-proven performance features include the full-synchromesh 4-speed gearbox, the stopping power of 10.75-inch front disc brakes and the roadability of racing-type suspension with radial-ply tires.

Then, too, you'll appreciate the kind of MG beauty that goes with MG performance. With its clean lines, mag-style wheels and recessed grille, the MGB has the kind of look that sets you apart as a winner. Test drive the MGB '71, the unique sports car for people who are a breed apart.

For the name of your nearest Austin-MG dealer and information about overseas delivery, dial (800) 631-1971 except in New Jersey where the number is (800) 962-2803. Calls are toll-free, of course.

MG AUSTIN

BRITISH LEYLAND

British Leyland Motors Inc., Leonia, New Jersey 07605.

LETTERS

Sirs:

"Physiological Tremor," by Olof Lippold [SCIENTIFIC AMERICAN, March], is a fine presentation of many aspects of muscle control. However, the notion that the brain's alpha rhythm may be generated by the extraocular muscles (voluntary muscles that control movements of the eyeball) needs to be corrected. It is natural that such a keen interest in muscle activity could lead to the aggrandizement of the role of muscles in neurophysiology. Data cited recently ("Alpha and Kappa Electroencephalogram Activity in Eyeless Subjects." Robert M. Chapman, C. R. Cavonius and J. Terry Ernest in *Science*, Vol. 171, No. 3976, pages 1159–1161; March 19, 1971) eliminate tremor of the extraocular muscles as a necessary source of alpha electroencephalogram activity. A subject with both eyes removed showed alpha activity, and marked left-right differences in alpha activity were not found in one-eyed subjects whose eyes and extraocular muscles were removed on one side.

Although alpha brain activity and the muscle action studied by Lippold are predominantly 10 cycles per second, the physiological tremor of the eye has been reported to be primarily between 30 and 70 cycles per second.

ROBERT M. CHAPMAN

Eye Research Foundation
Bethesda, Md.

Sirs:

I am sorry to have worried Dr. Chapman by my mild statement that "the alpha waves may be generated by the extraocular muscles." I was referring to the theory that alpha rhythm does not originate in cortical gray matter but is due to an interaction between the corneoretinal potential (a standing potential across the eye) and tremor in the extraocular muscles (voluntary muscles that move the eye).

This theory was suggested by the similarity between physiological-tremor records from any muscle in the body and human electroencephalograms obtained when the eyes are closed. They have a similar frequency spectrum; they behave similarly in response to various stimuli; both are absent in sleep and during anesthesia; in the child, tremor and alpha waves both have a frequency of around five hertz (cycles per second), increasing to 10 hertz in adult life.

Four main lines of experimental evidence support the new theory:

1. Tremor (at 10 hertz, having a frequency spectrum similar to alpha waves) can be recorded from the eye when alpha waves are present in records from the occipital regions of the scalp. This tremor is mainly an in-and-out translational movement of the eye and not a rotatory one; cross-covariance functions show the tremor to be significantly correlated with alpha waves.

2. The properties of the servo loop controlling eye muscle contraction can be altered. Injecting a local anesthetic into both eyes of an experimental dog abolishes the major part of the surface-recorded electroencephalographic rhythms (presumably the equivalent of alpha waves in humans). Cooling the orbits in humans slows the alpha rhythm by several hertz, just as finger tremor is slowed by cooling the arm. Fatigue of extraocular muscles, induced by making forced, rapid eye movements for some time, increases the amount of alpha rhythm that can be recorded. Fatigue increases ordinary muscle tremor.

3. Single or repetitive mechanical displacements applied to the eyes can either produce alpha waves or synchronize preexisting alpha rhythm. This phenomenon is just like the effects of "prodding" the finger described in my article.

4. The corneoretinal potential can be identified as the "source" of alpha waves. It is a potential difference of about 80 millivolts across the pigment layer of the retina, and it can be altered experimentally in magnitude by changing the level of illumination of the eye. When the potential is, say, increased in the left eye and decreased in the right, it is found that the amplitude of alpha waves is larger on the left side of the scalp and smaller on the right. This must imply that nervous mechanisms are not responsible for alpha rhythm because the optic nerve from each eye is distributed to both visual cortexes.

The purely "observational" approach to this question is useless. One cannot say: "I once had a patient with both eyes removed who had alpha waves, therefore the theory is wrong." If Dr. Chapman examines his eyeless patients carefully, he will find that they still have a potential difference across the empty sockets that behaves just like the normal corneoretinal potential, although it is usually about 10 to 20 percent of the normal magnitude (and the alpha

rhythm to be found in eyeless people is usually 10 to 20 percent of the usual amplitude). He will also find that the inside of the eye socket is in continuous oscillatory movement, roughly in synchrony with the alpha waves present in the scalp record.

O. C. J. LIPPOLD

University College London
London

Sirs:

I fear that the speculations of Thaddeus M. Cowan in the letters department of *Scientific American* [December, 1970] may mislead your readers. The answer as to why the light-sensitive rods and cones of the retina point away from the source of light in the eye almost certainly lies in the development of the eye.

The retinal portions of the vertebrate eye arise first as a hollow, ball-like structure whose inner cavity is continuous with the cavity of the brain. At earlier times in development the cavity of the brain is continuous with the spaces outside the embryo. Hence the outermost cells of the embryo (called ectoderm), the cells that line the cavities of the brain and the cells that line the cavity of the eye can be thought of as portions of a single continuous tissue layer. It is well known that surface ectodermal cells often possess cilia and that cells lining the cavity of the brain and spinal cord also possess those hairlike structures.

It seems reasonable to suppose that, long ago during the evolution of the animals that were giving rise to the vertebrates, the cilia already present on the inner side of the future eye became modified to act as more efficient light receptors. Hence the outer processes of the rods and cones came to be formed during the development of those organisms instead of the more primitive cilia. Thus the mutations and forces of natural selections that led to the origin of the vertebrate visual system probably operated by modifying preexisting cilia and not by creating new structures that point away from the light for no apparent reason.

NORMAN K. WESSELLS

Stanford University
Stanford, Calif.

Sirs:

In the May issue of *Scientific American* ["Letters"] Joel R. Cohen correctly points out that the Hebrew word for king is *melech*, not *melek*. He adds that the guttural sound *ch* bears no relation to the letter *k*, "which has a different symbol in the Hebrew alphabet." This latter statement is misleading. The last letter of *melech* is a terminal *kaf*, and the normal form of this symbol is pronounced *ch* or *k* depending on a simple *dagesh* (dot). Cohen's wording suggests that the letter or sound *k* is represented in Hebrew by the "different symbol" *kof* (or *qof*), which is not always the case.

Now, if we are to reject the cabalistically impressive *melech* because of its guttural sound, the same objection clearly applies to *cholam* (dream), particularly since this word starts with a different type of *ch*, namely the symbol *chet*, which is never pronounced *k*, unlike *kaf*.

Let's face it, cabalists! If we eliminate *kelimah* (shame) because it has four letters (KLMH), there is simply no good solution for KLM other than the obvious Koninklijke Luchtvaart Maatschappij (Dutch for Royal Dutch Airlines).

EMILE C. VAN REMOORTERE, M.D.

University of Nevada
Reno, Nev.

Scientific American, July, 1971; Vol. 225, No. 1. Published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Gerard Piel, president; Dennis Flanagan, vice-president; Donald H. Miller, Jr., vice-president and 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 Allan Wittman, Advertising Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017.

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

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

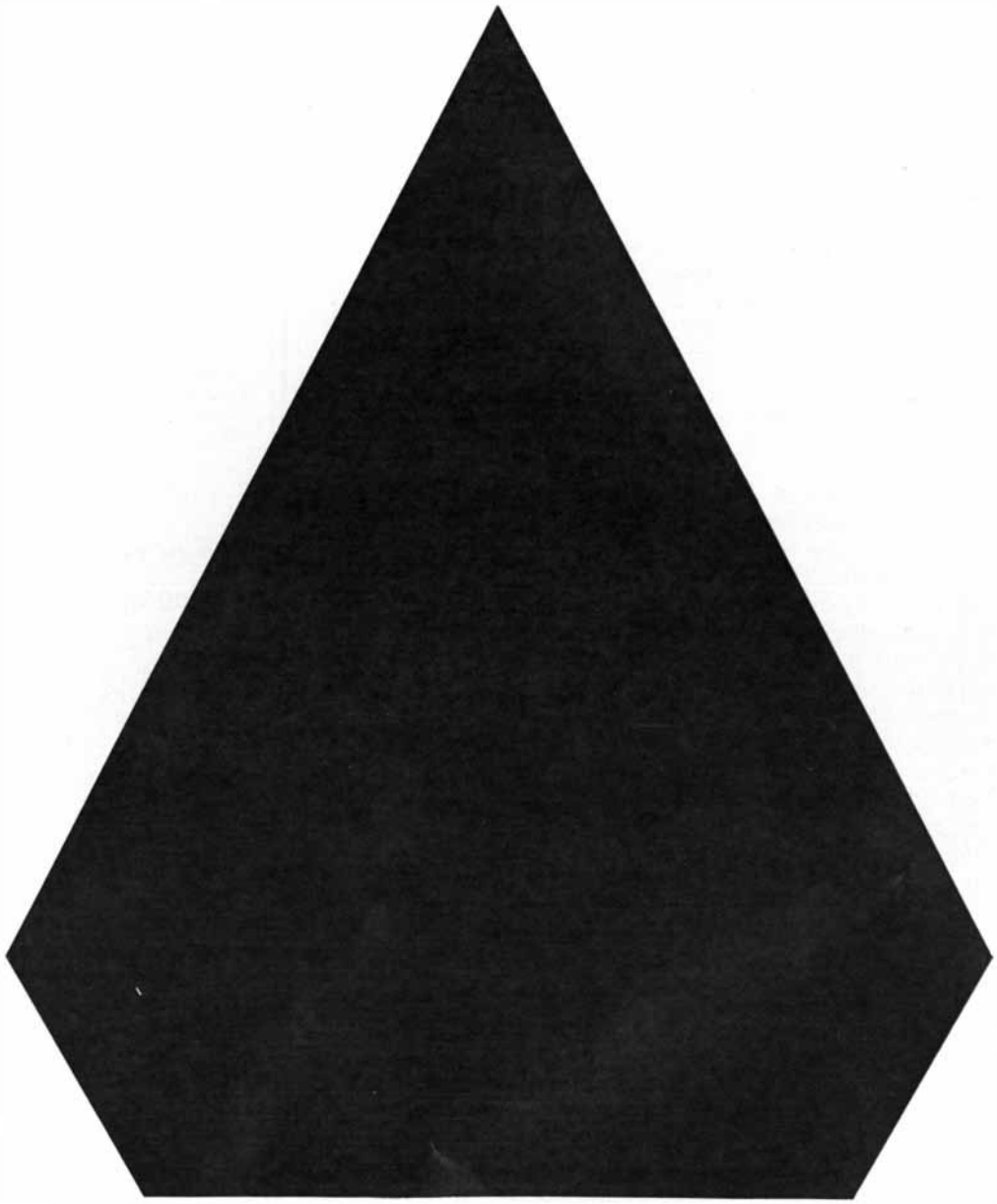
Please enter change of address ☐
Please enter subscription ☐
☐ 1 year, $10 ☐ 2 years, $18 ☐ 3 years, $25

NAME

NEW ADDRESS

OLD ADDRESS

This is how most oil companies see things.



This is how we see it.

Black is the absence of vision . . . light is our thinker's tool.

It's the light he carries in his mind that lets him look at what he's seen before . . . and see something new.

Something no one ever thought to see before.

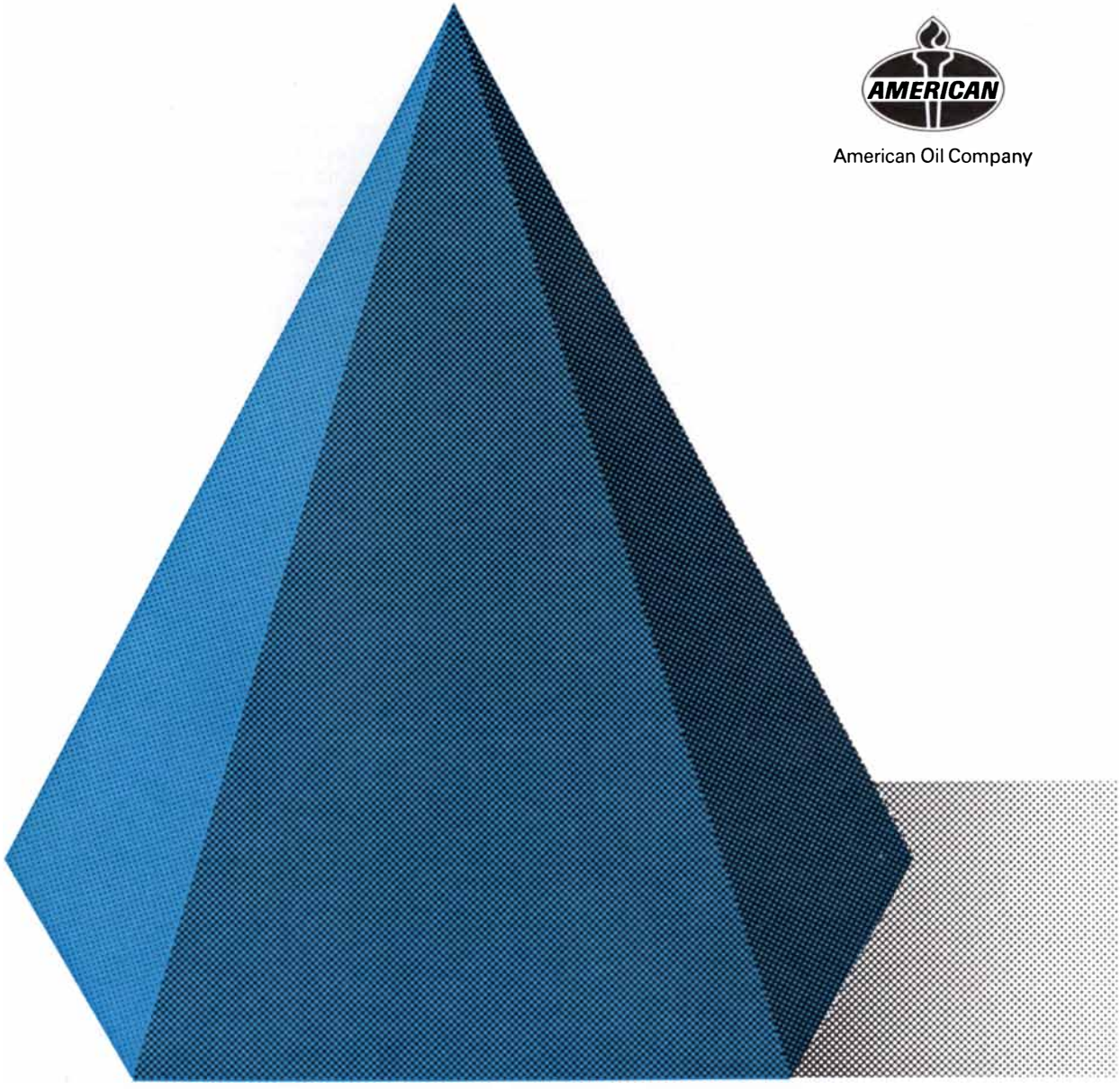
An American Oil thinker can look at something . . . like an industrial oil, and then see how to make it better.

That's how you get more oils for industry that are newer and better.

The way we see it . . . that's why we're important to you.

AMERICAN

American Oil Company



50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

JULY, 1921: "Seldom, if ever, in history has one man been presented with so sublime an opportunity for doing a work both great and good as that which confronts the President of the United States at the present momentous era in the history of the world. We refer, of course, to the practically universal call that has gone up from a war-weary and war-hating world for our President to convene an international conference for the reduction of naval armaments and the prevention of future naval rivalry. It took a world war to drive into men's heads the utter stupidity of settling debated questions by war. Whatever the war has left behind, it has certainly rooted that conviction in men's minds. The world has turned to us as the nation that, by virtue of its wealth in men and natural resources, is best able to carry on war and has frankly asked us to take the initiative for peace. The people of this country have risen to that cry and have answered that they are for peace and for a reasonable reduction of armaments. More than that, both houses of Congress have voiced this call of the people in a joint resolution in which the President is asked to call a conference of the three great naval powers to discuss the question of reducing naval armaments."

"It is rather startling when men of unquestioned scientific standing tell us that all of the tissues of the body are essentially immortal and that, barring accidents, we ought never to die. This is the newest evidence the science of medicine has to offer, and it is evidence, mind you, not theory. A skillful surgeon has been able to keep alive by artificial means, outside the animal, a bit of tissue for a longer time than the natural span of life of the animal itself. The remarkable thing is that the tissue is no longer subject to the influence of time and there is no doubt that if properly cared for it will live on indefinitely. The surgeon is Dr. Alexis Carrel of the Rockefeller Institute in New York and his experiment is with the heart of an embryo chick, which he has kept alive for more than eight years."

"It is interesting to note that it is now possible to construct scales for the measurement of length directly from the fundamental wave lengths of light without the use of any intermediary standard. As an example of work of this kind that is now being regularly carried out, it may be mentioned that the Bureau of Standards recently completed the rulings on a six-inch scale using light waves from a tube containing neon as the length standard."

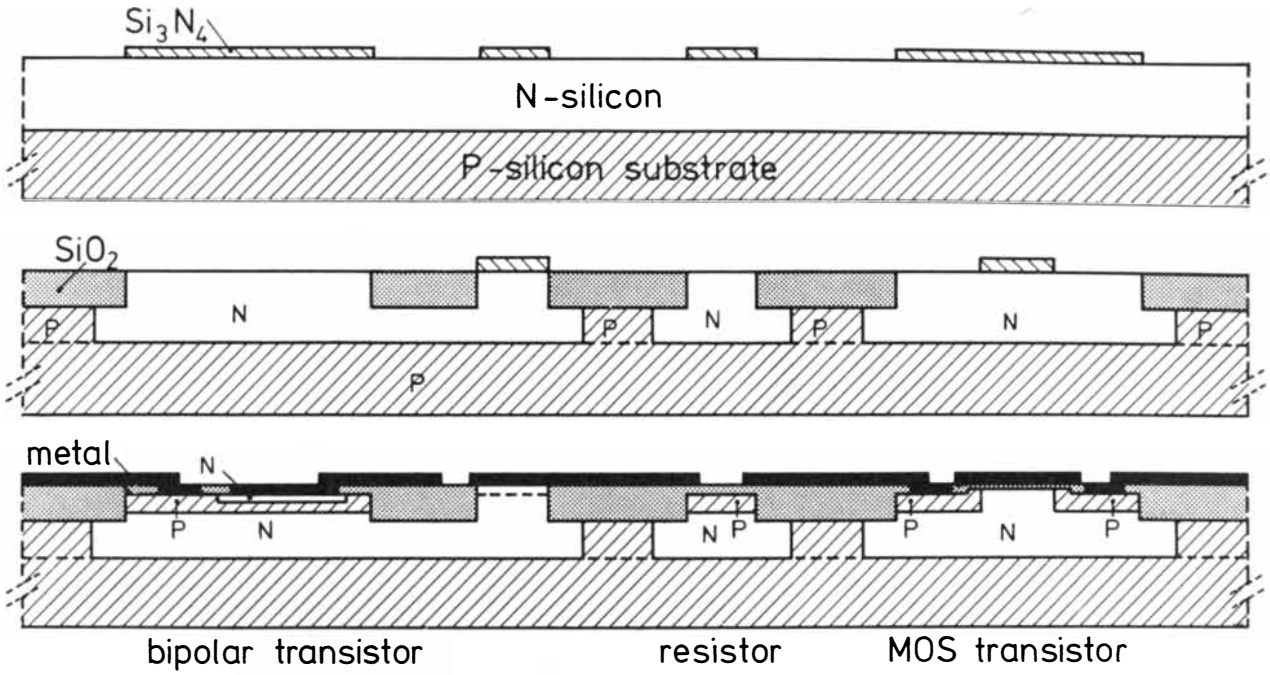
SCIENTIFIC AMERICAN

JULY, 1871: "Ten years have elapsed since the publication of the first paper on the wonderful discoveries of Bunsen and Kirchhoff in Poggendorff's *Annalen*, and since that time every year—we might say almost every month—has witnessed some new and unexpected application of the spectroscope to astronomical research. Having in a measure determined the nature of the atmosphere of the sun, we are startled by the announcement that the spectroscope can be used to measure the velocity of approach or recession in luminous bodies moving with great rapidity. We can detect and measure the movement of the stars, commonly called fixed, by this delicate instrument. It has been observed that if a luminous object approaches with great velocity, the position of the image of the slit will be removed toward the violet end of the spectrum; if the body recedes, the change of position is toward the red. If, therefore, we can establish a ratio between the rate of movement of an object and the amount of displacement observed in the image of the slit, we shall be able to measure or estimate the velocity of any moving body."

"It is estimated that there are 225 colleges for males in the United States entitled to confer degrees and that the total attendance upon them of undergraduate students is, in round numbers, 14,000. This will allow one student to every 2,850 of the population. It is said that the total number of students attending colleges in the United States is both relatively and absolutely less than it was 30 years ago. In other words, although the population of the country has greatly increased, the attendance upon colleges has either remained stationary or has actually diminished. While at one time in our history there was one person in every 1,000 in college, we now find only one person in nearly every 3,000. Our colleges would do well to consider whether the decrease in attendance may not be partly due to defects in their curriculum of instruction. Perhaps the colleges do not offer the kind of education that the times demand. It may be that a little less Latin and Greek and more of the physical sciences would be acceptable."

"The following illustration of the vibratory movement of matter, says Professor Henry, is attested by Professor Horsford. The top of the high tower that constitutes the Bunker Hill monument inclines toward the west in the morning, toward the north at mid-day and toward the east in the afternoon. These movements are due to the expanding influence of the sun as it warms, in succession, the different sides of the structure."

"Prof. Huxley, in responding to the customary compliments to science at the Royal Academy of the Arts dinner, said: 'I will be generous and acquaint you with a fact not generally known, to wit, that the recent progress of biological speculation leads to the conclusion that the scale of being must be thus stated: minerals, plants, animals, men who cannot draw, artists. Thence I conclude, Sir, that you, as President of the Academy, are the crown and summit of creation. My statement, however complimentary, may be a little startling, and you will therefore, I hope, permit me to state the grounds on which it takes rank as a scientific truth. We have long been seeking, as you may be aware, for a distinction between men and animals. The old barriers have long broken away. Other things walk on two legs and have no feathers, caterpillars make themselves clothes, kangaroos have pockets. If I am not to believe that my dog reasons, loves and hates, how am I to be sure that my neighbor does? Parrots talk what deserves the name of sense as much as a great deal that it would be rude to call nonsense. Again, beavers and ants engineer as well as the members of the noblest of professions. But, as a friend of mine discovered a few years ago, man alone can draw, or make unto himself a likeness. This then is the great distinction of humanity, and it follows that the most pre-eminently human of creatures are those who possess this distinction in the highest degree.' "



LOCOS - The new layer cake mix

Although our new layer cake is inedible, we've got a feeling that it is going to make some mouths water! In fact, the cross-sections above show stages in the construction of three different devices: a bipolar transistor, a resistor and a MOS transistor on one silicon slice. The resulting structure looks extremely complicated, yet it can be made with a relatively simple new technique which we call LOCOS: local oxidation of silicon.

LOCOS is the brainchild of Dr. E. Kooi and his support team at Philips Research Laboratories, Eindhoven, the Netherlands. A major advance in semiconductor technology, it enables designers to make new device structures and generally improves the performance and packing density of integrated circuits.

Basically, LOCOS involves the use of a silicon nitride mask to make well defined patterns of thick oxide on the silicon slice. When required, the oxidation stage is combined with a diffusion stage. The top figure shows the nitride mask on a thin N-type silicon layer which is grown epitaxially on the P-type silicon substrate. In the centre figure, which shows an intermediate step, the silicon oxide layer is embedded in the N-silicon, on top of P^+ diffused areas. N-type islands are thus formed, separated by the oxide walls. The nitride is then removed from those areas where P-diffusion is carried out. This is followed by oxidation and N-diffusion steps. During these, the N-type surface area of the MOS transistor continues to be protected by the nitride. This area still needs a thin oxide layer, which is made after the remaining nitride has been removed. Then contact windows are defined and an interconnecting metallized pattern is applied. The bottom figure shows the devices in their final form.

The number of likely applications for LOCOS is at least as great as that for the well-known planar technology. The main difference between the LOCOS and the planar techniques is that in the former, the thick oxide pattern is defined by a thin nitride mask and in the latter, the silicon is first completely covered by a layer of oxide in which windows are subsequently etched to give the desired pattern. As a result, LOCOS allows the use of much thicker oxide films whilst still giving an excellent window definition. In addition, the oxide is embedded in the silicon, so that the overall surface is relatively flat, permitting more efficient control of metallization.

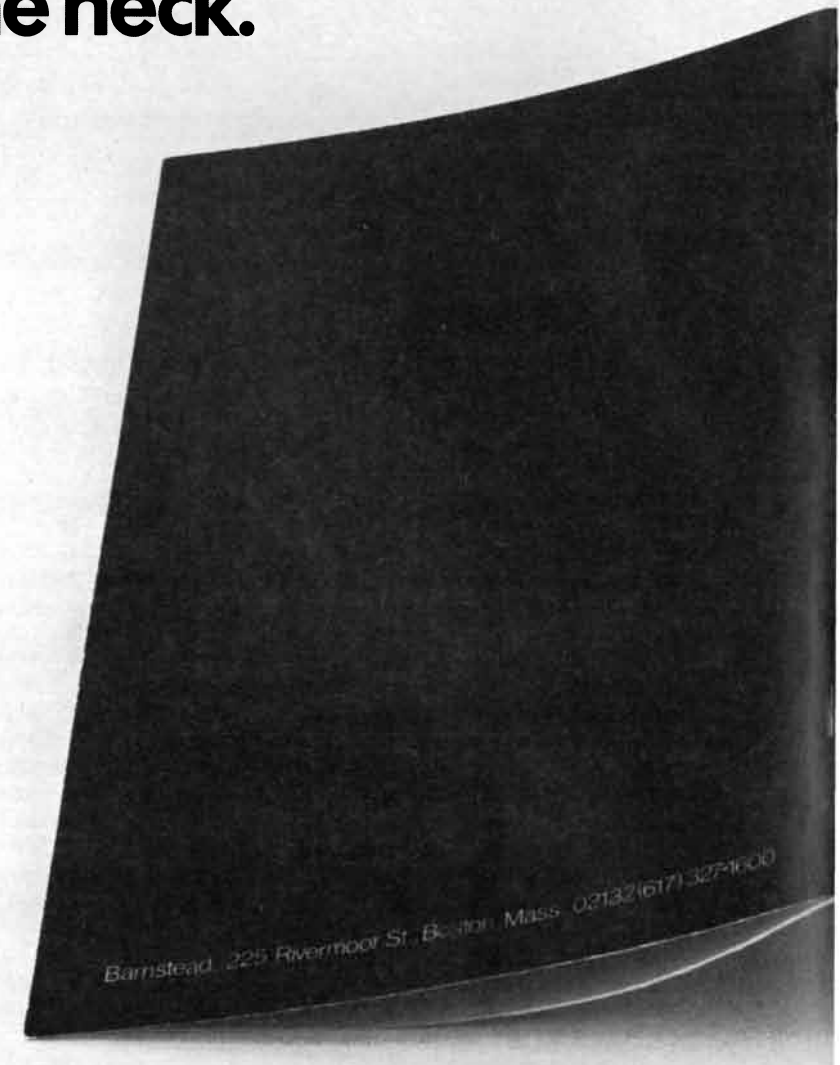
The example shows that LOCOS may be applied to both bipolar and MOS devices. Bipolar integrated circuits can have a very high packing density because the oxide walls are used for isolation. In MOS circuits the reduction in parasitic effects leads to improved reliability.

In all cases, the thick oxide pattern brings about a reduction in capacitances and thus gives better device performances at high frequencies. Generally speaking, we have found that the LOCOS technique offers valuable additions to the existing technology.

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

PHILIPS

We put this little book together because pyrogens can be a lot more than a pain in the neck.



We paid Grace D. Brown of totally independent G.B. Laboratories, Inc., Brookline, Mass., to do it.

To put down everything anybody knows about pyrogens.

What they are.

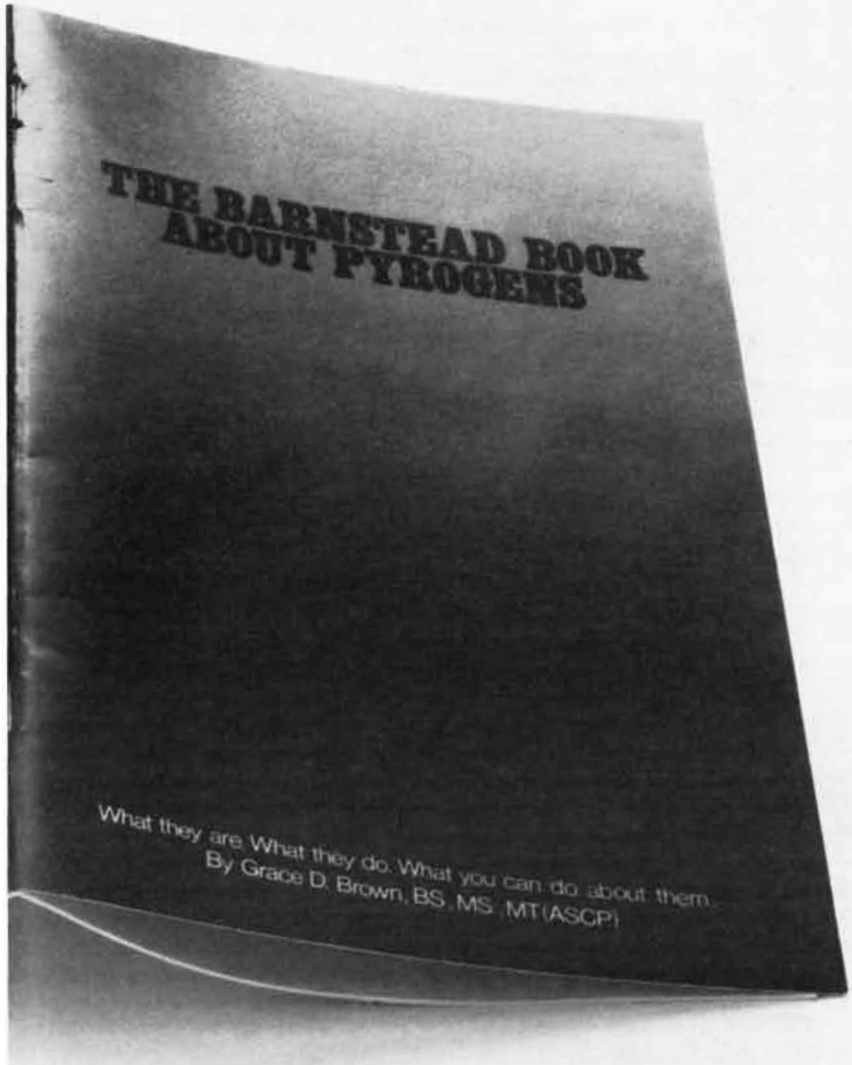
The damage they can do.

And what you can do about them.

Despite the fact that this little book proves that the best way to make sure you're not going to have a pyrogen problem is to run your water through one of our new TD-15 stills, it's much more than a Barnstead commercial message.

It's the last word on pyrogens.

And if you're making eye drops, mouthwash, toothpaste, pancake, lipstick, eye liner, intravenous or other medical solutions, anything that goes on or into a human being it's the end of your search to make it safer.



We're not in the printing business so we've got to charge you a 25¢ handling charge for handing your book to you.

But any information you want about our TD-15's is free.

Some of the best things you can get in life still are.

Even if little books about what you can do with it aren't.

Mr. Norman McLellan

BARNSTEAD
SYBRON CORPORATION

225 Rivermoor St./Boston, Mass. 02132/(617) 327-1600

☐ Enclosed is my 25¢ for The Barnstead Book About Pyrogens.
☐ Please send me my free information on your TD-15 stills.

Name_____
Title_____
Company_____
Address_____
City_____ State_____ Zip_____

THE AUTHORS

PHILIP M. HAUSER ("The Census of 1970") is professor of sociology and director of the Population Research Center at the University of Chicago. For the past several years he has been the chairman of two committees advising the Bureau of the Census: the Advisory Committee on Population Statistics and the Advisory Committee of the American Statistical Association. From 1938 to 1947 he was a member of the staff of the bureau. When he left the bureau in 1947 to join the University of Chicago, he was deputy director. In 1950–1951 he served as acting director of the Census, and from 1947 to 1951 he was the U.S. representative to the Population Commission of the United Nations. He received his undergraduate education and his Ph.D. from the University of Chicago.

MAURICE R. HILLEMAN and ALFRED A. TYTELL ("The Induction of Interferon") are respectively vice-president of the Merck Sharp & Dohme Research Laboratories and director of cell biology at the Merck Institute for Therapeutic Research. Hilleman is also director of virus and cell-biology research at the Merck Institute. After receiving a B.S. from Montana State College in 1941 and a Ph.D. from the University of Chicago in 1944, he worked principally at E. R. Squibb & Sons (1944–1948), the Army Medical Service Graduate School (1948–1955) and the Walter Reed Army Institute of Research (1956–1958). He went to the Merck Institute in 1958. Tytell studied chemistry and chemical engineering at Tufts University (B.S. 1936) and microbial chemistry and organic chemistry at the Massachusetts Institute of Technology (Ph.D. 1940). After teaching for 12 years at the University of Cincinnati College of Medicine, he went to the Merck Institute in 1952.

MORTON B. PANISH and IZUO HAYASHI ("A New Class of Diode Lasers") are members of the technical staff of the Bell Telephone Laboratories. Panish, who is head of the Materials Science Research Department, was educated at the University of Denver and Michigan State University. He received his Ph.D. in physical chemistry from Michigan State in 1954. Thereafter he worked at the Oak Ridge National Laboratory and the AVCO Corporation. He joined Bell Laboratories in 1964. Hayashi was born in Japan, received his Ph.D. in physics from Tokyo University, worked for eight years at the Tokyo University Institute for Nuclear Study and came to Bell Laboratories the same year as Panish. The two men began their collaborative effort in 1966, investigating the chemical and luminescence properties of gallium arsenide. Soon afterward they took up the chemistry and physics of the aluminum gallium arsenide system, which led to the evolution of the injection laser through the steps described in their article.

LENNART HEIMER ("Pathways in the Brain") is research associate in the department of psychology at the Massachusetts Institute of Technology. A native of Sweden, he was awarded an M.D. at the University of Göteborg in 1963. From 1962 to 1965 he was senior instructor in the Göteborg department of anatomy. Encouraged by Knut Larsson of Göteborg, he became interested in the organization of the olfactory areas of the brain, which is discussed in his article. In 1966 he went to M.I.T., where he has been working with Walle J. H. Nauta on techniques of staining brain tissue and on the organization of the forebrain.

SIMONE-ANTOINETTE DEYTS ("The Sacred Source of the Seine") is director of the Archaeological Museum of Dijon and has been in charge of the most recent archaeological excavations (1963, 1966 and 1967) at the source of the Seine. In addition she heads the laboratory where the wood sculpture uncovered in these excavations is preserved and restored. Born in Tunisia, she received her higher education at the University of Dijon, where she was awarded a doctorate in history in 1967. She has also participated in several other archaeological investigations, including the discovery and excavation of the ancient Greek port of Marseille.

PAUL GORENSTEIN and WALLACE TUCKER ("Supernova Remnants") are senior staff scientists at American Science and Engineering in Cambridge, Mass. They are members of the astrophysics group of American Science and Engineering, which has made notable contributions to X-ray astronomy. Gorenstein is an observer and experimental worker; Tucker is a theoretician. Gorenstein received his undergraduate education at Cornell University and his Ph.D. from the Massachusetts Institute of Technology (1962). After studying as a Fulbright postdoctoral fellow in Italy in 1963 and 1964, he joined American Science and Engineering in 1965. He is one of a group of "principal investigators" supported by the National Aeronautics and Space Administration who are responsible for an orbiting geochemistry experiment that will be conducted during the voyages of *Apollo 15* and *Apollo 16;* the experiment is designed to map gross differences in the chemical composition of the moon. Tucker studied at the University of Oklahoma and the University of California at San Diego, which awarded him a Ph.D. in 1966. From 1967 to 1969 he was assistant professor of space science at Rice University. He went to American Science and Engineering in 1969, and he is a part-time associate of the Harvard College Observatory.

RICHARD H. V. BELL ("A Grazing Ecosystem in the Serengeti") is an ecologist who is currently working in Zambia. Born in Kenya, he studied at Winchester College in England and at Magdalen College, Oxford. He received a B.A. in zoology from Oxford in 1963 and (after three years of fieldwork in the Serengeti National Park of Tanzania) a Ph.D. from the University of Manchester in 1969. He comments: "An ecologist has to be interested in everything from human behavior to geology, and the way that it is possible to trace causal relationships from rocks through soils, plants and animals to man has a great aesthetic appeal to me. At the same time I take great pleasure in visual objects: painting, sculpture, photographs, architecture and natural scenes. Combining these two outlooks, an ecologist can help to create an environment that man can live in, work in and enjoy." The work described in Bell's article was financed by the Nuffield Foundation.

FREDERICK MURPHY and DAVID E. YOUNT ("Photons as Hadrons") are respectively a member of the Experimental High-Energy Physics Group at the University of California at Santa Barbara and associate professor of physics at the University of Hawaii. Murphy, who is based at the Stanford Linear Accelerator Center (SLAC), received a B.S. from Georgetown University in 1959 and his Ph.D. from Princeton University in 1967. Yount did his undergraduate work at the California Institute of Technology and was awarded his Ph.D. at Stanford University in 1933. From 1962 to 1964 he taught physics at Princeton, where he worked in the same group with Murphy. Yount went to SLAC in 1965 and to the University of Hawaii in 1969. He is currently involved in photon and neutrino experiments proposed for the National Accelerator Laboratory at Batavia, Ill.

A SURPRISE CAN KILL

A hurricane.

Six hundred people were killed by one in New England in 1938.

Four years later, eleven thousand were killed by one in India.

The winds of a hurricane are over 75 miles an hour.

It takes the form of a circle or an oval, sometimes as much as 500 miles in diameter.

There is no prevention.

But if we reduce the element of surprise, we can reduce its deadly potential.

And we've taken steps so that it will never surprise us again.

RCA has designed, developed and produced the world's most advanced family of weather satellites. It's called the ITOS program.

It's the most sophisticated weather detector yet devised by man, and perhaps one of the most beneficial facilities man has created since he broke the bonds of earth.

From the detailed photographs transmitted back to earth, we can see the formation and plot the course of hurricanes, typhoons, storm centers. We can spot sea ice and snow cover. Calculate wind speeds.

Anywhere on earth.

This program exemplifies the importance of interdisciplinary engineering, the engineering concept of the future.

It is becoming increasingly more apparent that in the future, the engineer will touch virtually every aspect of our lives. From medicine to the arts to education to leisure products.

And he will do this in conjunction with other disciplines to complement, overlay, and unite his output for an end result that is both balanced and complete.

Think for a moment about an endeavor which, like meteorology, is seemingly unrelated to classical engineering: the graphic arts industry.

Recently, RCA engineers, working in conjunction with that industry, developed the world's most advanced electronic typesetting machine. It's called the VideoComp.

It can set type as fast as 900 lines per minute. Compare this with the 15/minute of mechanical typesetting, or the 300/minute of photosetters.

It is our goal ultimately to produce an RCA electronic printing system that will accept a manuscript as input and deliver printed copy — packaged, addressed and sorted — to the shipping dock.

Now consider another field which had previously been outside the realm of the engineer: medicine and health services.

We've been involved in the development of a mobile laboratory to provide comprehensive physical examinations in on-the-spot locations to people otherwise out of the mainstream of our health services.

And we've researched a high-speed turbine drill to increase the efficiency and decrease the pain in dental work.

And we've done extensive work in developing facilities for teaching speech to the retarded.

Communications? Computer education? Oceanography? We're involved with them. And with so much more.

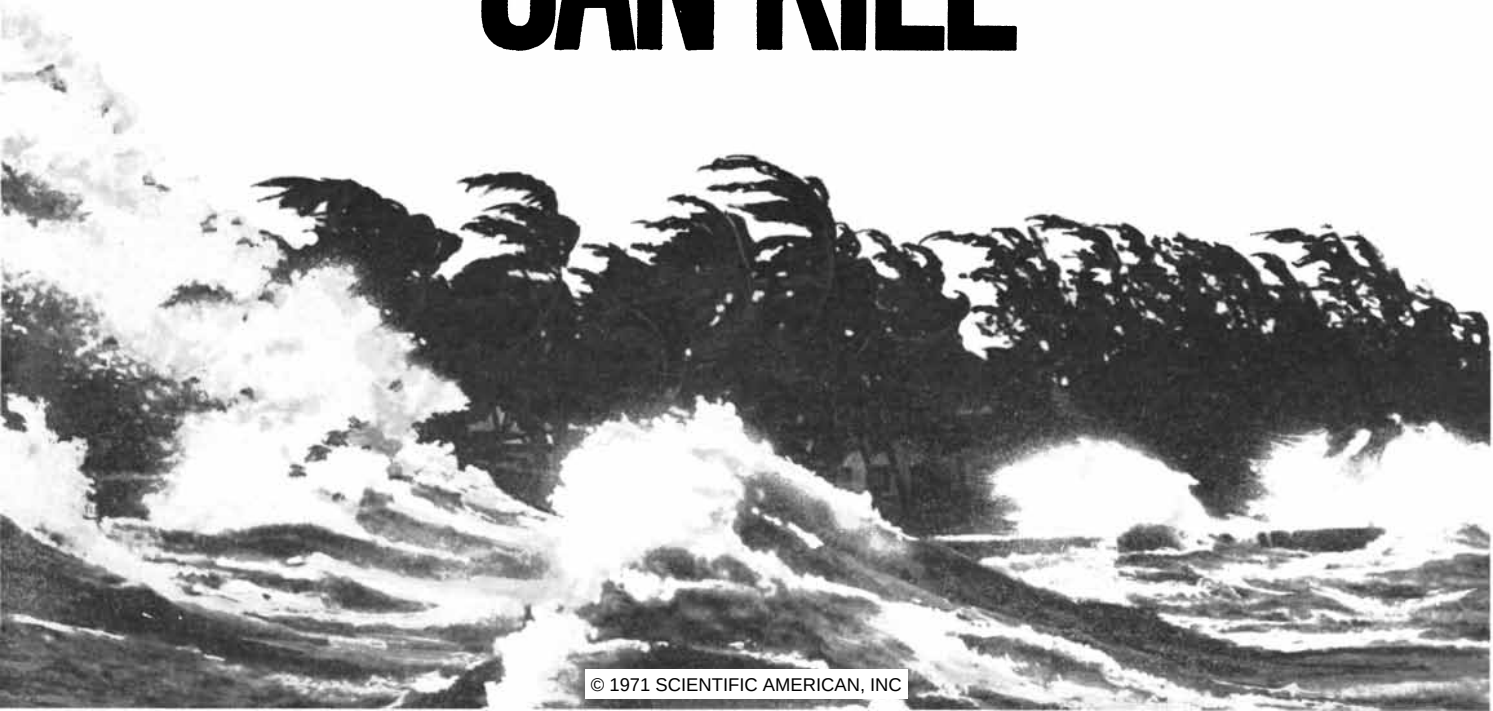
We like to think that at RCA the drive toward innovation should be everybody's concern.

Because this concern is needed if we are to continue the forward direction of creative technology for people.

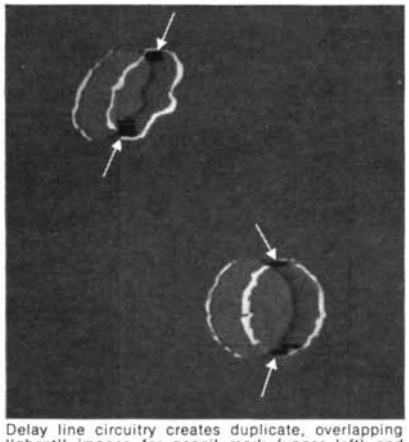
If you're interested in a comprehensive index of over 1100 technical papers published recently by RCA scientists and engineers, let us know.

Write to Mr. D. L. Crosier, RCA, Bldg. 2-2, Camden, New Jersey 08102. Of course, we're an equal opportunity employer.

RCA



How Western Electric spots dots with TV.



Delay line circuitry creates duplicate, overlapping "ghost" images for pencil mark (upper left) and regular .025″ dot (lower right). The points where the images cross can be made to produce black areas (marked by arrows). The ghosts are moved right or left, and the positions of the black areas tested

How would you use automation for drilling holes when every hole has to be in a precise spot, but every spot isn't exactly where you expect it to be?

That became a critical problem for Western Electric when we started using a polyethylene with very superior electrical properties as the base laminate in printed circuit boards. Unfortunately this material showed a tendency to shrink when the circuit pattern was etched in the copper on the board—not enough to affect its electrical properties, but enough to dislocate the dots which indicated where holes were to be drilled for placing components. The shrinkage was unpredictable but could move a dot by as much as 1/32″—more than its own diameter.

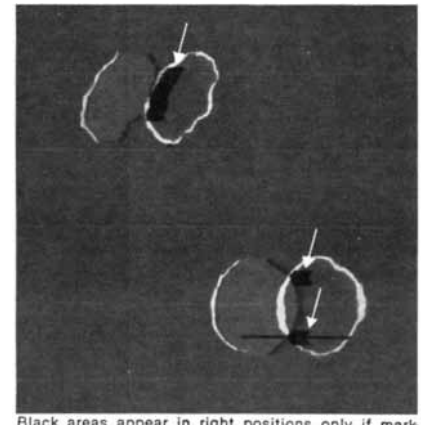
What that meant was that we could not use a conventional tape-controlled drilling machine, which put the drill in precisely the same spot every time. We needed a machine that could, in effect, "see" a dot, no matter where it was, tell it apart from an accidental marking of about the same size and shape, and put a drill right through it.

Engineers at our Greensboro, N.C., Works made such a machine. It consists of a TV camera hooked up to circuitry which removes the grays from the picture, turning it into a series of true on-or-off digital pulses; and logic circuitry which can respond to such pulses and activate a mechanism which moves the board around.

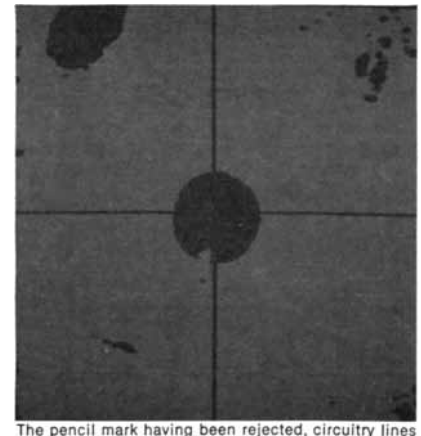
Once the dot has been roughly centered beneath the camera, delay line circuitry creates a second image which is first superimposed on the original, then moved by successive stages to the right. The points where the two images meet—i.e., where the two circles cross—will be in certain positions if and only if the original dot is a perfect circle of the proper size. Logic circuitry tests these positions, and if they are not exactly right, it won't drill and the circuitry sets itself to search for another dot. If they are, it lines the dot up, moves the camera aside, moves the drill into position, and there you are.

A simple, fast, economical, and very, very precise solution to a vexing problem. The kind of solution that helps us keep the quality of the equipment we make for the Bell System high, and the costs low.

Western Electric



Black areas appear in right positions only if marking is a circle of the proper size.



The pencil mark having been rejected, circuitry lines up acceptable dot for drilling.

Established 1845 SCIENTIFIC AMERICAN July 1971 Volume 225 Number 1

The Census of 1970

The early figures from the 19th count of the U.S. population show a reduced rate of growth, large migrations to the West and South and a continued shift of whites to the suburbs and blacks to cities

by Philip M. Hauser

The 1970 U.S. census, the nation's 19th decennial inventory of its human resources, will undoubtedly reap the richest harvest of useful information in our history. For the first time computer tapes of the statistical data (carrying no identification of individuals) are being made widely available to research organizations for analyses of one kind or another. The Bureau of the Census itself, universities and various other public and private research agencies will be tabulating and mining the data throughout the coming decade. From the great variety of studies that will be undertaken we can expect to obtain an unprecedented wealth of insights into the demographic, social, economic and educational trends under way and the health and life style of the American people.

The extensive information that was obtained from a sampling of the population has not yet been tabulated. What we have so far is a set of statistics based on the questions that were asked of everyone in the census. In this article I shall discuss these findings against the background of the trends shown in the censuses of the preceding decades.

First let me comment on the reliability of the figures that follow. Obviously no census covering so large a country and population can ever be 100 percent accurate or complete. The limitations in the state of the art of census-taking and in available funds, and the independent spirit of the American people that leads many to refuse to cooperate, result in an unavoidable margin of error for the overall count. In the past three decades, however, the census staff has made great progress in developing methods for estimating the reliability of the responses and the degree of completeness of the enumerations. The evaluation after the 1960 census indicated that the total population was undercounted by 3.1 percent, the count having failed to reach an estimated 2.2 percent of the white population and 9.5 percent of the nonwhite. New procedures designed to improve the completeness of coverage were adopted for the 1970 census, but the improvements in technique may have been offset by the effect of the rising alienation of many Americans, particularly among blacks and members of other minorities. No doubt again, as in the past, the blacks and the poor were undercounted somewhat in relation to the rest of the population. The degree of accuracy of this census has not yet been evaluated. We can be confident, however, that the margin of error was small enough for the actual counts to truly reflect the basic trends over the past decade.

The broad-scale tabulations show that the 1960's were a period of dramatic changes in the distribution and structure of the U.S. population, some of them already well known, others not so well known. There were significant changes in the population's rate of growth, in its regional and local distribution, in its metropolitan patterns, in its racial composition and in its age structure. Let us look at these changes and their implications.

In the 20th century we have seen a series of ups and downs in the U.S. population growth. In the first decade of the century the growth rate, boosted by a wave of immigration, reached a high of 21 percent. In the depression decade of the 1930's it fell to a low of 7.2 percent. In the 1940's the economic recovery occasioned by World War II raised the population increase to 14.5 percent, and the postwar "baby boom" and rapid growth of the economy in the 1950's pushed the population growth rate up to 18.5 percent. Now the 1970 census shows that in the 1960's the rate fell back to 13.3 percent (although in absolute numbers the increase was nearly 24 million, second only to the record 28-million increase in the 1950 decade). The total of U.S. residents as of April 1, 1970, was 203,184,772, not counting an additional 1,580,998 citizens who were living abroad in the armed forces and in U.S. Government employment.

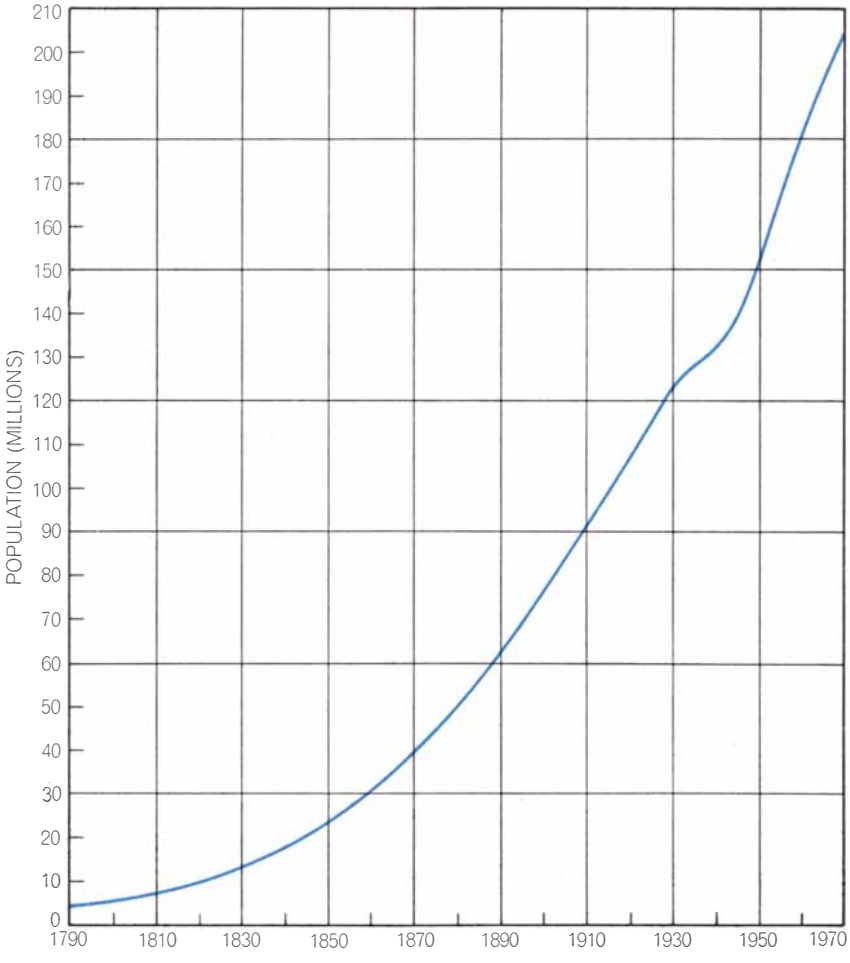
It seems, then, that in the 1960 decade the U.S. may have returned to the long-term downward trend in population growth that had been interrupted in the 1940's and 1950's by the effects of World War II and its aftermath. What were the reasons for the sharp drop in the growth rate in the 1960's? Immigration was not a factor; during

the decade there was a net immigration of three million persons, about 400,000 above the number in the 1950's. The overall growth rate fell because of a decline in natural increase, compounded of a drop in the number of births and, more important, an increase in the number of deaths. Births in the 1960's totaled 39 million, two million less than the preceding decade's total, and the number of deaths rose to 18.2 million, more than 2.5 million higher than in the 1950's. Thus there was a drop of about 4.5 million in the natural increase. At least part of the rise in deaths must be attributed to the increase of the population and of the number of older people in it.

All four of the nation's distinct regions—the West, the South, the Northeast and the North Central states—participated in the population growth, and also in the slowing of the growth rate, during the 1960's. In general the trends of recent decades continued: the West remained the fastest growing region with a 24 percent increase (down from 39 percent in the 1950's); the South was again the second fastest growing region with a 14.2 percent increase (compared with 16.5 percent in the 1950's); the Northeast had a 9.7 increase (down from 13.2 percent), and the North Central states, which had been third in growth rate in the 1950's, dropped to fourth in the 1960's with a 9.6 percent increase (down from 16.1 percent). In total population the South remained first with 62.8 million, the North Central states second with 56.6 million, the Northeast third with 49 million and the West fourth with 34.8 million.

There was a continuation of the main recent trends in the migration of people from region to region. The North Central states, continuing a trend that began in the 1930's, had a net loss of 752,000 by migration in the 1960's; of the 12 states in this region only three (Michigan, Wisconsin and Missouri) showed an excess of in-migration over out-migration, and their gains were small. On the other hand, the West continued to show the largest gains by migration, as it has in every census since 1920; in the 1960's its net in-migration was 2.9 million, California accounting for most of the gain with 2.1 million.

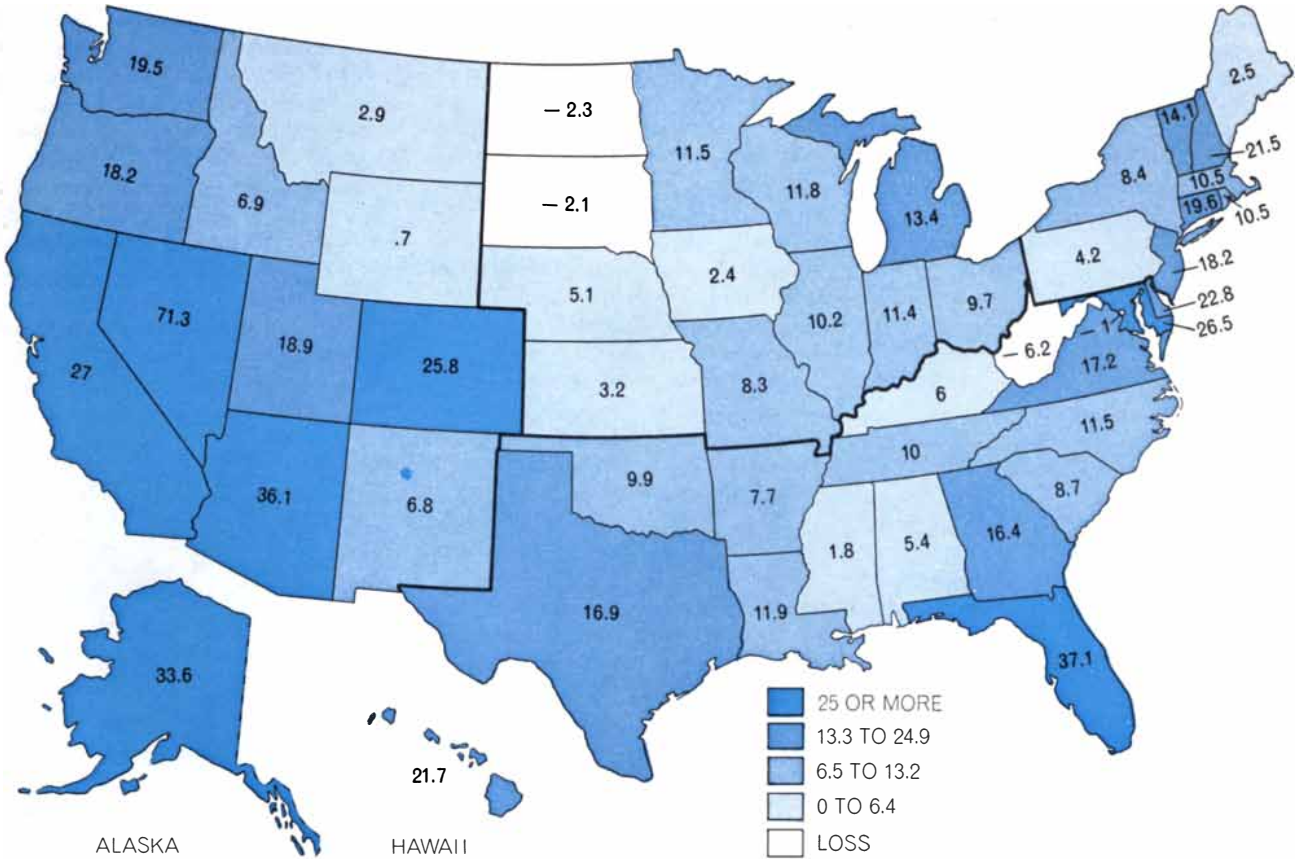
There were, however, a few dramatic shifts in the migration picture. In the Northeast, although the region as a whole continued to experience a net gain (324,000 in the 1960's), the state of New York for the first time had more out-migrants than in-migrants. This development was matched by an equally dramatic one in the South: for the first time since 1880 the region as an entity had a net gain by migration. The increase, amounting to nearly 600,000 for the region as a whole, was attributable to gains by just a few states. Florida had a net in-migration of more than 1.3 million, and there were substantial gains in Maryland, Texas and Virginia. Nine of the 16 Southern states and the District of Columbia continued, however, to experience net migration losses.



U.S. POPULATION HAS GROWN from 3.9 million at the first census in 1790 to 203.2 million at the census of 1970. The intermediate figures are: 1800, 5.3 million; 1810, 7.2 million; 1820, 9.6 million; 1830, 12.9 million; 1840, 17.1 million; 1850, 23.2 million; 1860, 31.4 million; 1870, 38.6 million; 1880, 50.2 million; 1890, 63 million; 1900, 76 million; 1910, 92 million; 1920, 105.7 million; 1930, 122.8 million; 1940, 131.7 million; 1950, 150.7 million; 1960, 179.3 million. The dip in the curve occurred during the depression years of 1930 to 1940.

In the U.S. as elsewhere the main attractions for migratory movement are greater economic opportunity and a less rigorous climate. The strong streams of migration to California and Florida represent a combination of these factors. Usually the free movement of people toward areas of better economic opportunity serves to improve the balance between the population and the nation's economic resources. Thus over the long run the shifts of population in the U.S. can be expected to smooth out regional differences in the standard of living.

At the state level the variations in population change were striking. California had a 4.2 million increase (27 percent) in the 1960's and, with nearly 20 million residents, passed New York as the most populous state. Florida was next in increase in numbers (1.8 million), and other states gaining more than a million persons were Illinois, Texas, New York, New Jersey and Michigan. The states with the highest percentages of growth were Nevada (71 percent), Florida (37 percent), Arizona (36 percent) and Alaska (34 percent). At the other end of the scale, three states declined in population. West Virginia's



CHANGES IN POPULATION by states are indicated by percentage of gain or loss and by tone of color (*see key at right*). Heavy lines delineate the four regions into which the Bureau of the Census divides the nation: Northeast, South, North Central and West.

shrinking economic base, which started a downtrend in population in the 1950's, caused a further loss of 6.2 percent, or 116,000 people, in the 1960's. In North Dakota and South Dakota the population decline that was initiated by the depression and the drought in the 1930's was resumed in the 1960's after a small gain in the 1950's.

Even more striking than the differences among the states was the shifting of populations at the county level—the changes in what might be called the fine structure of the population distribution. No fewer than half of the counties in the U.S. lost population during the 1960's, and of the counties that did not lose, half had only small gains: less than 10 percent. In other words, the nation's large increase in population was concentrated in a fourth of the counties, as it had been in the 1950's. This succinct statistic reflects the massive metropolitanization of the American people, steadily gravitating from the farm and rural areas to the cities and their suburbs.

In effect urban life swallowed up the nation's entire population increase: 24 million. In 1970, 149.3 million persons, 73.5 percent of the total population, were living in cities and towns (of 2,500 or more), compared with 69.9 percent in 1960. The rural population (persons living on farms or in places with a population of less than 2,500) declined from 54.1 million to 53.9 million, and the farm population dropped from 13.5 million in 1960 to about 10 million in 1970.

Interestingly, the Northeast, which is commonly regarded as the most urbanized region, lost this distinction in the 1960's. The West succeeded to the lead; its population in 1970 was 82.9 percent urban (compared with 77.7 percent in 1960), and the Northeast was second with 80.4 percent (about the same percentage as in 1960). In the North Central states the urban population rose to 71.6 percent and in the South it reached 64.6 percent. The South actually had the largest increase in proportion urban: 10 percentage points.

There was a slight increase in the number of large cities; the number with a population of more than a million, for example, rose from five to six (Houston joined the club), and the cities with a population between half a million and a million increased in number from 16 to 19. Some cities exhibited spectacular gains: Jacksonville more than doubled its population to 529,000 (having annexed within its limits the entire large county of which it was the center); Indianapolis gained 56 percent, and large advances in population were experienced by several cities in the Southwest, Phoenix increasing by 32 percent, Houston by 31 percent and Dallas by 23 percent.

Most of the nation's oldest and largest cities, however, stopped growing or showed a decline. Of the 25 largest cities, 12 lost population during the decade. Among them were Chicago and Philadelphia, the latter dropping below two million for the first time since 1940. Four of the 12 (Detroit, Cleveland, St. Louis and Pittsburgh) lost more than 10 percent; St. Louis, with a 17 percent loss, fell back to its smallest population since the beginning of the century.

When we switch our attention from the cities themselves to the metropolitan complexes they have generated, including their rings of suburban settlements, we see a different picture. Of the 50 leading metropolitan areas in the nation

all but one showed substantial growth in the past decade. (The St. Louis metropolitan area, notwithstanding the heavy loss in the central city, grew in population by more than 11 percent in the 1960's.) The nation's 243 Standard Metropolitan Statistical Areas (SMSA's), defined as cities of 50,000 or more together with their counties and the counties they dominate, had a net population growth of 16.6 percent during the decade. This was not as great as the SMSA growth in the 1950's (26.4 percent), but it still accounted for absorption of 84 percent of the total national growth in the 1960 decade. The 1970 count shows that 140.2 million people, 69 percent of the total population, now live in the metropolitan areas, compared with 67 percent in 1960 (based on SMSA's as defined for 1970). As in the 1950's, the metropolitan areas continue to grow considerably faster than the rest of the nation, whose population growth during the past decade was only 6.5 percent.

Analyzing the metropolitan areas in detail, we find, not surprisingly, that the suburbs grew much more rapidly than the central cores. Whereas the aggregate growth of the 243 central cities themselves was only 5.3 percent, in their suburban rings it was 28.3 percent. For the first time the suburban population outnumbered that in the cities: 76.3 million in the suburbs against 63.9 million within the cities. Between 1960 and 1970 the proportion of the metropolitan population living in the outer rings rose from 49.5 percent to 54.4 percent. In large part, of course, the change in the statistical picture betokens local outward movement from the city to the suburbs. In eight of the 10 largest U.S. metropolitan areas the central city lost population during the 1960's. Of these cities only Los Angeles registered a substantial gain: 12.2 percent.

According to the 1970 count the 10 largest metropolises are the New York metropolitan area (11.4 million), Los Angeles–Long Beach (about seven million), the Chicago area (6.9 million), the Philadelphia area (4.8 million), the Detroit area (4.2 million), San Francisco–Oakland (3.1 million), the District of Columbia–Maryland–Virginia area (2.8 million), the Boston area (2.7 million), the Pittsburgh area (2.4 million) and the St. Louis area (2.3 million). It is interesting to note that in the last four of these SMSA's about three-fourths of the population lives outside the central city. During the past decade in metropolitan areas of the one-million-to-three-million class the growth rate in the suburbs was 31.9 percent, eight times the growth rate in their central cities.

As in the 1950's, the growth rate of SMSA's, with the exception of the very largest ones, varied according to size: the larger metropolitan areas had the higher growth rates. SMSA's with a population in the range of one to three million grew by 20 percent in the 1960's, whereas in those above three million the growth rate was only 10.7 percent. This seems to suggest that three million inhabitants may represent an upper limit of reasonable size, beyond which urban problems begin to become unmanageable. The question will undoubtedly be of increasing concern to the nation and warrants intensive research.

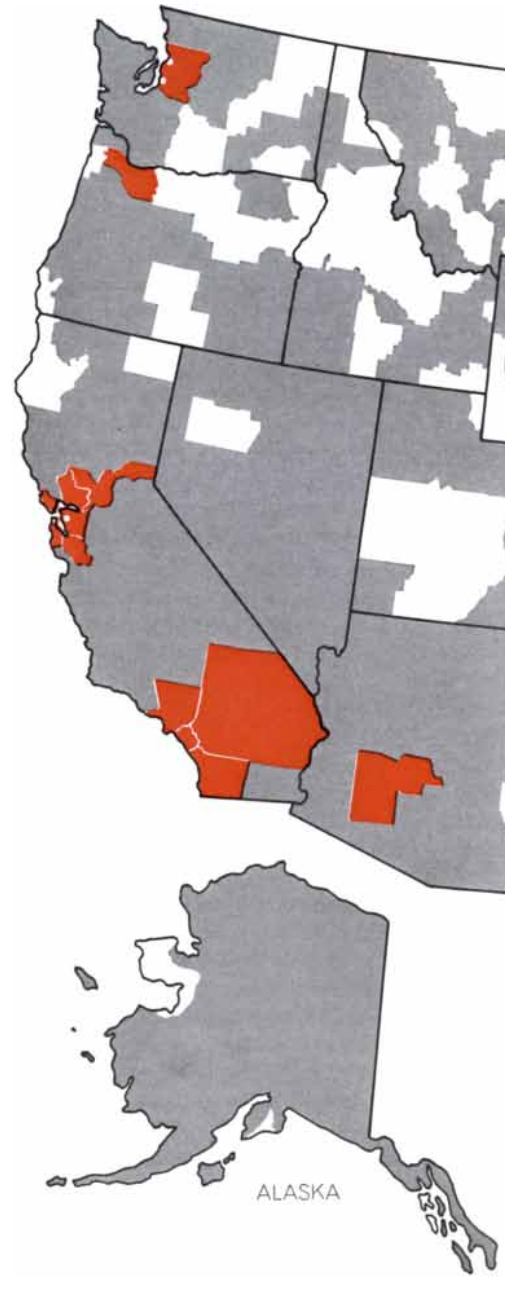
The story the 1970 census tells about the U.S. population's racial composition and movements is one that in the main is already familiar. During the past decade the black population grew by 20.1 percent, compared with 11.8 percent for the white population, so that the proportion of blacks in the total U.S. population rose from 10.5 percent in 1960 to 11.2 percent in 1970. The number of blacks was 22.7 million.

The migration of blacks from the South to other regions continued during the 1960's, although at a lower rate than in the preceding two decades. The net out-migration of blacks from the South was 1.4 million; the region showed a small overall gain through migration because of a net in-migration of about 1.8 million whites, mainly in Florida. By states the exodus of blacks was about 279,000 in Mississippi, 231,000 in Alabama, 197,000 in South Carolina, 175,000 in North Carolina, 163,000 in Louisiana, 154,000 in Georgia, 112,000 in Arkansas and lesser numbers in other Southern states. In the South only Delaware, Maryland, Kentucky and the District of Columbia showed net gains of blacks by migration.

Through natural increase there was a small net rise (6.7) in the total number of blacks in the South, but as a result of the exodus to other regions the proportion of the nation's blacks who live in the South fell to little more than half (53.2 percent) by 1970, against 59.9 percent in 1960 and 77 percent as recently as 1940. Blacks still constitute 19.2 percent of the South's population, compared with 8.9 percent of the population in the Northeast, 8.1 percent in the North Central states and 4.9 percent in the West. The blacks' migrations are rapidly reducing the regional differentials, however.

In the 1960 decade the Northeast gained 1.3 million blacks (a 43.4 percent increase), the North Central states 1.1 million (a 32.7 percent increase) and the West 609,000 (a 56.1 percent increase). The states experiencing the largest increases of blacks through net in-migration were New York with 396,000, California with 272,000 and Illinois, Michigan and New Jersey with more than 100,000 each.

The census figures vividly document the well-known fact that the black migrations were mainly to central cities in

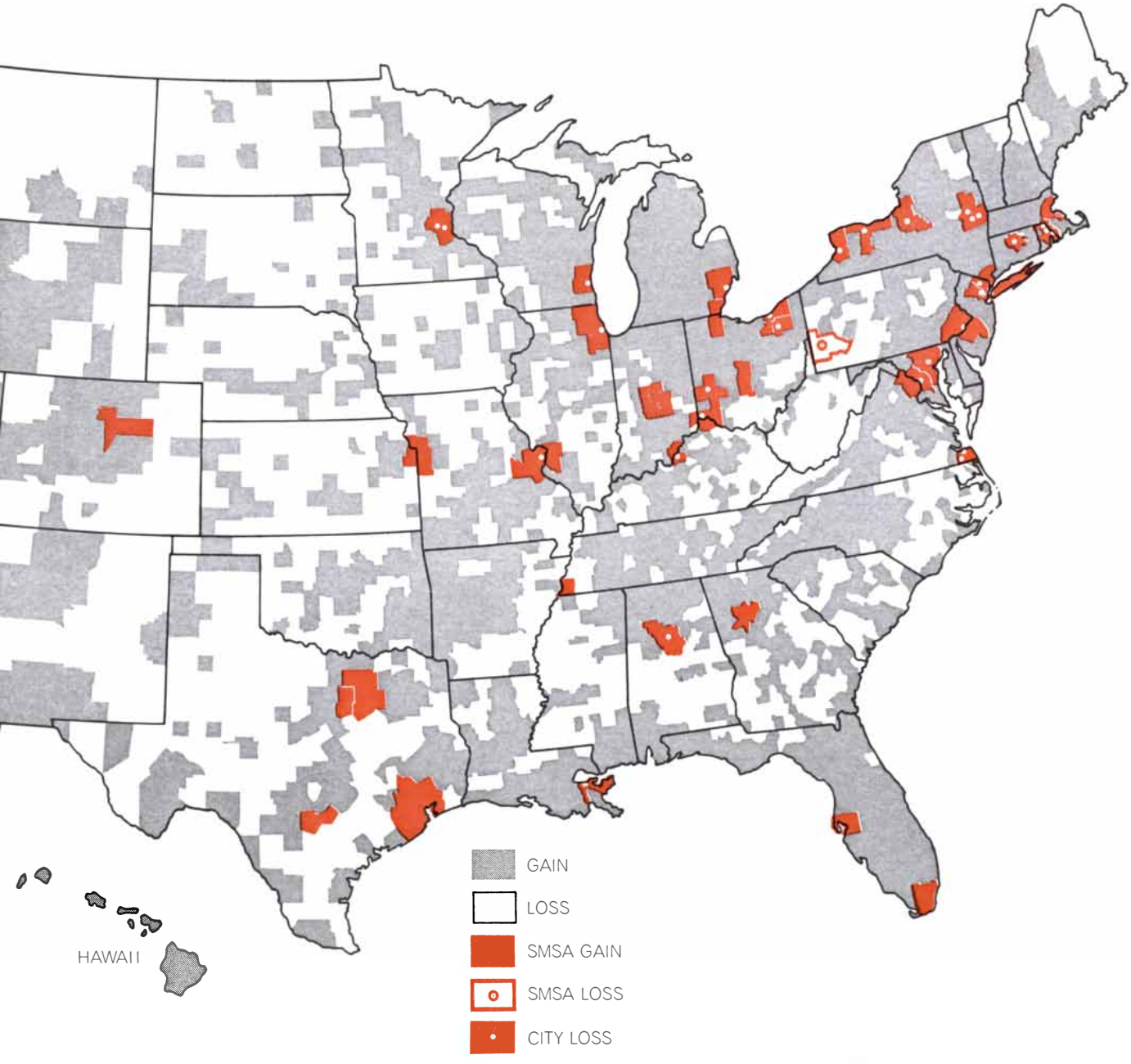


REDISTRIBUTION OF POPULATION is portrayed by counties and 50 largest Standard Metropolitan Statistical Areas (SMSA's).

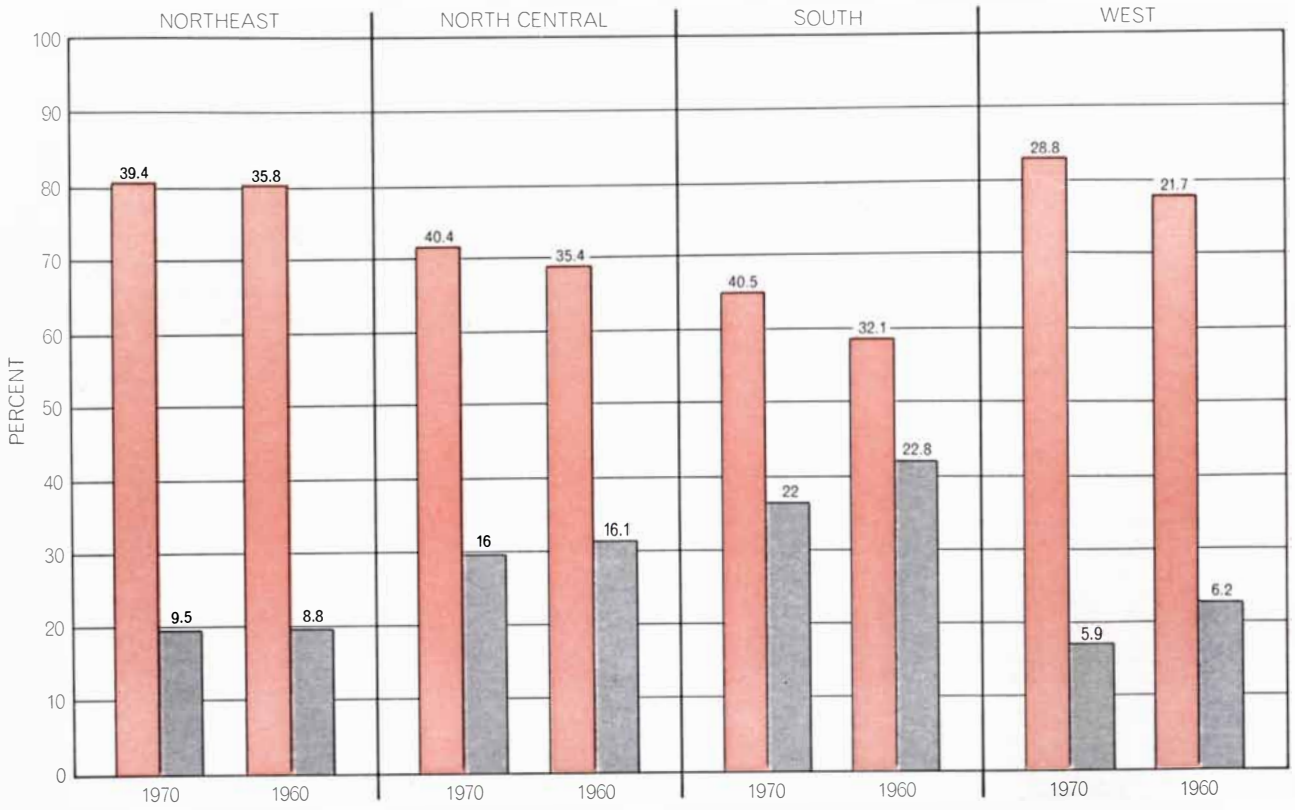
the large metropolitan areas. In four cities blacks now constitute a majority of the population. In Washington, D.C., the percentage of blacks rose during the decade from 53.9 to 71.1; in Newark, N.J., from 34.1 to 54.2; in Gary, Ind., from 38.8 to 52.8, and in Atlanta, from 38.3 to 51.3. In seven other cities the population is more than 40 percent black; these are Detroit, St. Louis, Baltimore, New Orleans, Wilmington, Del., Birmingham, Ala., and Richmond, Va. The rising ratio of blacks in some cities was due in part to the exodus of whites. In Chicago, for example, the white population declined by more than half a million, or 18.6 percent, in the 1960's. Several other cities showed even larger percentage losses of whites: 29.2 percent in Detroit, 31.3 percent in St. Louis and 36.7 percent in Newark.

The growth pattern in the suburban rings around the central cities is, of course, quite different. With conditions in the suburbs generally unfavorable for the influx of blacks, these areas usually experienced only small increases of blacks and large increases of whites, so that in many cases the proportion of whites rose. Even in suburban areas where blacks made large gains in percentage terms the actual increases in number of persons were relatively small. For example, in the Los Angeles suburbs the blacks doubled their number with an increase of 124,000; around Washington they doubled in number with an increase of 84,000, and in the New York suburban area an increase of 77,000 blacks represented a gain of 55 percent in the total black population.

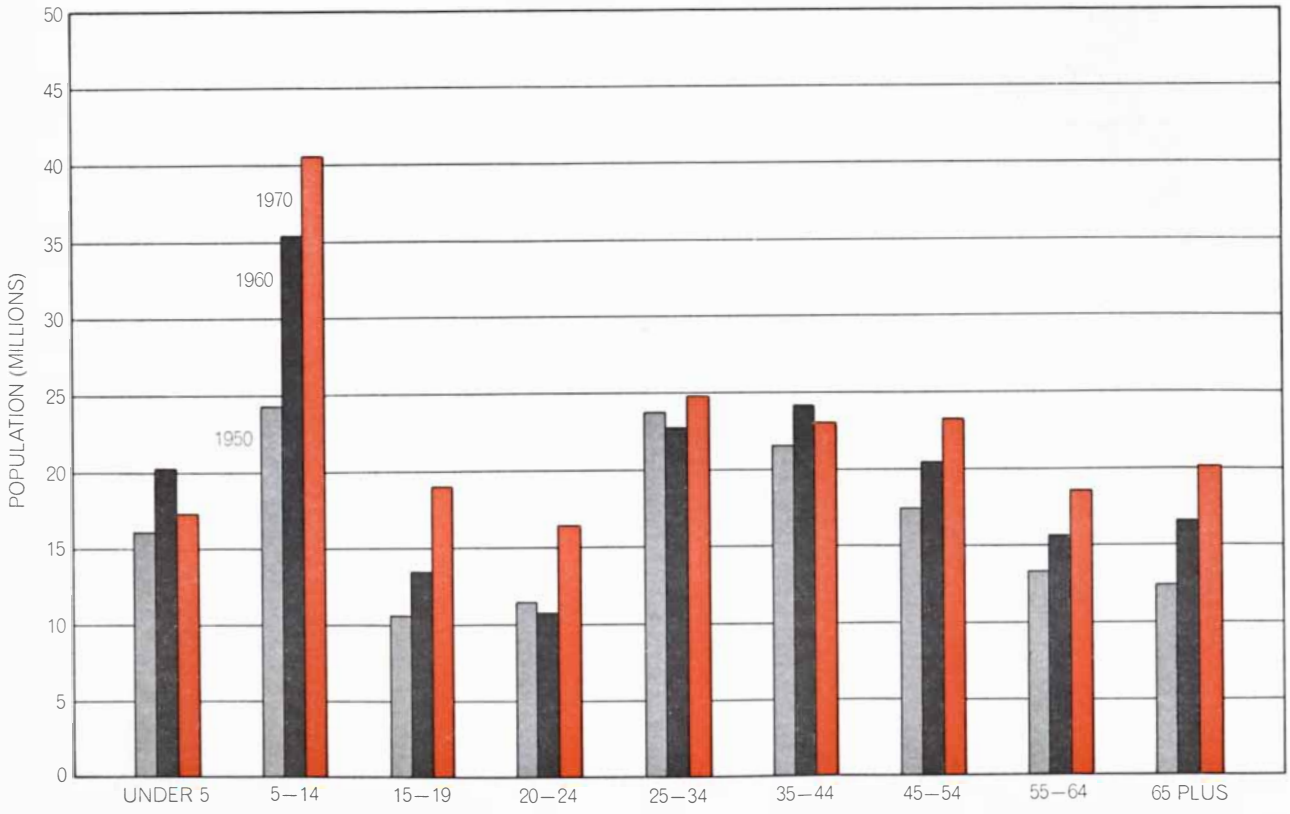
The 1970 census emphasizes in cold



Shown here are counties that gained population (*gray*), counties that lost (*white*), SMSA's that gained (*solid color*) and the single SMSA that lost (*white area bordered in color*). The white dots in most of the SMSA's indicate that the central city of the SMSA lost population during the decade while its suburban ring gained population. Boundaries of individual counties do not appear on map.



URBANIZATION OF POPULATION continued in all four census regions between 1960 and 1970. The bars show the percentage of urban (*color*) and rural (*gray*) population in each region for the years of the two censuses. Numbers give population in millions.



AGE DISTRIBUTION of the U.S. population showed significant shifts over three censuses. The distributions in the nine age groups regularly tabulated by the U.S. Bureau of the Census are compared for the censuses of 1950 (*gray*), 1960 (*black*) and 1970 (*color*).

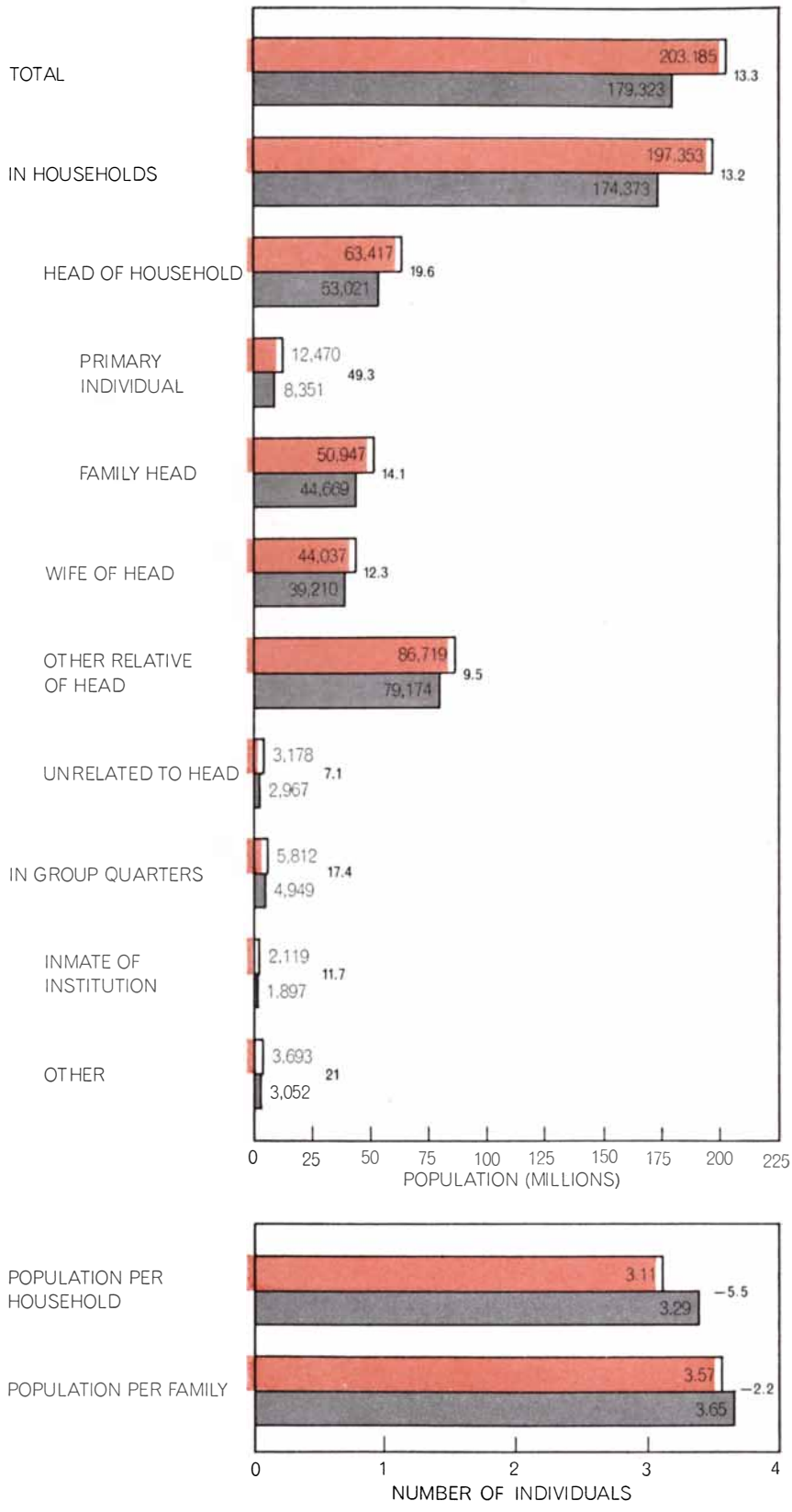
statistics the dimensions of the developing racial problem that reached an explosive climax in the 1960's. More and more the problem of race relations is extending into every region and population center of the nation, as the combination of civil-rights conflicts in the South and greater economic opportunity for blacks elsewhere continues to spur migration. In all likelihood the black exodus from the South will continue on a large scale in the 1970's, intensifying the problems of adjustment in other regions and involving the entire nation in these problems.

Among the general social questions on which the census can be expected to throw some light is the question of what is happening to the American family. So far we have only a few preliminary tabulations on this subject from the 1970 census. Their salient message is that the long-time trend to smaller families and smaller households in general is continuing. Whereas the population count of individuals increased by 13.3 percent, the number of families (defined as two or more related persons living together) increased by 14.1 percent, from 44.7 million families in 1960 to 50.9 million in 1970. This works out to a decrease in the average family size from 3.65 to 3.57.

Two factors are responsible for the declining size of the family as the census defines it: (1) the trend toward bearing fewer children and (2) the "undoubling," or splitting, of the traditional extended family so that older couples beyond the child-raising age now generally live apart from their children in their own household.

The number of nonfamily households (one person or a group of two or more unrelated persons living together) shows a sizable increase during the past decade: 49.3 percent. Adding these 12.5 million units to the family domiciles, we have a total of 63.4 million households in the nation, up from 53 million in 1960. The average size of all the households then is 3.11, down from 3.29 in 1960. In part the overall trend toward smaller households stems from changes in the age structure of the population, a factor I shall discuss below.

There were also increases during the 1960's in the numbers of inmates in institutions and of other persons living in group quarters (establishments containing five or more unrelated persons). Inmates in institutions numbered 2.1 million, an increase of 11.7 percent in the decade, and other persons in group quarters totaled about 3.7 million, up 21 per-



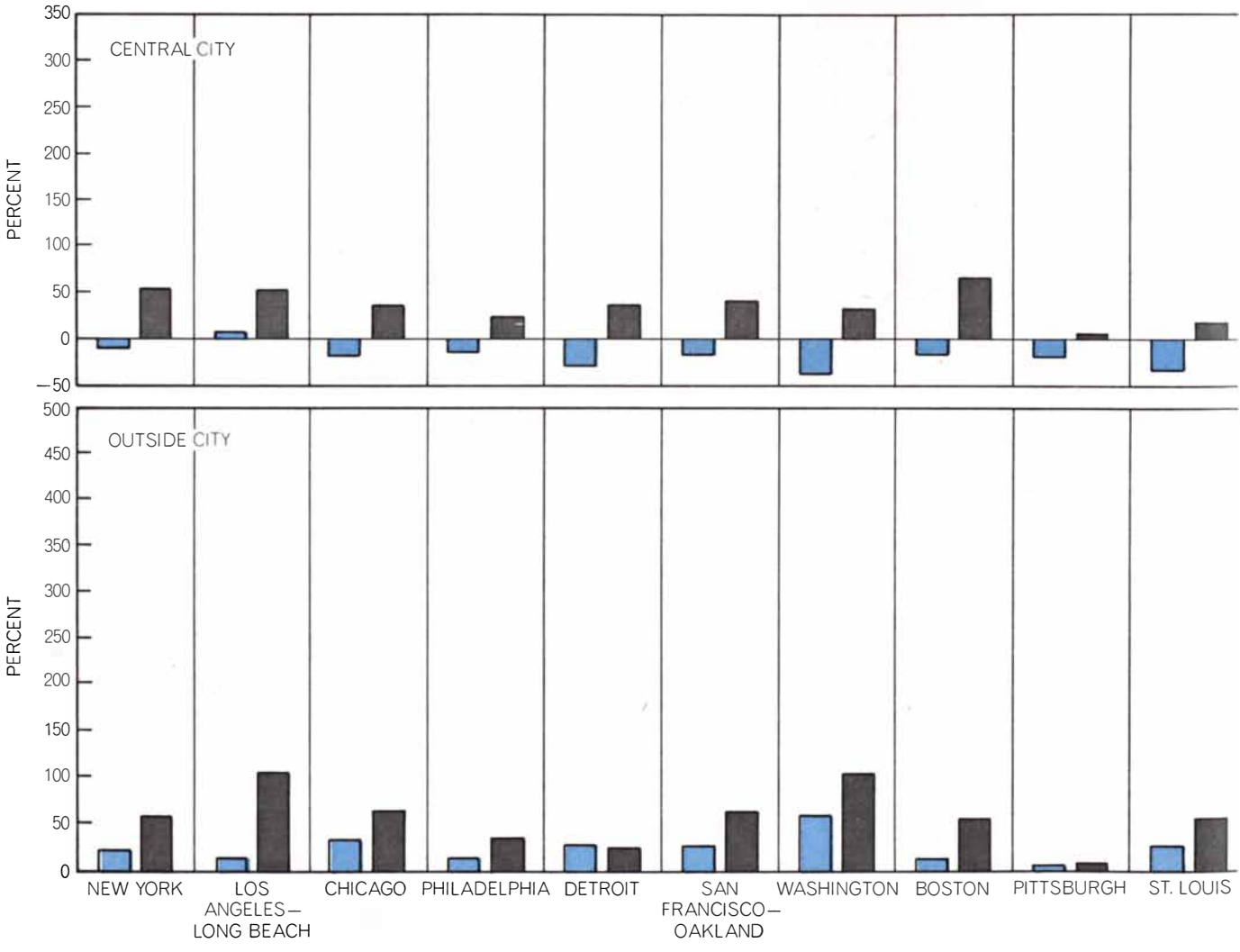
TREND TOWARD SMALLER FAMILIES is reflected in analysis of households. The total population (*top pair of bars*) increased by 13.3 percent between the 1960 census (*gray bar*) and the 1970 census (*colored bar*); the population living in households (*second pair of bars*) increased by about the same percentage. Heads of households, however, increased by a larger percentage (*third set of bars*), indicating a larger percentage of households and therefore smaller ones. "Primary individual" refers to a person living alone or with a nonrelative only; "Family head," to the head of a household living with one or more relatives.

cent from 1960. The reasons for these increases remain to be investigated.

The age composition of the U.S. population has exhibited wide swings in recent decades. The swings have resulted from the sharp fall in the marriage and birth rates during the depression 1930's, the baby boom following World War II, a drop in the birthrate between 1957 and 1968 and a new baby boom that began in 1968 as babies born after World War II reached reproductive age. Underlying these various episodes there is a basic long-term decline in the birthrate. There will be large changes in the age composition during the coming decades as these groups move through their life cycle. Such swings generate severe problems for the nation, particularly in the areas of education and employment.

Among the changes shown in the 1970 census was a 44.3 percent increase in the age group between 15 and 19, following a 23.9 percent increase in this group that had taken place in the decade ending in 1960. Since the age span roughly represents the population of high school age, the sharpest impact of these increases fell on the facilities for high school education. The pressure on those facilities became severe in the 1960's and will continue into the 1970's. The swelling of the late-teen-age population also had other important consequences. This is the main part of the age group that, by definition, is involved in juvenile deliquency. Consequently just from the increase in the size of this population we should have expected an increase of more than 40 percent in the amount of juvenile delinquency during the 1960's, even without a rise in the group's *rate* of delinquency. This consideration has generally been overlooked by police authorities, political officials and others who lament the conduct of the young generation.

The population in the age group from 20 to 24, after a 6.5 percent decline in the 1950's, shot up to an increase of 51.6 percent during the 1960's. Thus the late-teen-age and young-adult groups together constituted the greatest bulge in the population, representing nearly half of the decade's total increase. The combined increase of the two groups has had pronounced effects in a number of specific areas, notably the rate of automobile accidents and the problem of unemployment. Since drivers under 25, particularly males, account for a high proportion of the automobile accidents and fatalities, the recent increases in insurance rates are in large part a reflection of the increase in the size of this population group. The entrance of this group into the employment market has also made a major demand on the economy, a challenge that the economy has not yet successfully met, as unemployment is particularly high among the nation's young people.



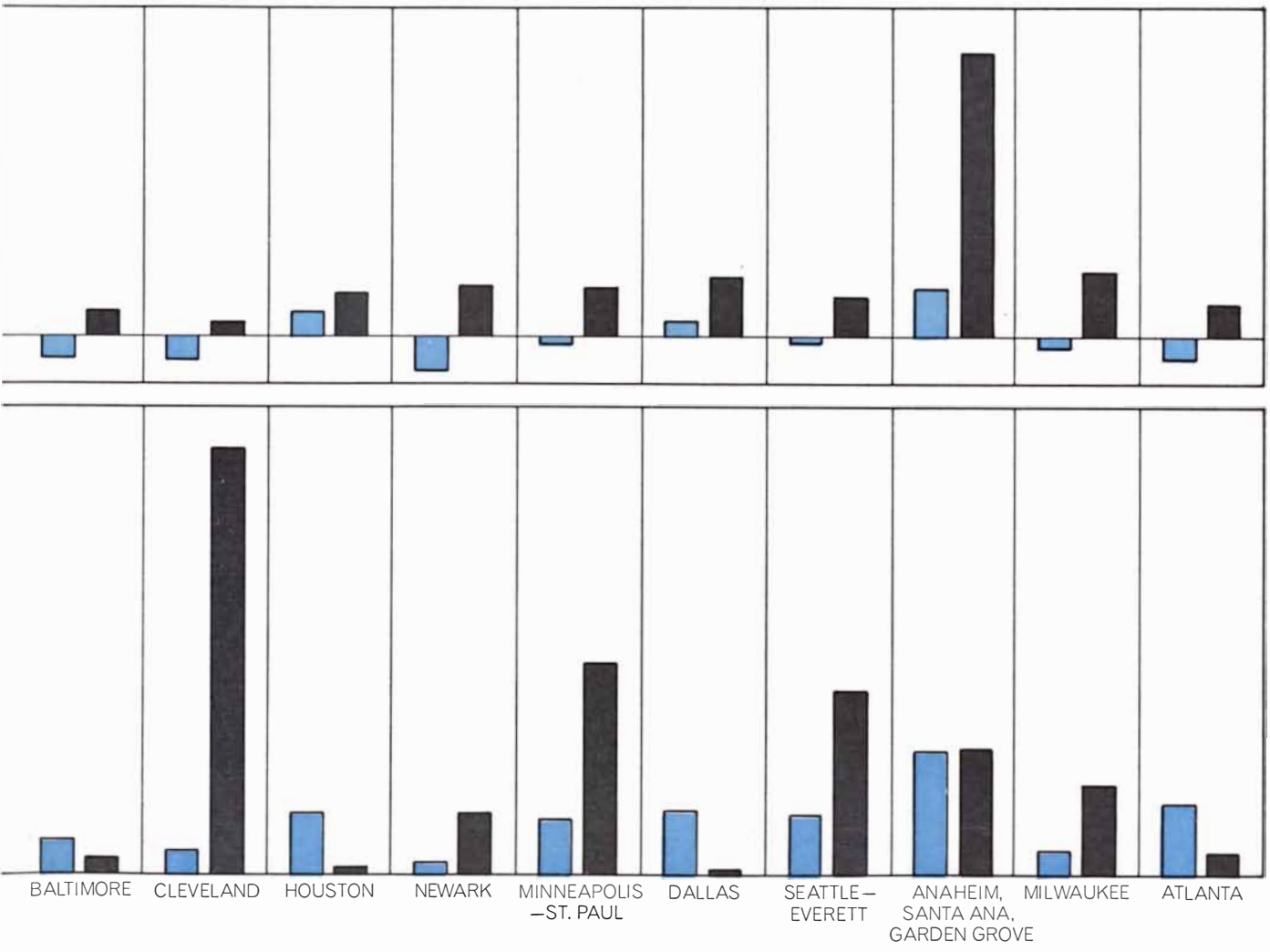
WHITE AND BLACK POPULATIONS of the 20 largest SMSA's showed a consistent pattern of change between 1960 and 1970. With a few exceptions the white population of the central city declined and the black population increased (*top row of bars*). In the sub-

Since young adults are chiefly responsible for the formation of new households, the large increase of this group has greatly stepped up the demand for housing and for the many products and services needed for the establishment of homes and families. In addition to the increase in the group under 25, during the 1960's there was a substantial rise (9.2 percent) in the number of people in the age span between 25 and 34. In the age groups between 35 and 64 most of the changes were less dramatic [*see bottom illustration on page 22*]. It is obvious, however, that the fluctuations in the various age levels of the working and consuming population have had and will continue to have profound effects on every sector of the economy, in terms of consumption patterns and the training, recruitment, employment and promotion of the labor force. To cite an example, during the 1950's, when there was a shortage of young men aged 20 to 29, we saw a change in the pattern of promotion of male employees and an increase in the employment and promotion of women.

At the polar ends of the age spectrum—children and "senior citizens"—the trends at the moment are following different courses. The number of children under five showed a sharp drop (15.5 percent) between 1960 and 1970, and in the age group from five to 14 the increase slowed to 14.9 percent (from 45.2 percent in the 1950's). The tidal wave of postwar babies that inundated the elementary schools in the 1950's and early 1960's has receded. This decline will continue into the 1970's, but the increase in the birthrate since 1968 indicates that there will again be a great rise in the elementary school population in the 1980's.

In contrast to the loss during the 1960's at the baby end, the number of citizens aged 65 and over continued to increase at a high rate, although the rate slowed from 34.7 percent in the 1950's to 21.1 percent in the 1960's. The senior citizens now number more than 20 million, about 10 percent of the total population. Their increase in numbers is forcing additional national attention to the problems and needs of the aged.

Clearly the changes in the population's age structure, and the lowering of the voting age to 18, foreshadow important shifts in political weight. Youth and age will gain at the expense of middle age. The reapportionment of Congress dictated by the 1970 census will bring about regional shifts, principally adding power to the West. California gains five seats in the House of Representatives, Florida three and Arizona, Colorado and Texas one each. On the other hand, New York loses two, Pennsylvania two and Alabama, Iowa, North Dakota, Ohio, Tennessee, West Virginia and Wisconsin one each.



urbs the white population increased and so did the black population (*bottom row*). In some SMSA's suburban black population rose by a huge percentage because it was building on a small base. In Cleveland suburbs black population increased by 452.8 percent.

The Induction of Interferon

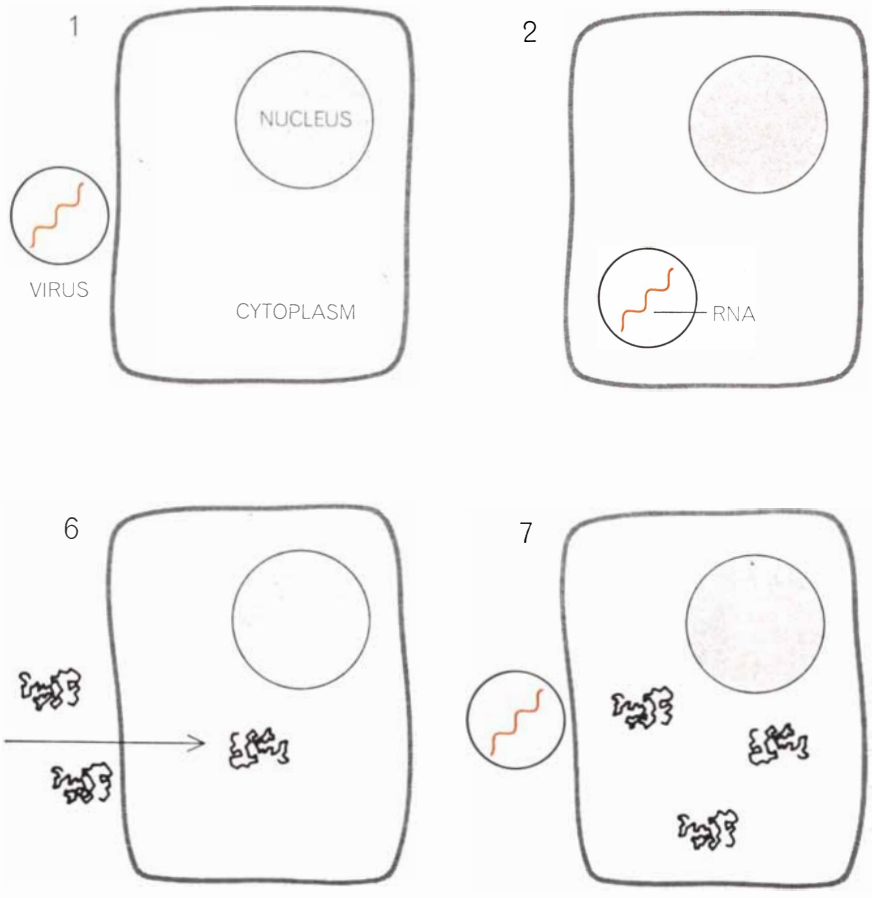
A synthetic RNA called poly I:C may provide broad-spectrum protection against virus diseases. It makes cells form interferon, the protein that defends other cells against infection by viruses

by Maurice R. Hilleman and Alfred A. Tytell

There are potentially three ways in which virus diseases can be controlled in man. To date the only clinically practical way has been to administer a vaccine: a preparation of killed or attenuated virus that stimulates the body to form antibodies against that virus. The specificity, or narrow spectrum, of vaccines is a limitation; it means that a different vaccine is required for each virus or strain of virus. Another possible method, and potentially one with broad-spectrum action, is chemotherapy. No safe and effective chemical agents have been discovered, however, that act against viruses in the way that sulfonamides or antibiotics act against large groups of bacteria. The third possibility is to rely on what is apparently the cell's own first line of defense against virus attack: interferon. Since 1957 it has been known that a cell infected by a virus produces a substance, interferon, that protects other cells against virus infection. Now an effective method has been found for mobilizing the body's interferon system and the way is open for the assessment of the method's safety and efficacy in clinical trials.

In 1935 Meredith Hoskins of the Rockefeller Foundation observed that infection of a monkey with an attenuated yellow-fever virus prevented death from a virulent yellow-fever virus administered at the same time. Two years later George M. Findlay and F. O. MacCallum of the Wellcome Research Institution in England found that monkeys infected with the virus of Rift Valley fever were protected against yellow-fever virus. Because it operated against a second virus given at the same time as the attenuated one and also against a different virus, this protective effect was clearly not related to known immune reactions involving the antigen-antibody response. Findlay and MacCallum called it "virus interference," and in the next few years many examples of interference between pairs of viruses were reported. No one found a practical way to adapt the phenomenon to the prevention of viral disease, however.

Then in 1957, at the National Institute for Medical Research in England, Alick Isaacs and Jean Lindenmann made the epochal discovery that the interfering action of viruses was brought about by a substance, which they called interferon, that was produced by the virus-infected cell and that protected uninfected cells against viral infection. Interferon could be destroyed by proteolytic enzymes and so it was presumed to be a protein; it was of relatively low molecular weight, and it was more resistant to



INTERFERON SYSTEM is believed to be induced and to provide protection against viral infection in two phases. The virus containing single-strand RNA attaches itself to the cell wall (*1*) and enters the cell (*2*). The outer coat of the virus is removed, leaving the single

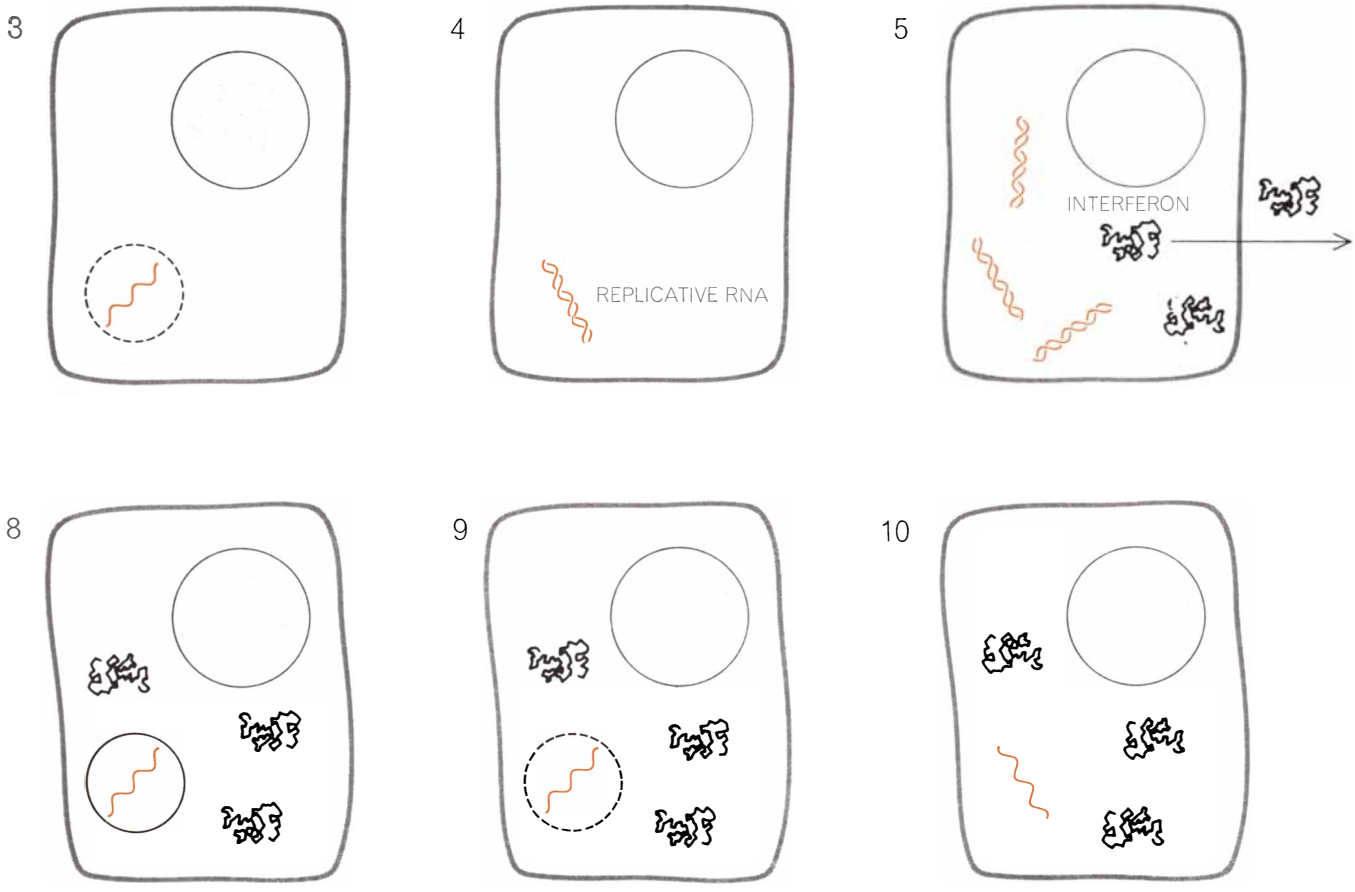
acid and alkali than most proteins. Interferon's antiviral activity was found to be very broad: the protein was effective against essentially all viruses. On the other hand, it had a narrow spectrum in a different sense: it was effective only in the animal species in which it was produced. Mouse interferon was active only in mouse cells, chick interferon in chick cells and so on [see "Interferon," by Alick Isaacs; SCIENTIFIC AMERICAN, May, 1961].

The immediate response to the discovery of interferon was the hope that the protein might be produced in cells, purified and administered to prevent and treat viral diseases. There was even the possibility that the interferon molecule might turn out to include a simple chemical group that was responsible by itself for the antiviral activity and that could be synthesized. The first necessary step was purification, a task that was undertaken by our group at the Merck Institute for Therapeutic Research. In 1963 we were able to report a significant degree of purification of interferon produced in embryonated hens' eggs that were infected with influenza virus. The refined interferon was extremely active against viruses in test-tube experiments; it was effective as an antiviral agent, for example, in smaller quantities than the usual antibiotics are effective when they are assayed against highly susceptible bacteria. It turned out that interferon's antiviral activity stemmed from the complex molecule as a whole rather than from any simple subgroup of the molecule. Our observations were confirmed and extended by Thomas C. Merigan, Jr., of the Stanford University School of Medicine and by Karl H. Fantes of Glaxo Laboratories Ltd., in England.

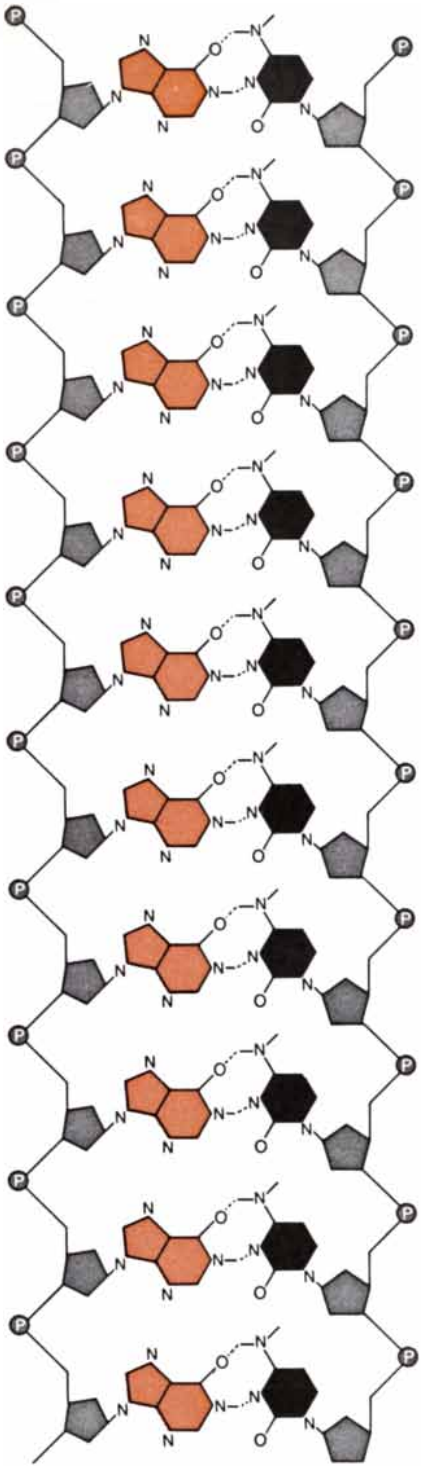
It is doubtful that absolutely pure interferon has yet been isolated. Moreover, there is no practical way to produce interferon in human cells that is safe for general use in man. Even if a safe way is found, it is not likely to be practical quantitatively. That is, the relation between the required dose and the possible yield of interferon by cells is so unfavorable at present as to seem to rule out the practicality of producing human interferon, purifying it and administering it to humans. Monto Ho and Bosko Postic of the University of Pittsburgh have arrived at a similar conclusion on the basis of extrapolation from a mouse-interferon system, but we should note that not all workers share this judgment.

In view of the limitations on the use of exogenous interferon in medicine, the obvious alternative was to find a suitable inducer of interferon: a substance that could be given a patient to cause his body to manufacture its own endogenous interferon. The search for inducers in a number of laboratories turned up many different kinds of substances that stimulated interferon production in animals. These included bacteria, parasites, viruses, polysaccharides, agents such as phytohemagglutinin that promote cell division, bacterial endotoxins, synthetic plastics and other substances. For various reasons none of these agents was suitable for clinical purposes.

Our group at the Merck Institute, consisting of the authors and A. Kirk Field, George P. Lampson and Marjorie M. Nemes, concentrated on an attempt to find and then imitate the means by which viruses naturally induce inter-



strand of RNA (*3*), which replicates to form a double strand (*4*). The presence of a foreign double-strand RNA causes the cell to synthesize interferon (*5*) but the RNA continues to replicate. The interferon leaves the cell and enters a new one (*6*). It cannot keep a virus from entering the new cell (*7, 8, 9*) but somehow serves to prevent the replication of the single-strand viral RNA (*10*).



POLY I:C is a synthetic polynucleotide, a laboratory analogue of RNA formed by the complexing of two homopolymers: polyriboinosinic acid (*left*) and polyribocytidylic acid (*right*). Each strand, like the strands of a natural RNA, consists of a backbone (*gray*) of alternating phosphate (*P*) and sugar (ribose) groups to which nitrogenous bases are attached as side groups. In these synthetic homopolymers the bases are all the same: the purine hypoxanthine (*color*) and the pyrimidine cytosine (*black*). The two single strands are linked by hydrogen bonds between the bases (*broken lines.*)

feron. We decided on two approaches. The first was to fractionate a number of virions, or virus particles, and to try to find a component or components responsible for inducing interferon. The second was to isolate the active principle from helenine, an extract of the fungus *Penicillium funiculosum* that Richard E. Shope and his colleagues at the Rockefeller Institute for Medical Research had shown would induce interferon in cell culture and induce resistance to viral infection in animals. The event that happened to put us on the right track was the finding by Werner Braun and Masayasu Nakano of Rutgers University that antibody responses in animals were enhanced by certain synthetic polynucleotides. These are analogues of the nucleic acids DNA or RNA that are made in the laboratory by combining nucleotides (the subunits of nucleic acids) in arbitrary ways instead of in the naturally occurring arrangements that encode hereditary information in living cells.

We wondered whether such polynucleotides might also induce interferon, and we tested a number of synthetic RNA's. One of them was extremely active: in small amounts (micrograms) it induced interferon in rabbits and made cell cultures and animals resistant to viral infection. The active substance was poly I:C, consisting of paired homopolymer strands of polyriboinosinic acid (poly I) and polyribocytidylic acid (poly C) [*see illustration at left*]. Poly A:U, a complex based on adenylic and uridylic acids, was also active in cell cultures but was much less active in animals. The individual single-strand homopolymers were not active.

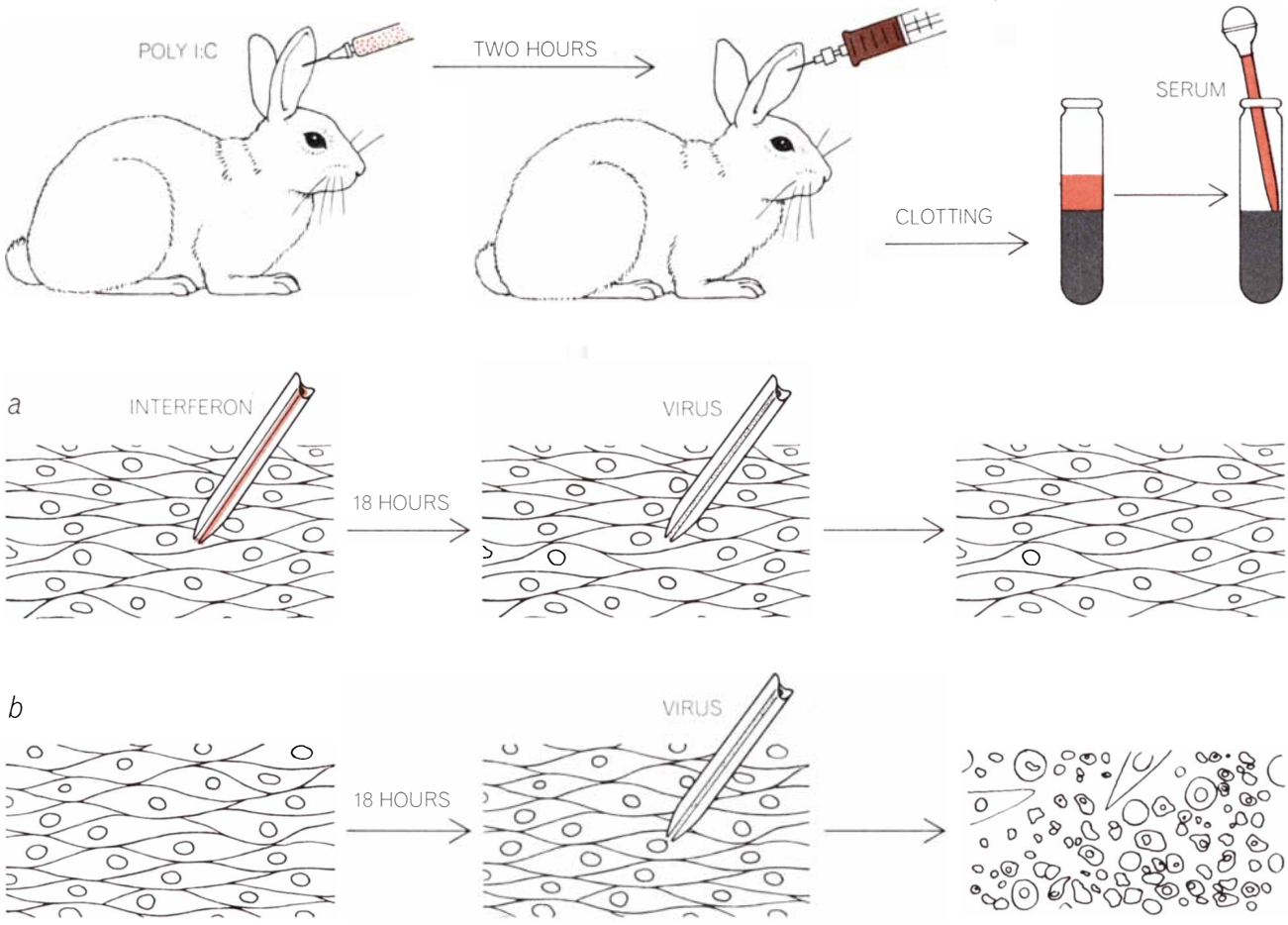
Meanwhile the work on helenine was well on its way, and soon after we had established the interferon-inducing property of poly I:C we were able to report that the active principle in helenine was also a double-strand RNA. We suggested that the mold from which the helenine was derived must have been infected with a virus, which had presumably supplied the double-strand RNA. Our suspicion was confirmed when Walter J. Kleinschmidt and his associates at Eli Lilly and Company demonstrated the presence of a fungal virus in statalon, an extract of another penicillium species and, like helenine, an inducer of interferon. Later G. T. Banks and his co-workers at the Imperial College of Science and Technology in London isolated a fungal virus from *P. funiculosum* itself.

We thereupon undertook an extensive screening program, testing polynucleotides of varying origins for interferon induction. We did this by injecting the polynucleotides intravenously in rabbits, withdrawing blood and testing the ability of the blood serum to protect a cell culture against virus infection [*see illustration on opposite page*]. A large number of double-strand RNA's, of natural and synthetic origin, were found to be active inducers. Single-strand RNA's were inactive, as was double-strand DNA [*see illustration on page 30*].

The question arose: Where does the double-strand RNA come from in the case of an ordinary virus infection? Whereas the genetic material is DNA in some viruses, in bacteria and in higher plants and animals, it is RNA in many viruses. In most RNA viruses it is the single-strand form of RNA that is present in the virion, not the double-strand form. We realized that such viruses therefore must not contain an interferon inducer to begin with; the interferon-inducing activity must arise only when the double-strand replicative form of RNA is produced in a virus-infected cell. It has since been demonstrated by immunofluorescence and other techniques that the double-strand form does indeed appear in the normal course of virus replication. The same thing is true of the synthetic polynucleotides: the homopolymers become active only when they are complexed to form a double strand. (Under certain conditions poly I or poly C alone may show weak interferon-inducing capacity, but such activity is trivial compared with that of poly I:C and is of no practical importance.)

The discovery that interferon is induced by double-strand RNA provided a breakthrough in the understanding of interferon induction by viruses as well as a possible new approach to the control of viral infections. The exact mechanism for interferon induction and utilization has yet to be fully elucidated, but in our working hypothesis the interferon system is conceived of as operating in two distinct phases [*see illustration on preceding two pages*]. In the first phase a single-strand RNA virus penetrates the cell. Its outer coat of protein is removed, releasing the RNA, which sets about producing the complementary strand of RNA that is essential to its replication. The double-strand RNA that is formed is not normal to the cell, and so in some undisclosed way it alerts the cell to synthesize interferon.

The interferon itself is not active against viruses. It leaves the cell in which it was produced and, in the second phase, enters uninfected cells,



TEST FOR INDUCTION of interferon is conducted by injecting poly I:C (or any other potential inducing substance) intravenously in rabbits. After two hours the rabbits are bled and the blood is allowed to clot. Interferon induced in the rabbit is contained in the blood serum, which is tested at various dilutions by a cell-culture assay (*a*): the culture is treated with the diluted serum, incubated for 18 hours, washed and then infected with the virus of vesicular stomatitis. If the interferon level is adequate, the cells are unharmed. Meanwhile an untreated control culture is exposed to the virus in the same way (*b*). Its cells are killed by the virus.

where it exerts antiviral activity. It may function as a "derepressor" that causes the new cell to produce a new protein that prevents the formation of viral nucleic acid when the cell is later attacked by a virus. Alternatively, it is possible that the interferon itself is altered by the new protein, perhaps enzymatically, to form an active antiviral molecule. There is only indirect evidence for the presence of this "new protein," and its precise mechanism of action in preventing viral replication is not known. The few viruses (such as the reoviruses) that contain double-strand RNA induce interferon immediately on cell penetration without the need for replication. At least one DNA virus, vaccinia, for some reason produces a double-strand RNA, which is active in inducing interferon; how most DNA viruses stimulate interferon induction is not yet known.

The usefulness of an interferon inducer cannot be predicted on the basis of a demonstration of induction alone. We went on to more meaningful tests that showed actual protection against cell destruction by a virus and prevention of disease and death in animals infected with viruses. In cell-culture trials of protective efficacy the inducer being tested is added to the culture, and about three hours later a measured challenge dose of virus is administered. Activity is demonstrated by the drug's ability to prevent viral damage to the sheet of cells in the culture. All the double-strand RNA's found to be active in inducing interferon in rabbits were also active, in microgram amounts, in preventing cell death by vesicular stomatitis virus in a wide variety of cells of avian and mammalian origin, including human cells. When the same inducers were tested in mice, all of them prevented disease and death in animals infected with the pneumonia virus of mice (PVM).

Because of the ready availability of poly I:C and its considerable promise in these early tests, further studies in our laboratories were centered on poly I:C alone. Following a common practice for appraising the potential usefulness of a drug, we established a "model system" in mice, infecting them with a standardized and ordinarily fatal dose of various viruses so that the prophylactic and therapeutic effect of poly I:C could be measured. In a typical experiment poly I.C was instilled into the nostrils of mice before or after nasal inoculation of a standardized lethal amount of PVM virus [*see illustration on page 31*]. Nearly all untreated control mice died but all animals given poly I:C before the virus survived. So did nearly all mice that were given poly I:C two days after the virus. In other words, the poly I:C had a therapeutic as well as a prophylactic effect. Apparently poly I:C given soon enough after infection induces sufficient interferon to protect a large proportion of the cells; in effect the interferon outraces the virus, and the animal survives. Starting treatment four days after virus inoculation was too late and all the animals succumbed.

Experiments along these lines have re-

SOURCE	DOSE (MICROGRAMS)	INTERFERON TITER
DOUBLE-STRAND RNA		
POLY I:C	2	640
POLY A:U	25	20
PENICILLIUM FUNICULOSUM	8	640
REOVIRUS 3	8	640
MS2 COLIPHAGE (REPLICATIVE)	8	160
MU9 MUTANT COLIPHAGE (REPLICATIVE)	2	40
RICE DWARF VIRUS	20	640
CYTOPLASMIC POLYHEDROSIS VIRUS	22	160
SINGLE-STRAND RNA		
POLY I	25	0
POLY C	20	0
MS2 COLIPHAGE	8	0
ESCHERICHIA COLI	100	0
NEWCASTLE DISEASE VIRUS	10	0
INFLUENZA VIRUS	10	0
TOBACCO MOSAIC VIRUS	40	0
YEAST RIBOSOME	1,000	0
YEAST SOLUBLE RNA	200	0
DOUBLE-STRAND DNA		
CALF THYMUS TISSUE	200	0

SCREENING OF NUCLEIC ACIDS for interferon induction yields these results. Double-strand RNA's, both natural and synthetic, were effective in microgram amounts. The titers (*right*) are the dilutions of the rabbit serum that were effective in protecting 50 percent of the cell cultures tested. Single-strand RNA's and double-strand DNA were ineffective.

vealed a strong protective effect of poly I:C in mice against the encephalomyocarditis virus, vaccinia virus and parainfluenza viruses. Minor protective effects were obtained in animals against the oncogenic (cancer-producing) SV-40 virus, adenovirus and Friend leukemia virus. This ability to protect against both DNA and RNA viruses and against oncogenic as well as cell-killing viruses emphasized the broad-spectrum antiviral activity of poly I:C. Workers in other laboratories have greatly extended the studies with poly I:C. They have shown poly I:C to be effective against severe herpes simplex eye infections in rabbits, Semliki Forest virus (an insect-borne virus) infections in mice, rabies in rabbits and certain respiratory virus infections such as influenza. Moreover, marked antitumor effects have been observed for poly I:C against several transplant tumors in mice, particularly the L1210 ascites tumor. Activity against experimental malaria and other parasitic diseases has also been recorded. The mechanisms of the effects against tumors and malaria are not known and may well involve properties other than interferon induction. Poly I:C, as noted above, is known to stimulate the ordinary immune mechanisms, including cell-mediated immunity, and this is believed to be important in defense against tumors.

Before any drug can be used in clinical trials in man, its safety must be assessed in animal tests, and so we carried out trials in mice, rats, dogs and monkeys with different doses and routes of injection. When the drug was given intravenously, destructive side effects were noted in the small blood vessels, liver and blood-forming organs of dogs; these effects were minor or absent following administration by the subcutaneous or respiratory routes. The toxic effects were far less in mice and rats than in dogs, and they were essentially absent in monkeys. The tests indicated that there would be a satisfactory margin of safety for human subjects treated with doses of poly I:C large enough to stimulate interferon.

Other tests for safety showed that poly I:C did not sensitize guinea pigs to anaphylactic reactions. The drug was not carcinogenic when injected into newborn animals, which were observed for 18 months after injection. It did not cause normal human cells to acquire cancer-like properties when they were assayed in cell culture or tested by implantation in the cheek pouch of hamsters. Finally, it did not bring about changes in the chromosomes of normal human cells grown in cell culture.

When the time came for testing in man we decided, because of the remarkable effect shown by poly I:C against certain cancer-producing viruses in animals and the striking beneficial action of the drug against certain transplanted tumors, to conduct trials in patients with cancer. Accordingly we carried out a study in cooperation with Charles W. Young and Irwin H. Krakoff of the Memorial Hospital and Sloan-Kettering Institute for Cancer Research in New York. Twenty patients with advanced cancer of various types were given poly I:C intravenously—from two micrograms to four milligrams per kilogram of body weight. Serum samples for interferon assay were taken prior to injection and up to 72 hours thereafter. Fourteen of the 20 patients responded with the production of interferon; six failed to respond. The induction of interferon and the magnitude and duration of the interferon response did not appear to be related strongly to the size of the poly I:C dose. Interferon appeared in the blood as early as two hours after the poly I:C injection and persisted in some patients for as long as 72 hours. Peak production of interferon was usually attained between 12 and 48 hours after the administration of the drug. Unfortunately there was no significant beneficial effect against the tumors in these patients.

Two significant problems that relate to the clinical application of interferon inducers are the relatively short duration of resistance following the administration of inducer (usually about six days in animals) and the failure to respond to a second injection of inducer within a certain period after successful induction. Fortunately this state of hyporeactivity seems to disappear at about the same time as the resistance to viral infection is lost—just when one needs to give a new dose of inducer. In the studies with human patients successful reinductions of interferon were obtained when the poly I:C injections were given six or seven days apart in two subjects and even within three days in one patient. Another patient showed lack of response to interferon induction only after the fourth of a series of daily doses of poly I:C, and this patient became fully responsive again after a 10-day period. It appears, then, that although resistance to induc-

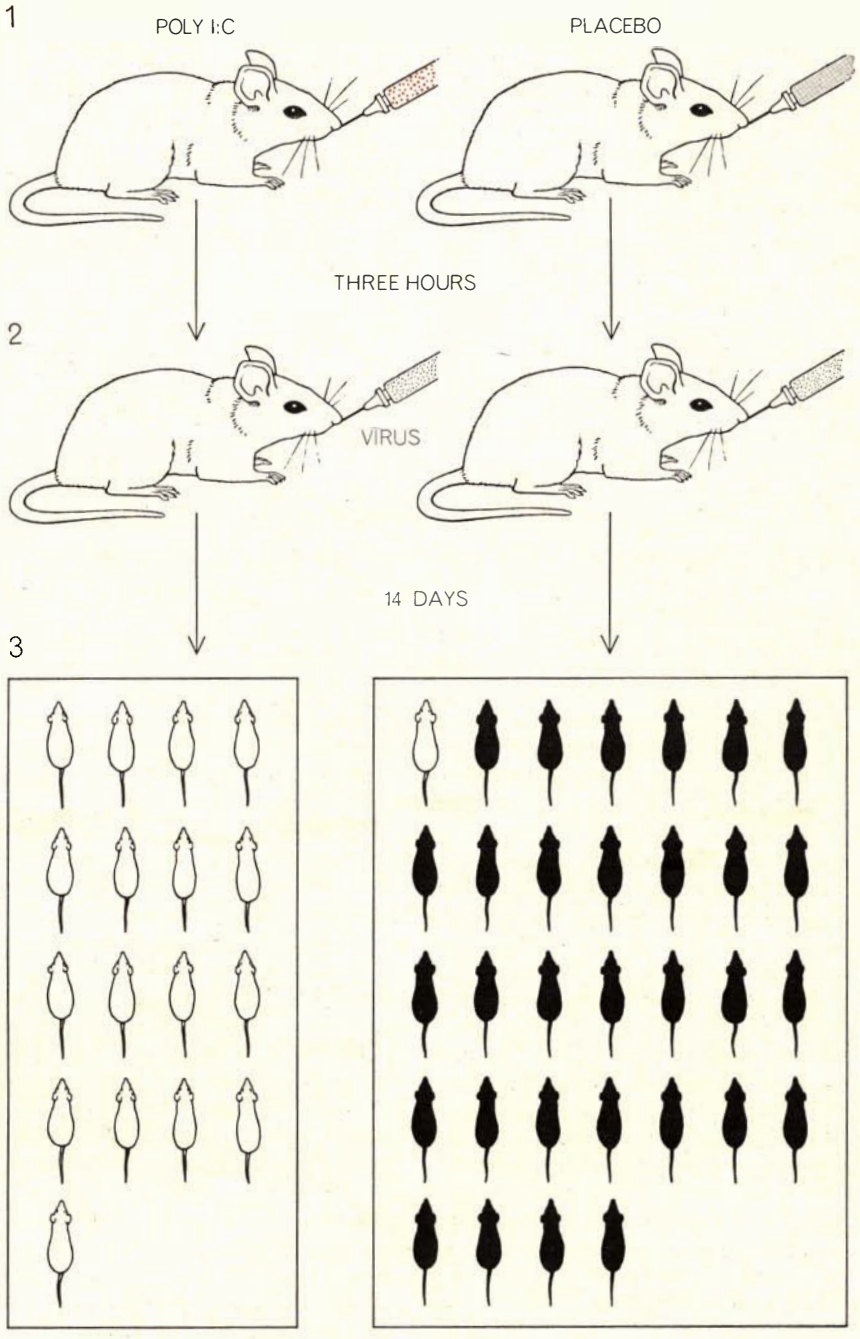
tion of interferon does occur in man, its duration is relatively short. It may be that the administration at frequent intervals of small but therapeutically adequate doses of poly I:C will allow the continuing induction of interferon and thus maintain continuous resistance to viral infection.

Clinical reactions to poly I:C in man were remarkably few. The most consistent reaction to intravenous injection of poly I:C was fever. The temperature elevation ranged as high as seven degrees Fahrenheit, reaching a peak from six to 15 hours after injection. The fever was generally accompanied by interferon induction, although the responses did occur independently in some patients. The amount of fever was not correlated with the poly I:C dosage. Extensive laboratory observations revealed no disturbance of liver, kidney or bone-marrow functions. No effects on the blood-clotting mechanism were observed. In sum, there were no clinical symptoms that placed a limit on the doses we administered.

We were concerned about the possibility of an antibody response to the poly I:C, but tests showed that the patients did not develop antibody against either poly I:C or DNA. One related possibility remains to be investigated. It has been noted that the development of an autoimmune disease that resembles systemic lupus in man, and that appears spontaneously in a particular breed of mice, may be accelerated by multiple injections of poly I:C. The dose requirements and route of administration of poly I:C in these mice are now being studied in order to establish another safety limit for the drug in human subjects susceptible to autoimmune disease.

Most viral infections in human beings are self-limited, suggesting that natural defense mechanisms are operative in recovery from viral diseases. Antibody and cell-mediated immunity, the ordinary specific immune mechanisms, seem to come into effect rather late in the infectious process; they may be of greater importance in preventing later reinfection by the same virus or in removing virus-infected cells than in the early recovery of the host from a particular virus infection. Interferon is produced early in the process of infection and there is substantial reason to believe that it is at least one of the important factors involved in the recovery from viral disease.

The administration of exogenously produced interferon has shown very limited promise up to the present time because of limitations in source, safety and cost. The more practical approach seems to lie in the administration of interferon inducers. The fact that beneficial secondary effects, such as stimulation of the ordinary immune mechanisms, may also occur is an added reason for utilizing polynucleotides. Poly I:C shows considerable promise as a means for exploiting the interferon mechanism. The remarkably low toxicity for human beings and ready availability of this synthetic product open the door to large-scale exploratory investigations. After some final tests to rule out the danger of autoimmune disorders, poly I:C will be ready for cautious trials in human subjects for preventing infections, such as the common cold, that are caused by a variety of viruses.



PROTECTIVE EFFECT of poly I:C was tested in laboratory animals. Poly I:C was instilled into the nostrils of some mice (*left*) and a control group (*right*) was given a placebo. After three hours both groups of mice were infected with a lethal dose of pneumonia virus of mice (PVM). The animals were then observed for 14 days. All 17 of the treated mice survived but only one of 32 mice in the control group survived. The drug was also effective when it was given two days after the virus; the survival rates were 15 of 15 treated animals and one of 25 controls. Given four days after the virus, poly I:C was ineffective.

A New Class of Diode Lasers

Minute sandwich-like "heterostructures" composed of two or more different semiconducting materials show great promise as cheap, efficient carrier-wave generators for use in mass communications

by Morton B. Panish and Izuo Hayashi

Soon after the invention of the laser a decade or so ago it was recognized that one of the most important potential applications of this new tool would be as a generator of carrier waves for a long-distance optical communications system. In principle a single laser beam has many thousands of times the signal-carrying capacity of any other transmission medium then or now in service. Nevertheless, the large-scale commercial application of lasers to communications has not yet occurred, and it may not occur for quite a while. This situation is of course partly a result of the fact that such an application requires the development of a vast and complex associated technology. In addition, however, there has been the fact that most of the lasers available so far have been too large, too delicate, too costly and too complicated to operate for this approach to compete economically with present mass-communications techniques. The one exception to many of these drawbacks is represented by the semiconductor diode laser.

The first success in obtaining laser light from semiconductor diodes was announced almost simultaneously in 1962 by groups working at the General Electric Company, the International Business Machines Corporation and the Lincoln Laboratory of the Massachusetts Institute of Technology. These three groups achieved success with diodes made from the semiconducting compound gallium arsenide. The first announcement was greeted with considerable enthusiasm by workers in the field, because the new lasers, unlike their larger predecessors, could be made very small and were quite rugged, cheap and simple to operate. They were expected to work, in principle at least, at a very high efficiency in comparison with other lasers. Moreover, they required an operating voltage of only 1.5 to two volts! It is in fact tempting, where certain applications are concerned, to compare the more conventional large lasers to the diode laser in much the same way that one would compare the vacuum tube to the transistor.

The early diode lasers did, however, suffer from a major disadvantage. They required such a high current density to operate that, in order to protect them against damage resulting from overheating, they could not be run continuously except at very low temperatures. It was possible to operate them in very short pulses (approximately a millionth of a second) at room temperature, but pulsed operation severely limits the usefulness of such devices in communications as well as in other applications.

In the discussion that follows, the properties of a new class of semiconductor diode lasers that are capable of being operated continuously at room temperature are described. The evolution of this new class of lasers—called heterostructure lasers—brings closer to reality the achievement of a competitive optical transmission system for telephone signals and other communications. Some of the ideas involved in the development of the heterostructure lasers were first presented in 1963 by Herbert Kroemer, who is now at the University of Colorado. In addition to our work at the Bell Telephone Laboratories important contributions have been made by Zh. I. Alferov and his co-workers at the Ioffe Institute in Leningrad and by H. Kressel and his co-workers at the RCA Corporation.

The diode laser is a refinement of the electroluminescent diode [see "Light-emitting Semiconductors," by Frederick F. Morehead, Jr.; SCIENTIFIC AMERICAN, May, 1967]. Such a device produces light in a manner quite different from the way the familiar incandescent lamp does. In the incandescent lamp the electrically heated tungsten wire emits light over a single band of wavelengths thousands of angstroms wide as a result of the fact that some of the thermal-excitation energy of the atoms is transferred to photons with a wide range of energies characteristic of the temperature of the emitting wire. Luminescent diodes produce a narrow band of radiation, usually less than several hundred angstroms wide, at a peak wavelength characteristic of the material that is luminescing. Luminescent diodes require no increase in temperature; in fact, their efficiency almost invariably increases when their temperature is decreased. The electroluminescence of these devices is a direct result of the semiconducting properties of the materials of which they are made.

Semiconductor crystals, like any other crystal, are materials in which the individual atoms are bound to each other in a repetitive three-dimensional array. Such an array is called a crystal lattice. The bonding of the individual atoms in the lattice involves primarily the interaction of the outer electrons of the constituent atoms. These electrons, in addition to binding the crystal together, give the material its characteristic optical properties. In an isolated atom (for example an atom in a gas) each electron may have any one of a number of distinct energies. These energies are called energy levels, and an electron at a particular energy is described as occupying a particular energy level. What level an electron occupies depends on its position in the electronic structure of the atom, on the temperature, on the degree of external excitation and on other factors.

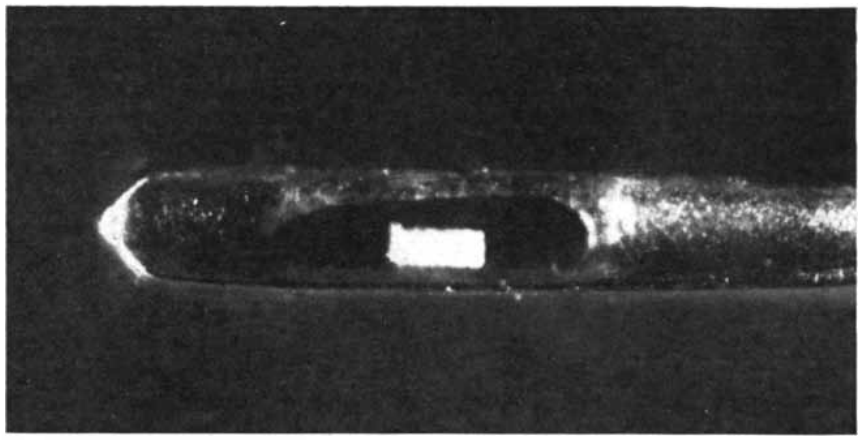
The energy-level structure of an isolated atom can in fact be described as a kind of ladder [*see upper illustration on page 34*]. For an atom in a low state of excitation the upper rungs (or levels) are

all empty. With the addition of energy in discrete quantities electrons can be made to jump from one rung to a higher one, so that at higher states of excitation of the atom some of the lower rungs are empty and some of the upper ones are occupied. Luminescence results from the transfer of the excitation energy from the electron to a discrete quantum of light, or photon, when the electron goes from a higher to a lower level. The photons can be thought of as a wave with a given wavelength or as particles with a given energy. Either term describes the color of the light if it is in the visible part of the spectrum.

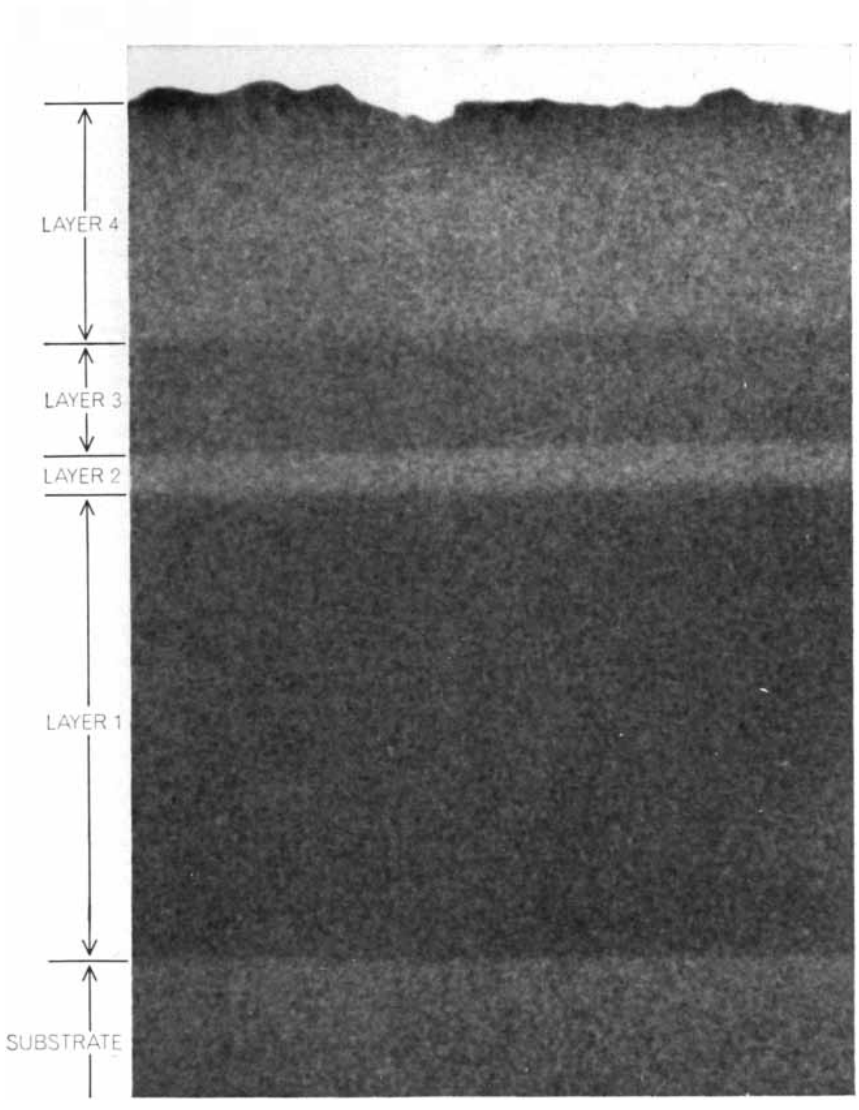
In crystals the individual energy levels of the outer electrons of the isolated atoms have merged to form almost continuous bands of very closely spaced energy levels. These levels in the bands can be regarded as a continuum of energy states, each of which can be occupied by an electron. The number of states in a band is equal to the number of atoms in the crystal multiplied by a small whole number (usually less than 10). One describes the bands as empty, full or partly full depending on whether none of the states is occupied by electrons, all are occupied by electrons or only some of them are occupied by electrons. The bands are separated from each other by "forbidden" energy regions. Except under special circumstances electrons in the crystal cannot have energies in the forbidden range, which is called an energy gap or a band gap.

It is sufficient for the purposes of this discussion to regard a typical semiconductor as having two such energy bands. The higher-energy one, called the conduction band, is separated from the lower one, called the valence band, by a forbidden energy gap with a width designated E_g. The size of E_g has a controlling effect on the optical properties of the semiconductor. In general the semiconductor will be transparent to light with photon energies smaller than E_g and will absorb light with photon energies greater than E_g. Light with photon energies near E_g is absorbed to a greater or lesser extent depending on many factors but particularly on the amounts of various foreign atoms present as "dopants" in the crystal.

The conduction band contains almost no electrons unless the crystal contains "donor" impurities, which contribute them. The valence band is nearly filled with electrons. It has a few empty states, called holes, which act as though they were positively charged particles. These imaginary particles result primarily from the presence of foreign "acceptor" atoms,

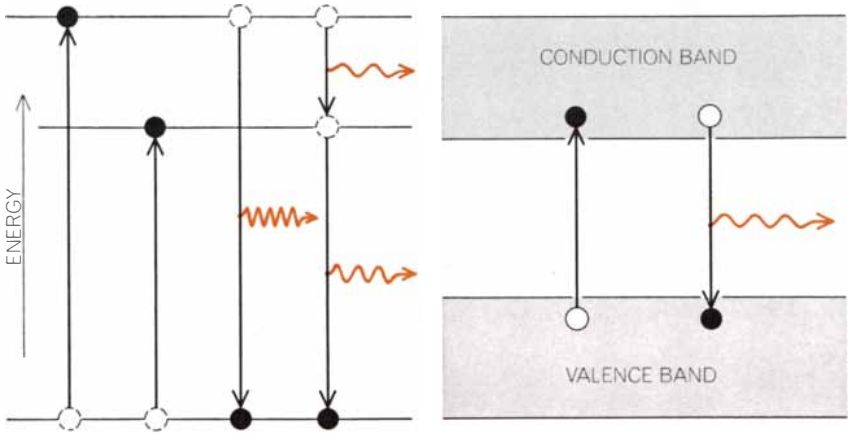


NEW LASER designed by the authors at the Bell Telephone Laboratories is shown here in the eye of a needle to accentuate its extremely small size. The device, called a double-heterostructure diode laser, is the first semiconductor laser capable of operating continuously at room temperature. Low-voltage leads are normally connected to two opposite faces.

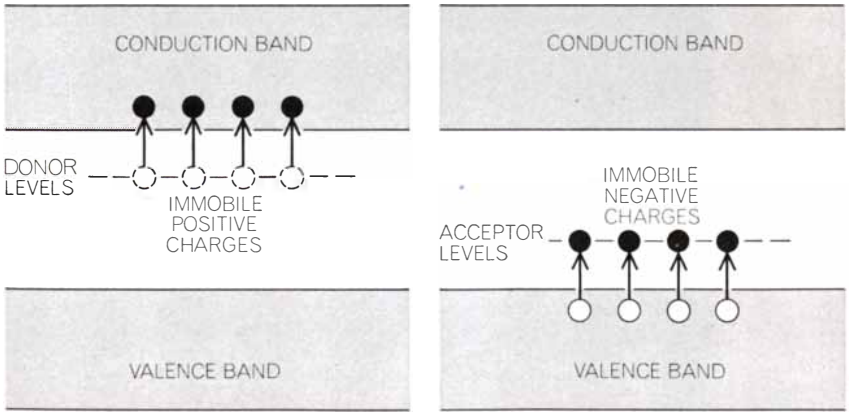


CROSS SECTION OF NEW LASER appears in this scanning electron micrograph of a cleaved end of the device. The particular double-heterostructure design shown is essentially a stack of very thin impurity-"doped" semiconducting layers grown on a gallium arsenide substrate. Layer *1* is tin-doped aluminum gallium arsenide. Layer *2*, the active lasing region, is silicon- and zinc-doped gallium arsenide. Layer *3* is zinc-doped aluminum gallium arsenide. Layer *4*, which serves as a contact layer, is germanium-doped gallium arsenide.

which have accepted some electrons from the valence band, leaving the empty states, or holes, behind. The holes in the valence band and the electrons in the conduction band are the only current-carrying entities present in the semiconductor. If there are more donors than acceptors, current in the material is carried by conduction-band electrons (negatively charged) and the material is designated an *n*-type semiconductor. If there are more acceptors than donors, current is carried by holes (positively charged) and the material is designated a *p*-type semiconductor [*see lower illustration on this page*].



ALLOWED ENERGY STATES of an electron in an isolated atom (*left*) and in a semiconducting solid (*right*) are compared in these two diagrams. In the atom an electron (*black dot*) may occupy one of a number of discrete energy levels depending on its degree of excitation. When the electron absorbs energy, it moves to a higher level (*upward arrow*); when it loses energy, it falls to a lower level (*downward arrow*) and in the process emits a photon of light whose wavelength is inversely proportional to the energy-level transition (*colored arrows*). This phenomenon is called luminescence. In a crystalline solid, on the other hand, the energy levels merge to form bands, each of which contains approximately the same number of levels as there are atoms in the crystal. These bands are separated from each other by a forbidden-energy gap (E_g) within which electrons cannot exist in the crystal. A pure semiconductor has a pair of adjacent energy bands; the lower one, called the valence band, is usually filled with electrons, whereas the upper one, called the conduction band, is usually empty. In a semiconducting crystal electric current is carried only by carriers in partially filled bands, such as the one that results when an electron is excited from the valence band to the conduction band by absorbing the band-gap energy either thermally or by some other means of excitation. This leaves an unoccupied energy level, or "hole" (*open circle*), in the valence band. Both the negatively charged excited electrons in the conduction band and the positively charged holes in the valence band can act as current-carriers. The electrical conductivity of a semiconductor is normally rather low, because only a comparatively few carriers are created thermally. Luminescence occurs in a semiconductor when an excited electron in the conduction band "recombines" with a mobile hole in the valence band, releasing its excess energy in the form of a photon of light.

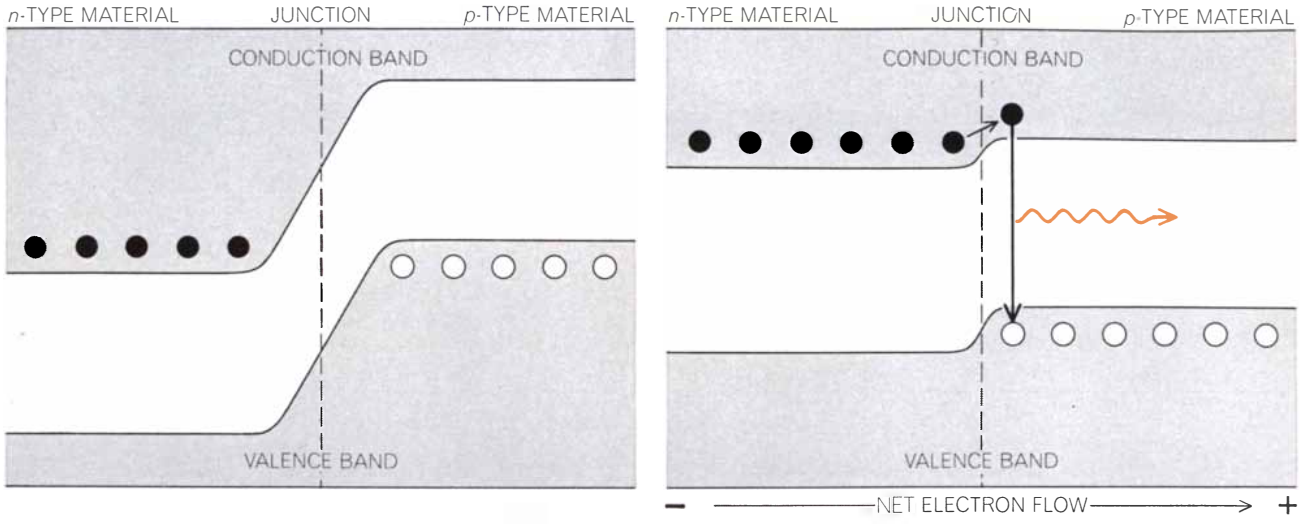


IMPURITY ATOMS with energy levels in the band-gap region of a semiconducting material can be used to improve the electrical conductivity of the semiconductor. Donor atoms (*left*) contribute an electron to the conduction band, creating immobile positive charges at the donor levels. Acceptor atoms (*right*) accept an electron from the valence band, making the electron immobile but leaving a mobile hole behind in the valence band.

It is possible to create pairs of electrons and holes in *n*-type or *p*-type material in several ways. These extra carriers are present in a "nonequilibrium" way, and the crystal tends to eliminate them. This elimination of the injected carriers involves the loss of the excitation energy by the excited electron. Since the excitation of the electron from the valence band to the conduction band means giving it the energy E_g, this is the amount of energy it can lose. It then "reoccupies" the injected hole in the valence band in a process called hole-electron recombination. Among the ways in which this energy can be lost is the emission of a photon (luminescence). That is the process with which we are primarily concerned. It occurs quite efficiently in semiconductors such as gallium arsenide.

When part of a semiconductor crystal is *n*-type and part is *p*-type, the boundary between the two regions is called a *p-n* junction. Such junctions form the basis of transistor action and also of the action of luminescent diodes. Since a *p-n* junction is a boundary that has mobile negative charges (electrons) on one side and mobile positive charges (holes) on the other, it might be expected that by ordinary thermal motion these charges would move toward the junction, where they would recombine, causing a net current to flow in the diode. Actually this process does start to happen, but it stops after a small number of holes and electrons have reached the junction and recombined, because the mobile holes and electrons have left immobile positively charged and negatively charged impurity centers behind. The electrostatic field due to the immobile charges opposes the motion and stops current flow. The result is a bending of the valence and conduction bands [*see top illustration on opposite page*].

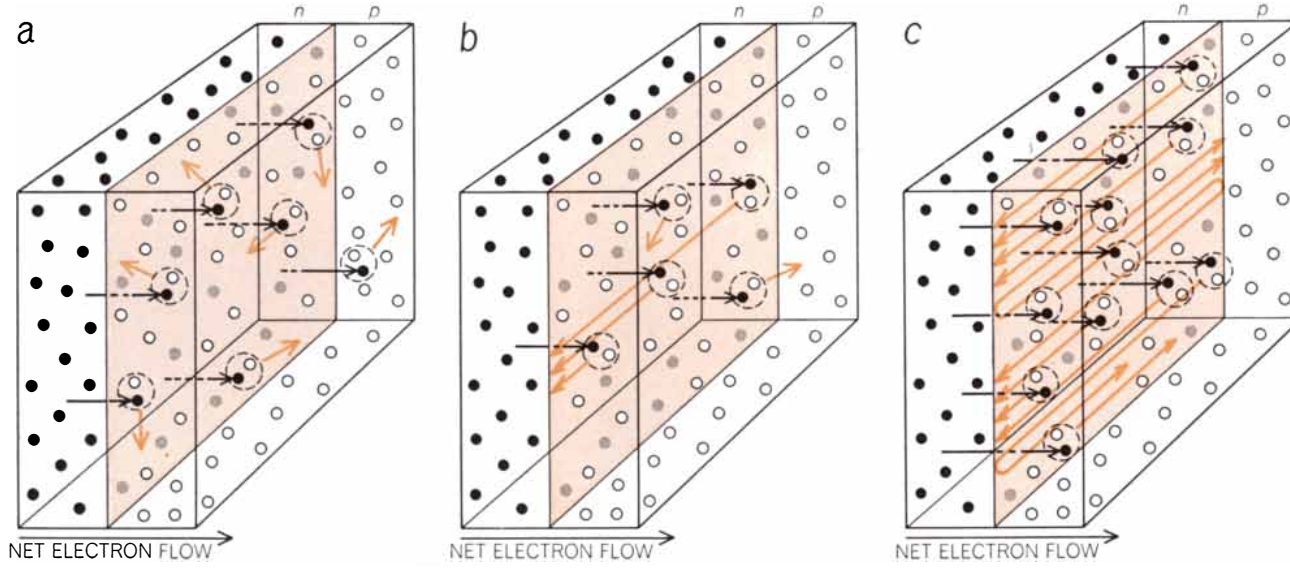
The bands can be unbent and a current can be induced to flow across the *p-n* junction by applying an external voltage to the crystal in such a way as to oppose the internally induced field. This is called "forward biasing" the diode. When the forward-bias voltage is applied to the diode, electrons are caused to flow into the *n* side from the external circuit and out of the *p* side into the external circuit. (To say that electrons flow out of the *p* side to the external circuit is exactly the same as saying that holes flow into the *p* side from the external circuit.) When the *p-n* junction is sufficiently forward-biased, electrons flow across



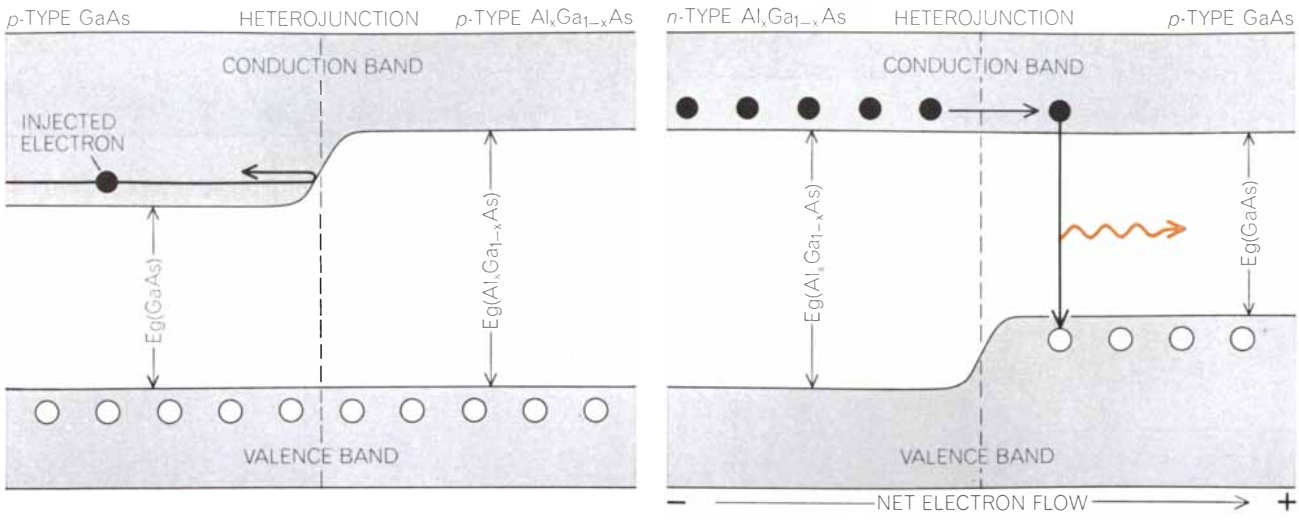
***P-N* JUNCTION** exists between one region of a semiconductor in which negatively charged electrons are the predominant current-carriers (*n*-type material) and another region in which positively charged holes predominate (*p*-type material). The induced electrostatic field bends the edges of the valence band and the conduction band at the junction, producing potential barriers that prevent electrons and holes from crossing the junction (*left*). By applying an external voltage of about E_g (a procedure known as forward-biasing the junction) one can unbend the bands and reduce the potential barriers to current flow (*right*). For gallium arsenide with appropriate impurities most of the current is carried across the junction by electrons that are injected by the external source into the *p*-type region. In that region the electrons are minority carriers and not at equilibrium. Hence they readily recombine with holes, releasing their energy in the form of light. Donor and acceptor impurity levels have been omitted for simplicity.

the *p*-*n* junction from the *n*-side conduction band to the *p* side, a process called electron injection. In many semiconductors holes from the *p*-side valence band cross the *p*-*n* junction into the *n* side of the junction, a process called hole injection. Usually one of these two processes predominates. For the semiconductor of interest here, gallium arsenide, the predominant process is electron injection.

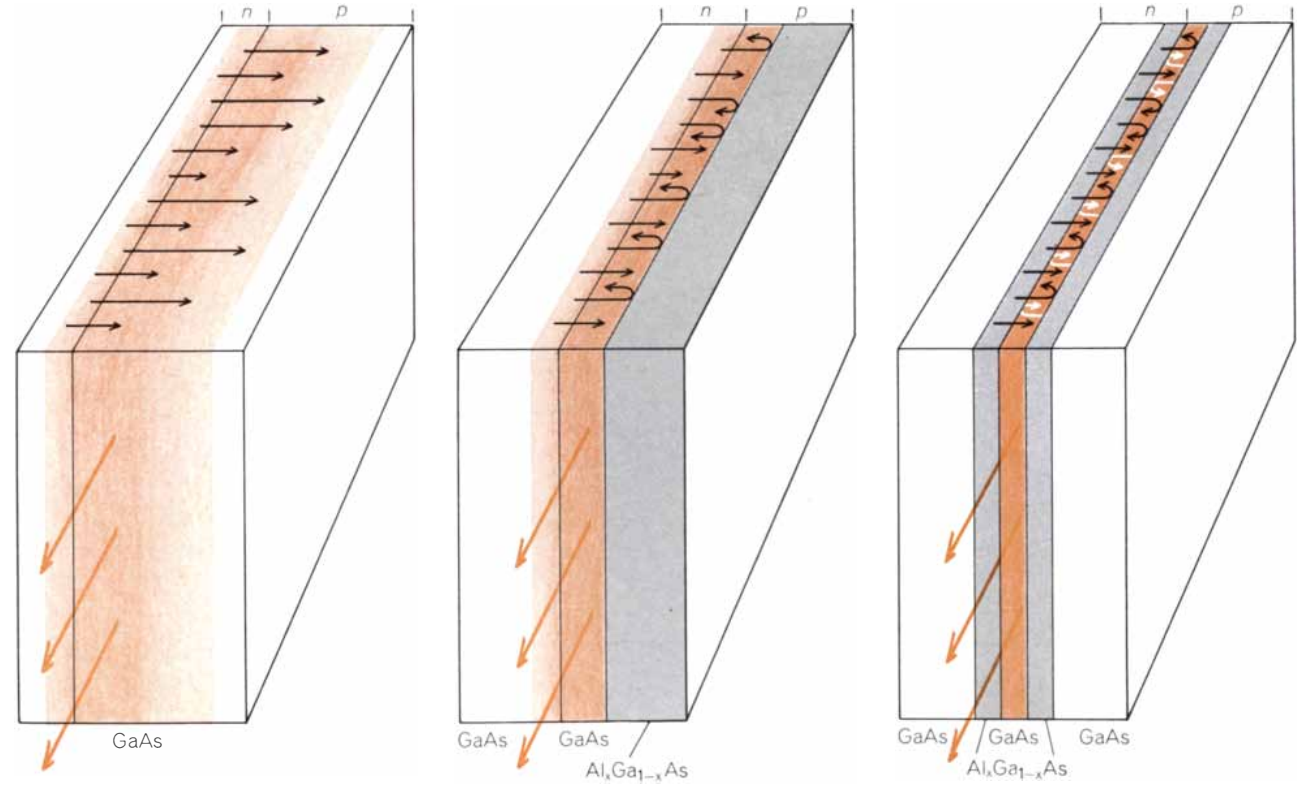
Once in the conduction band of the *p* region, the injected electrons have an energy above their normal energy in this region by an amount roughly equal to the band-gap energy. They exist in the conduction band a certain average length of time (their "lifetime"), and finally they recombine with holes already present in the region. They can recombine in various ways, but in gallium arsenide diodes a substantial fraction of



WHEN ELECTRONS ARE INJECTED across a *p*-*n* junction at low currents, they recombine spontaneously with holes and emit incoherent light in any direction (*a*). As the current is increased stimulated emission starts to take place near the junction where there are excited hole-electron pairs that have not had time to recombine spontaneously; the stimulation occurs when a photon from the recombination of one such pair strikes another pair (*b*). Since the stimulated light has the same phase and direction as the stimulating light a succession of such events results in amplification. When the crystal is made with a flat *p*-*n* junction and has two cleaved ends perpendicular to the plane of the junction, some of the light is reflected by the cleaved faces, which act as partial mirrors (*c*). This light is thus trapped between the mirrors and may contribute to further stimulation and amplification. At a sufficiently high current more light is being produced by stimulation than is being lost by absorption, scattering, leakage out of the mirrors and so on; at that point the amount of trapped light begins to increase rapidly and lasing, or laser action, ensues. This illustration shows these processes taking place inside a homostructure laser, that is, one constructed entirely of a single semiconducting material.



HETEROJUNCTIONS can be made between different semiconducting materials with the same or different conductivity types. In this illustration the potential profiles of the valence band and the conduction band are shown for a *p-p* heterojunction and a *p-n* heterojunction between gallium arsenide (GaAs) and aluminum gallium arsenide ($Al_xGa_{1-x}As$). At the *p-p* heterojunction (*left*) the wider band gap of aluminum gallium arsenide causes a step in conduction band that can act as a potential barrier to reflect electrons if they are injected from the left. Forward-biasing the diode has no effect on this barrier. At the *p-n* heterojunction with no forward bias the bands are already bent to prevent current flow. When forward-biased enough to inject electrons from the *n*-type aluminum gallium arsenide to the *p*-type gallium arsenide (*right*), a residual potential barrier remains on the valence band at the *p-n* heterojunction as a result of the band-gap difference. This barrier prevents holes from crossing the *p-n* heterojunction from the right.



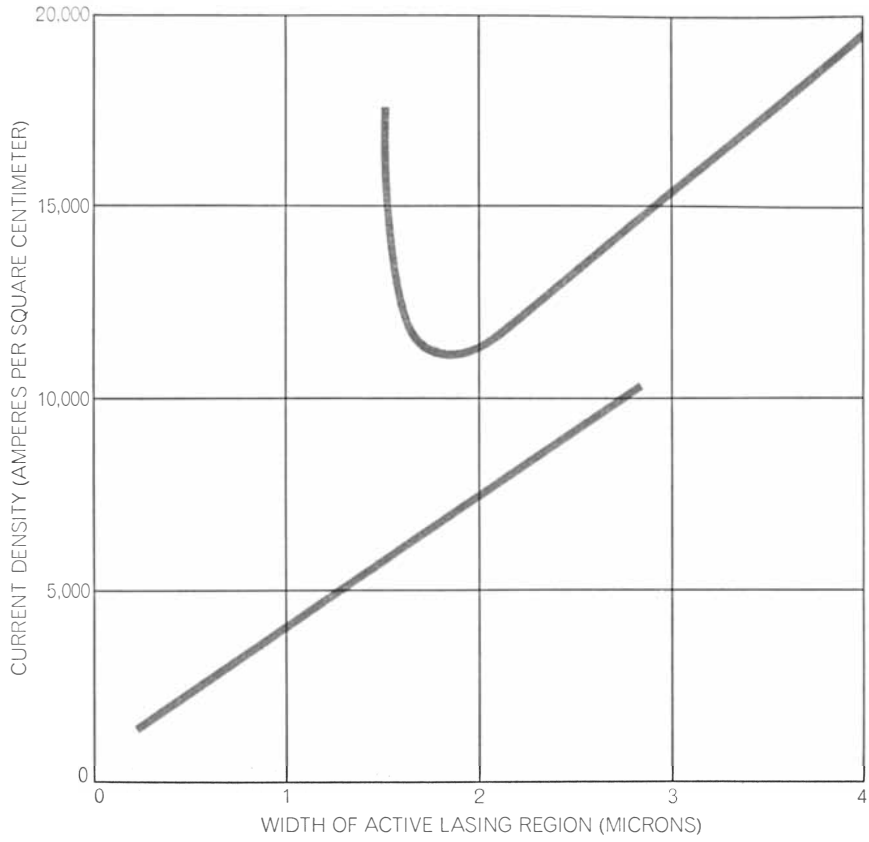
OPERATING PRINCIPLES of a homostructure laser (*left*), a single-heterostructure laser (*middle*) and a double-heterostructure laser (*right*) are compared in this illustration. In the homostructure laser electrons (*black arrows*) are injected across the *p-n* homojunction to varying distances into the *p* region. (Hole injection does not become important until much higher current levels are reached.) In the homostructure device light (*color*) penetrates into the adjacent material out of the region where most of the stimulation is taking place, causing a loss of light that would otherwise contribute to stimulation. In the single-heterostructure device the light does not penetrate significantly through the *p-p* heterojunction and the electrons are reflected. As a result a higher degree of carrier-confinement and light-confinement is attained. In the double-heterostructure device holes, electrons and light are all effectively confined in the very thin region between the two heterojunctions. Electrons are reflected by the potential barrier in the conduction band at the *p-p* heterojunction, and holes are reflected by the potential barrier in the valence band at the *p-n* heterojunction. Light-confinement results from the difference in refractive index between the gallium arsenide and the aluminum gallium arsenide.

the injected electrons give up their energy by the emission of a photon with a wavelength determined approximately by the band-gap energy of gallium arsenide. The process is called electroluminescence, and a device that produces light in this way is called an electroluminescent diode. In gallium arsenide the band gap is about 1.4 electron volts wide. The light emitted by gallium arsenide luminescent diodes has approximately that energy and therefore has a wavelength of about .00009 centimeter. An electromagnetic wave of this length is not visible but is in the near-visible part of the infrared region of the spectrum.

The light thus produced is said to be a spontaneous emission because the holes and electrons combine spontaneously after a certain average time interval without being affected by photons from other recombination events. As the injection current is increased a process called stimulated emission will start to become important. In general when an electron in an excited state is struck by a photon with the same energy as its excitation energy, it will be stimulated to emit that energy as a photon of light during a transition to a lower energy level; the stimulated light will have the same phase, wavelength and direction as the stimulating light. This idea was quantitatively developed by Albert Einstein in 1917. Injected electrons in the p region are, of course, excited electrons, their excitation energy being approximately E_g. Therefore when an injected electron that has not had a chance to recombine spontaneously is struck by a photon from a previous recombination event, it may be stimulated to emit its photon in the same direction and with the same phase as the stimulating photon; under these circumstances both photons will have the same energy, E_g.

As the injection current is increased the light produced by stimulated emission increases, because the number of hole-electron pairs present near the p-n junction increases; amplification, particularly along the junction plane, is the result. If the diode has been made with a flat p-n junction, it can be fabricated in such a way as to have two flat ends of the crystal perpendicular to the plane of the junction. These flat ends (actually cleaved crystal surfaces) act as partial mirrors and tend to trap some of the light produced parallel to the junction plane by reflecting it back into that region. The structure is now a laser diode [*see bottom illustration on page 35*].

The introduction of the mirrors feeds the stimulated light back into the plane of the junction and thus furthers amplification. When the injection current becomes high enough so that the light in the crystal making a round trip along the junction plane is amplified enough to offset losses due to absorption, to scattering, to leakage out of the mirrors and so on, the diode is said to be lasing. In the language of workers in the field, the gain now equals the loss. With a further increase in injection current the added hole-electron pairs recombine primarily by stimulation. An intense beam of light is observed to emerge from the mirror ends of the crystal; the entire process is called light amplification by stimulated emission of radiation (hence the acronym "laser").



THRESHOLD CURRENT for the onset of laser action at room temperature in a single-heterostructure laser (*top curve*) and a double-heterostructure laser (*bottom curve*) are compared in this graph as a function of the width of the active lasing region. In the single-heterostructure device the threshold current decreases with the width of the active lasing region to near two microns because the smaller widths correspond to smaller active-region volumes. The smaller this volume is, the fewer electrons are needed to reach the electron density required for the lasing to start. Below about two microns the forward bias needed to start lasing is large enough to initiate hole injection. This means that some current is being carried by holes going into the n region, where they do not contribute to the laser action. As a result the lasing threshold rises. This effect does not apply to double-heterostructure lasers because of the potential barrier at the p-n heterojunction. The thresholds of double-heterostructure lasers are lower overall because of their superior light-confinement.

Simple diode lasers of the type described above are called homostructure lasers because they are made of a single semiconductor material. Even when they are made of the best material available (gallium arsenide), lasing at room temperatures requires high current densities (25,000 to 100,000 amperes per square centimeter). In other words, if the area of the junction plane of the laser diode were a square centimeter, the current would in the best case be 25,000 amperes. Actually, of course, the areas are much smaller, typically .0005 centimeter, and the corresponding current is 12.5 amperes. The use of the "current density" notation enables us to compare diodes with different areas. The minimum current density needed to start lasing action in a diode is termed the threshold current density. It is because of their high threshold current density that such lasers are always run under pulsed conditions at room temperature in order to keep the power dissipation low enough to prevent overheating. Typically pulses of a millionth of a second are used with an interval between

pulses of a thousandth of a second (to allow cooling). Continuous, or nonpulsed, operation is possible only at much lower temperatures (typically 77 degrees Kelvin, the boiling point of liquid nitrogen), because the lasing threshold current is much lower at lower temperatures.

In recent years the maximum temperature at which homostructure lasers could be run continuously has been 205 degrees K. This temperature was achieved in 1967 by J. C. Dyment and L. A. D'Asaro at Bell Laboratories in a test where the laser was mounted on a diamond to allow very efficient heat removal. Such a heat-removal device is called a heat sink. From that work it was recognized that, in order to operate a semiconductor diode laser with direct current (continuously) at room temperature, not only was an efficient heat sink necessary but also a large reduction in the threshold current density at room temperature had to be achieved.

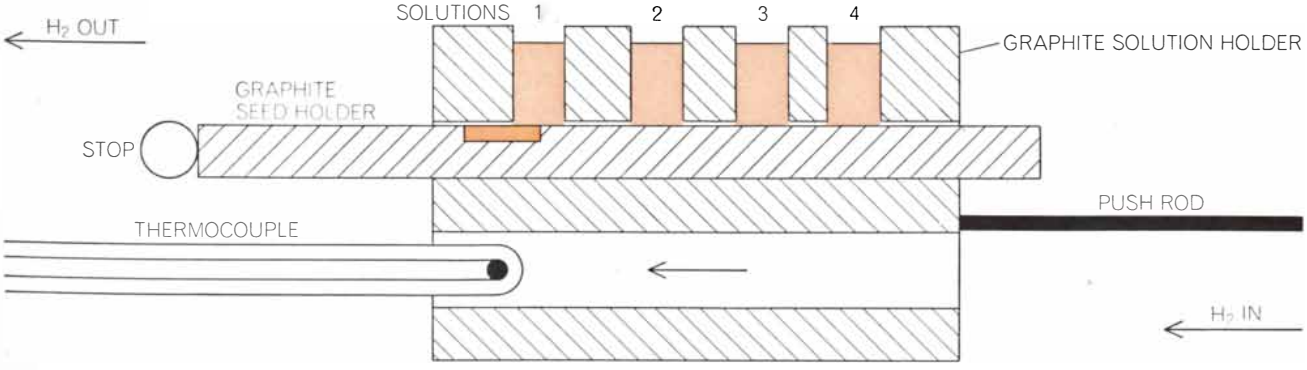
In retrospect one can appreciate the reasons for believing that a considerable reduction in the threshold current density could be obtained, on the basis of the following simplified argument. In the region near the plane of the *p-n* junction, where stimulated emission is taking place in a semiconductor laser diode, a critical density of light and electrons must be attained in order for the amplification due to stimulated emission to overcome losses. In gallium arsenide laser diodes, however, the distance that the electrons can diffuse away from the *p-n* junction after injection increases with temperature. At room temperature the electron's average diffusion distance from the *p-n* junction after injection may be as much as several microns. Thus the width of the region contributing to lasing is uncontrolled; it depends not only on the semiconductor used but also on the temperature and the impurities present. This means that the volume of semiconductor material that must reach the required electron density is not a parameter that one is able to vary in a homostructure laser. Clearly if this volume could be reduced, a lower threshold current density could be achieved.

Furthermore, much of the light is not well confined in such a laser; the light tends to spread into inactive regions of the crystal. That reduces the photon density available to contribute to amplification because the light occupies a larger volume and because some of it has spread out into regions where there are few excited electrons. The light is absorbed rather than amplified because its energy is approximately the same as the energy of the band gap of gallium arsenide. Therefore to a first approximation the high threshold current density required for lasing to occur at room temperature in homostructure lasers is demanded not only by the need to maintain a required electron density in a large volume of the *p* region into which the carriers diffuse but also by the high light losses and reduced photon density resulting from the spreading of the laser light.

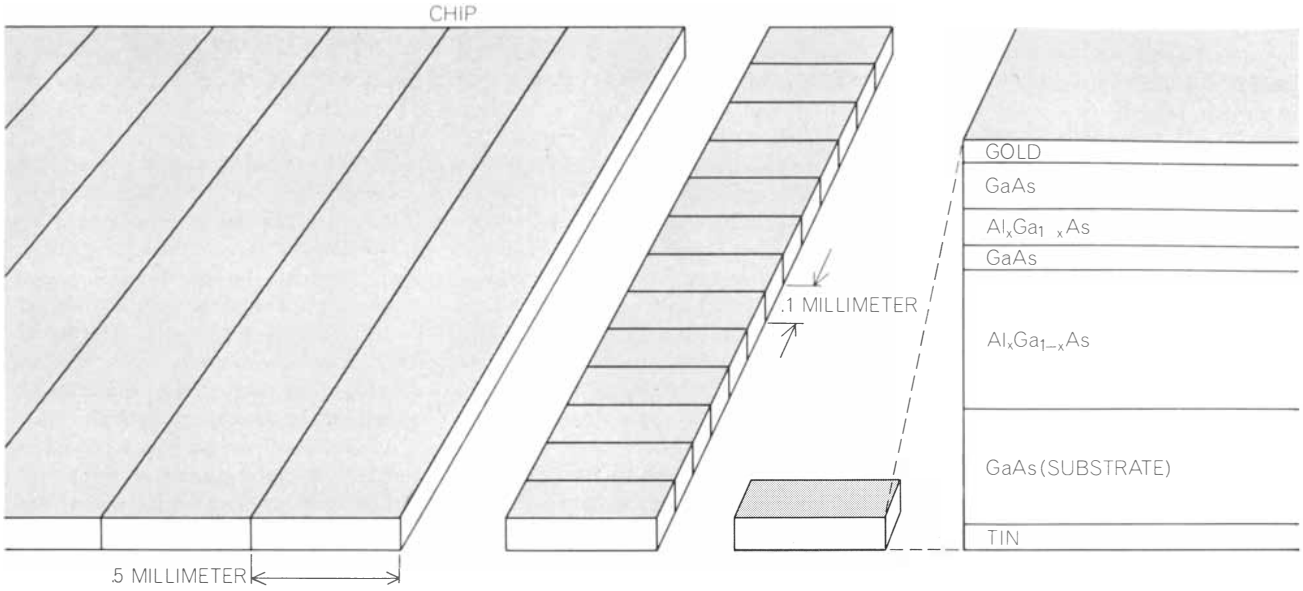
The achievement of low-threshold lasers that can be made to operate continuously at room temperature results directly from the use of modified structures that restrict both the light and the electrons to a small active region. This was done by utilizing crystals that had a region of pure gallium arsenide and a region where some of the gallium atoms in the gallium arsenide crystal had been replaced by aluminum atoms. The latter region consists of the semiconductor aluminum gallium arsenide, and the junction of the two materials is called a heterojunction. The chemical symbol for gallium arsenide is GaAs and for aluminum gallium arsenide is $Al_xGa_{1-x}As$, where x is 1 or less. This symbolism indicates the one-for-one replacement of gallium atoms by aluminum atoms in the lattice. If, for example, half of the gallium atoms are replaced by aluminum atoms, the formula of the resulting semiconductor is $Al_{.5}Ga_{.5}As$.

In aluminum gallium arsenide the band gap is larger than it is in gallium arsenide, the amount of this increase itself increasing with the fraction of the gallium atoms that are replaced by aluminum atoms. The increased band gap has profound effects not only on the behavior of the holes and electrons when they encounter these heterojunctions but also on the electroluminescent light that is generated. The holes and electrons are influenced because the presence of the heterojunction results in abrupt changes in the width of the band gap and hence in the shape of the bottom of the conduction band or the top of the valence band in its vicinity [*see top illustration on page 36*].

The electroluminescence is influenced because the aluminum gallium arsenide is transparent to light that has the band-gap energy of gallium arsenide (as a result of the larger forbidden gap of aluminum gallium arsenide) and because hand in hand with an increase in E_g with increasing x is a decrease in the refractive index. In the past, when other materials were used for the preparation of heterojunctions, these properties could not be utilized because another property of heterojunctions is that they tend to trap holes and electrons in such a way as to prevent recombination from occur-



APPARATUS used for growing multilayered structures to serve as diode lasers is depicted here. The graphite "boat" contains four wells filled with solutions of gallium, aluminum, arsenic and an appropriate impurity element. A wafer consisting of a slice of a single crystal of gallium arsenide, mounted in a sliding graphite holder, is successively brought under each solution. When a suitable cooling cycle is used, successive layers are grown from each solution. From one to four layers of gallium arsenide or aluminum gallium arsenide, each with a proper complement of impurities, can be grown on the wafer. A typical wafer measures four millimeters wide by 10 millimeters long by half a millimeter thick. Growth temperatures range from 800 to 1,000 degrees Celsius.



LASERS ARE FABRICATED from the multilayered semiconducting wafers by first lapping the substrate of the wafer so that the wafer has a total thickness of about 75 to 125 microns. Very thin metal electrical contacts are then applied by directing atomic beams of the proper elements onto the surfaces in a high-vacuum system. The crystal is then cleaved into bars approximately half a millimeter wide, and these in turn are sliced with a fine wire saw into the finished lasers, which are roughly .1 millimeter wide.

ring with the production of useful light. This does not happen with heterojunctions between gallium arsenide and aluminum gallium arsenide. We assume that this is because the two compounds have the same crystal structure and almost exactly the same spacing between atoms in the crystal lattice.

How do all these factors aid in achieving a reduced lasing threshold in diode lasers? To make such lasers one must first of all be able to make crystal structures with one or more heterojunctions between gallium arsenide and aluminum gallium arsenide. Such structures are generally made by growing layers of both compounds on slices of single crystals of gallium arsenide. At present these layers are grown on the gallium arsenide crystal at high temperatures from liquid solutions containing a mixture of gallium, aluminum and arsenic in much the same way that ice crystals grow larger when heat is removed from a mixture of ice and water.

The simplest use that can be made of the ideal nature of the gallium arsenide–aluminum gallium arsenide heterojunction for the reduction of laser thresholds is the single-heterostructure laser. This laser structure is made by growing a layer of aluminum gallium arsenide on an *n*-type gallium arsenide substrate wafer and simultaneously diffusing zinc a short distance into the wafer from the growth solution. Zinc is an acceptor impurity in gallium arsenide, and enough of it can convert *n*-type material into *p*-type material. Zinc also has the property that its atoms can, at sufficiently high temperatures, readily move into the gallium arsenide lattice from an external source, be it vapor, liquid or solid. The result then is a wafer consisting mostly of an *n*-type substrate but with a thin *p*-type region caused by the diffused zinc. On the surface of the *p*-type region there is a layer of *p*-type (Zn-"doped") aluminum gallium arsenide. The structure therefore has a *p-n* junction in gallium arsenide (a homojunction) separated by a small amount of *p*-type gallium arsenide from a layer of *p*-type aluminum gallium arsenide. We refer to the latter boundary as a *p-p* heterojunction. The *p-n* junction plane and the *p-p* heterojunction plane are parallel to each other. The structure has a single heterojunction, hence the name single-heterostructure laser.

In constructing the actual lasers from the wafer the substrate is thinned by lapping with an abrasive to a total thickness of about 100 microns. Electrical contacts (thin metal layers) are applied by vacuum deposition on the *n* side and *p* side of the wafer, and the resulting combination is sawed and cleaved into rectangular chips approximately 100 by 400 microns on a side [*see illustration above*]. These are the lasers. When a chip is forward-biased enough for electron injection across the *p-n* junction to occur, the injected electrons go into the *p*-type gallium arsenide region but are stopped (reflected) by the potential barrier formed by a bend in the conduction band at the heterojunction. The electrons are hence not free to diffuse into a larger volume, as happens with homostructure lasers. Moreover, as a result of the decrease in the refractive index at the heterojunction there is little penetration of the light out past the heterojunction boundary. The lasing threshold current is thus reduced below that of homostructure lasers because both the electrons and the light can be confined to a smaller volume in the single-heterostructure laser. The lasing threshold is now dependent on the width of the region between the *p-n* junction and the *p-p* heterojunction because smaller width means a smaller volume to be filled with electrons to the density required for lasing [*see illustration on page 37*].

With the single-heterostructure laser some of the light is still lost owing to its penetration across the *p-n* junction into the *n*-type gallium arsenide. Furthermore, when the voltage needed to cause the current to flow is raised sufficiently, holes are injected across the *p-n* junction into the *n*-type gallium arsenide. The hole injection occurs primarily in single-heterostructure lasers with a very narrow active region. Since this part of the current is not useful for lasing, the threshold current rises. Even with such defects single-heterostructure lasers with a room-temperature threshold as low as about 8,000 amperes per square centimeter have been made.

A further considerable reduction in current thresholds results if an additional heterojunction is located at the *p-n* junction. In order to achieve such a *p-n* heterojunction a three-layer struc-

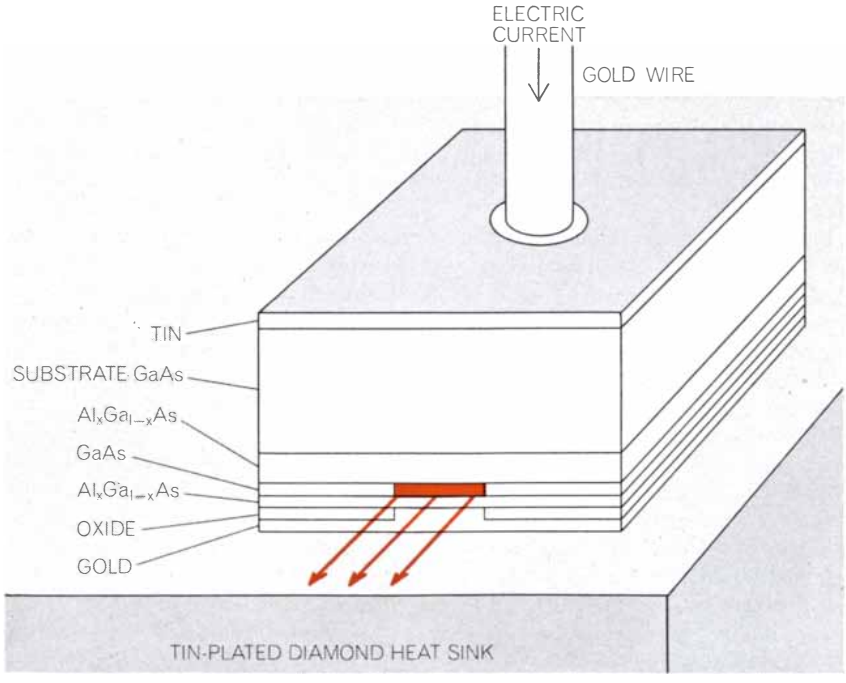
ture must be grown onto the gallium arsenide substrate. Actually four layers are grown [*see bottom illustration on page 33*]. The fourth layer (of gallium arsenide) serves to provide a convenient surface to make electrical contact.

The important part of this layered structure is a sandwich of *n*-type aluminum gallium arsenide, *p*-type gallium arsenide and *p*-type aluminum gallium arsenide; the sandwich comprises the first three layers on the substrate. An *n-p* heterojunction exists between the first two layers and a *p-p* heterojunction between the second two. The *p*-type gallium arsenide region between these two heterojunctions defines the active lasing region; such a structure is called a double heterostructure.

As before, contacts are applied and the wafer with the layers is cleaved into small rectangular laser chips. When these are forward-biased enough for electrons to be injected into the second layer, a residual step remains on the valence-band edge. As before, electrons are reflected at the *p-p* heterojunction, but now in addition holes are reflected at the *p-n* heterojunction. The result of this second reflection is that at no voltage near the voltage needed to obtain lasing do we get hole injection with its attendant increase in threshold current. It is therefore possible to construct lasers with much narrower active regions than was possible with the single-heterostructure laser. There are now steps in the refractive index at both the *p-p* and the *p-n* heterojunction and the light penetration out of the active region is restricted on both sides. We thus have a structure that for widths of the active region as small as .5 micron has almost complete light and recombining-carrier confinement. This means that it is possible to have a very thin active region and hence a very small volume in which holes, electrons and light are confined. Therefore in order to obtain sufficient carrier and photon density for lasing to start we need a much lower current than is required by homostructure lasers or single-heterostructure ones. Threshold current densities for laser action as low as about 1,000 amperes per square centimeter have been obtained with double-heterostructure lasers, and it is believed that lower thresholds will result from further research.

Although the achievement of such relatively low threshold current densities brought the achievement of continuous operation of diode lasers at and above room temperature much closer, one major problem remained. At currents of about 1,000 to 3,000 amperes per square centimeter much heat still must be removed from the laser during operation. Since most of the heat is generated in the active region it must be removed through at least one layer of aluminum gallium arsenide. Such ternary compounds have much poorer thermal conductivity than binary compounds such as gallium arsenide. The problem here is therefore worse than it was with the homostructure lasers. To overcome this difficulty at least one of the two aluminum gallium arsenide layers should preferably be very thin. Estimates indicated that by making one of these layers approximately a micron thick, continuous operation with the low-threshold lasers would be possible if the laser were bonded to an efficient heat sink. The multilayer-growth technique was therefore developed in such a way that the required thin layers were obtained.

An efficient heat sink was achieved by alloying the metal contact on the *p*-type surface of the laser to a tin-coated diamond. The first units to run continuously at and above room temperature were obtained in our laboratory in that way. The structure thus obtained is a simple parallel-plane configuration. More efficient heat sinks have recently been obtained at Bell Laboratories by another group with a "striped" heat-sink configuration [*see illustration at left*]. With this configuration only a small stripe in the active region is affected by the current. Less heat is produced because the active part of the device is smaller, and the heat that is produced is more readily dissipated because heat dissipation occurs laterally as well as vertically in the striped structure. A number of such units have already been tested. Continuous operation of such striped lasers on both diamond and copper heat sinks has been achieved; with diamond heat sinks and the stripe-geometry modification lasers have been run continuously at temperatures as high as about 100 degrees C.



STRIPE GEOMETRY characterizes the design of this double-heterostructure laser. The insulating oxide layer restricts the current to a narrow stripe along the active region, enabling heat to flow more readily away from the lasing region to the diamond heat sink.

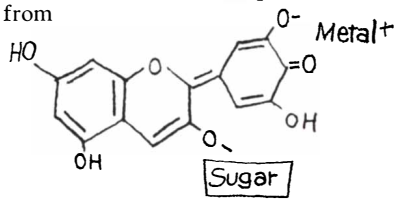
We have described the evolution of a new class of lasers, the heterostructure lasers, which have so far been made with output powers of tens of milliwatts while operating continuously with wavelengths between .00009 and (with slight modification of the structure) .000077 centimeter. Total efficiencies for the conversion of electrical to optical power of about 10 percent have already been achieved during continuous operation of several of these lasers at Bell Laboratories. The numbers will probably improve with further study. Although the devices are still in the research stage, they show great promise as small, cheap, efficient and rugged carrier-wave generators that can be utilized for optical mass-communications systems and other applications. We believe the evolution of other devices that utilize the ideal nature of lattice-matched heterojunctions will be accelerated by the results of the work described here.

We want to be useful ...and even interesting

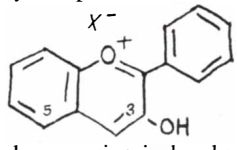
Kodak

Roses are red, violets are blue Sugar is sweet and attached to the 3- or 5-hydroxy group

Delphiniums, for example, are blue from



To make blues and reds, nature is very busy right now in flowers, fruits, and berries synthesizing pigments on this basic anthocyanin pattern:



Dye chemists keep prying jealously, even dye chemists who already ape the colors as well as ours do with pigments on film and paper that last longer than a few summer days.

We have recently found that certain man-made molecules built on this general plan join together by threes to form an intensely blue compound. No sugar.

Seeing no way that our shareowners can profit from keeping this a secret, we have published a short paper* on the details in the hope that *someone* can profit from the information in a way not yet clear. That's science.

—and from flowers to babies

To wander in the garden and spend one's days wondering what makes the roses red and the delphiniums blue is not quite pulling one's load. Not today it isn't. But was it irresponsible behavior 65 years ago? 65 years is not a very long span of moral history.

In 1906 a Russian botanist named Tswett at the University of Warsaw proved he could separate the components of coloring matter in plants by the same principle that leaves a ring on cloth after a partial failure at stain removal. He must have been pleased. He would have been more pleased to watch the following routine for testing blood or urine of everybody at birth for the purpose of detecting metabolic errors that can be treated by diet and otherwise lead to mental retardation and other impairment:

Tswett called it chromatography. He might not have recognized it in this version. The idea is new of applying the solvent in a gel and incorporating therein a color-forming compound that reacts with the separating compounds from the sample mixture to form identifying colors (after heating to 100 C in the case of amino acids). Many hands have carried chromatography a long way.

Our contribution has been analogous to what we did to photography a decade and more before Tswett's discovery and to what we did later in color photography (the better term "chromatography" having been appropriated by Tswett): minimizing the manipulation, not with fancy hardware, but by keeping the elegance back in the factory and financing it by broadening the market for the end result.

The new KODAK CHROMAT/O/SCREEN *Analysis Kits will soon be stocked by leading lab suppliers, complete with detailed directions, discussion, and bibliography. The one for amino acids serves the infant-screening market. Others separate respectively amphetamines, barbiturates, alkaloids, catecholamines, sugars. Clinical chemistry may not be the only place for such quick and easy thin-layer chromatography.*

For the enthusiastic teacher

Let's say that next year 30,000 students in North America (or in the English-speaking nations of the world) are to be taught some flavone chemistry. Or that 300,000 should learn how molecules and light interact. They could all be assembled and hooked together, as for watching a championship prize-fight. Alternatively, individual teachers, each greatly interested and each with a somewhat different viewpoint for a somewhat different level, could project a separate vision to a separate band of students.

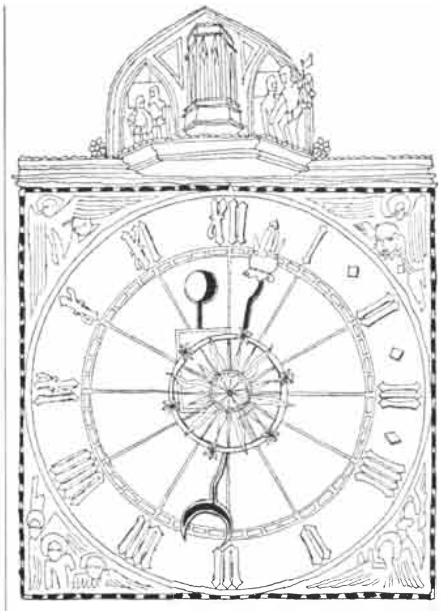
In *chemistry?* Yes, in chemistry. Or in other subjects.

The teacher too enthusiastic to stick to that dusty audiovisual medium chalk-and-talk can consider making a film. It's not that big a deal. Some of the students may well have made films themselves. Little kids are doing it. To express oneself with film is to match what our more articulate forebears might have tried with verse or well-turned phrase. Film language gets learned early. A KODAK PRESSTAPE Universal Splicer (list $8.50, at camera counters) is an instrument for eloquence through self-discipline. "Movies With a Purpose" (single copies free from Dept. 412-L, Kodak, Rochester, N.Y. 14650) teaches film grammar, without which communication is difficult. Animation, for subjects like molecules, may be the heart of the matter. For suggestions on how to do it, see *Journal of Chemical Education 48,* 143 (Feb. 1971).

**Journal of Heterocyclic Chemistry 7,* 1395 (1970)

Price subject to change without notice.

SCIENCE AND



Unreordered Priority

World military expenditures continued to rise in the past year, reaching a record peak in 1970 of $204 billion, or 6.4 percent of the total gross national product of the nations of the world, according to a report released by the U.S. Arms Control and Disarmament Agency. In summarizing the results of the agency's fifth annual survey on the size and impact of world arms expenditures, the report's authors observe that last year's world total was the equivalent in dollar value of a year's income produced by the 1.8 billion people in the poorer half of the world.

In the six years since the agency's first world survey military outlays in current dollars have risen almost 50 percent. In terms of constant dollars (that is, if prices had remained unchanged) military expenditures for 1970 would have exceeded those for 1964 by roughly 20 percent. This 20 percent "real" increase is reflected partly in increased numbers of personnel, which are up more than 10 percent since 1964 (to an estimated 23 million), and partly in investment in increasingly sophisticated military equipment, including both strategic (long-range nuclear) and conventional weapons.

In spite of a slight decline for the second straight year in the rate at which military expenditures are increasing, the current survey could find "no evidence that a widespread shift from military to civilian uses of resources has begun. Nor does the slower rate of increase in total military expenditures in the past two years reflect a worldwide tendency to restrict such outlays. In the vast majority of countries for which recent information is available, there was a further rise in military budgets in current prices in 1970."

Developing countries in particular continued to increase their share of world military outlays, devoting to military programs funds that could have been allocated to economic development. For example, it is estimated that in developing countries the increase in military spending since 1964 represented the equivalent of three years' expenditures on public education for the billion school-age children in those countries. In general more public funds continued to flow into military programs than into public-education or public-health programs for the world's population of 3.6 billion.

Although the share of world GNP going into military programs has declined from a high of 7.3 percent in 1967 to 6.4 percent in 1970, the report points out that "changes over time in the ratio of military spending to GNP must be evaluated cautiously, both as measures of relative burden and as indicators of arms restraint. Economic growth is population-related. Unless the civilian portion of world GNP—the investment and consumption component—grows with rising population, living standards fall. Unless the civilian portion adjusted for the population rise expands with rising GNP, living standards have not benefited fully from the growth dividend."

The record on this point for 1970, as compared with 1964, shows that military expenditures have indeed absorbed part of the growth dividend, even though the share of world GNP that they take is diminished. In 1970 the dollar value of world GNP was $1,252 billion higher than in 1964. Taking into account price rises and population growth, however, the "real" gain in world GNP was only $450 billion, or about a third of the apparent gain. The increase in world arms expenditures in constant prices took $28 billion of this dividend, representing a considerably larger share of real growth in developing countries than in developed countries.

As an indicator of arms restraint, the report concludes, "a downward trend in the ratio of military expenditures to GNP

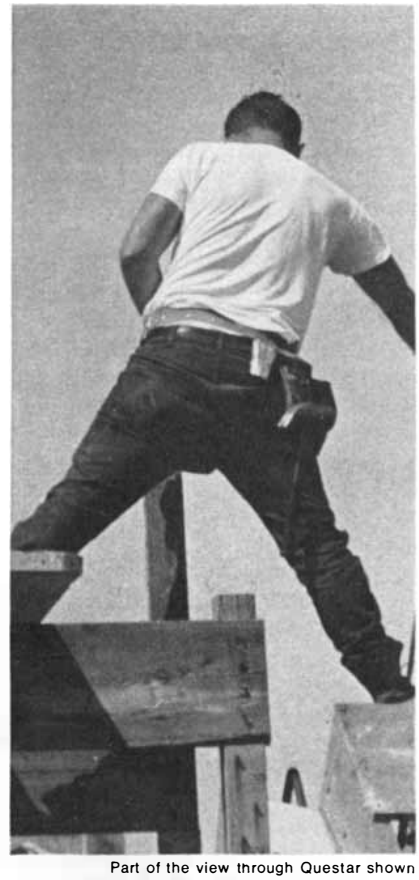
is not sufficient. With a rapidly growing economy, even a declining ratio to GNP can conceal an intensified arms race."

How Many People?

Japan is unique among the major nations in having steadily reduced the rate of growth of its population to a point suggesting that the population will begin to decrease in a generation or so if the present level of reproduction continues. Now concern is rising that the trend will inhibit the rapid economic growth the nation has experienced over the past decade. The situation is described by Minoru Muramatsu, chief of the Section of Public Health Demography of the Institute of Public Health in Tokyo, in "Country Profiles," a publication of the Population Council.

Muramatsu notes that the population of Japan was about 50 million in 1912 and had doubled by 1967. The annual rate of growth of the population during this period therefore averaged 1.3 percent. In the U.S. the population has grown at rates ranging from a high of 21 percent in the first decade of this century to a low of 7.2 percent between 1930 and 1940 (see "The Census of 1970," by Philip M. Hauser, page 17). In recent years the Japanese rate has been about 1.2 percent per year; by 1985, according to the Institute of Population Problems (an agency of the Ministry of Health and Welfare in Japan), it will decline to .7 percent.

The control of population in Japan has been largely voluntary. It is traced by Muramatsu back to the "rice riots" of 1918, when people in many parts of the country demonstrated against the high price of rice; the episode was widely interpreted as being an indication that there was not enough food to go around. "Regardless of official policy decisions or the pros and cons of birth control," he writes, "people proceeded to practice birth control on their own, primarily through induced abortion." Two government policies, however, were "notable influences." The first was the establishment of the Eugenic Protection Law; the second, the establishment of a national program for the promotion of family-planning practice. The chief aim of the law, which was enacted in 1948, was to liberalize induced abortion. The num-



Part of the view through Questar shown at lower left of photograph below

ber of reported induced abortions rose sharply to a peak of 1.17 million in 1955; thereafter it declined gradually to 744,451 in 1969. The family-planning policy, inaugurated in 1952, resulted from alarm in the Cabinet about the abortion rate; the policy was that contraception was to be recommended as a far more reasonable method of family limitation.

Then came the economic boom, which has kept Japan's economic growth rate above 10 percent a year since 1963 (except in 1965). During the same period, because of the downward trend in fertility, a shortage of young workers developed. The annual increment of the productive age group (15 to 64) was 960,000 from 1965 to 1970 but will be only 620,000 from 1970 to 1985.

As a result of these effects, according to Muramatsu, "a reaction to the low fertility trend and apprehension of 'diminishing' population are developing. Business magnates and industrial leaders have voiced concern over the shortage of young workers." The Population Problems Council gained wide attention with a report saying that Japan would have to raise its fertility rate slightly in order to bring about a stationary population in the future and recommending that efforts to raise fertility levels should be directed toward the strengthening of social-development problems so that the people encounter fewer economic, social and other deterrents to pregnancy, childbirth and child-rearing. In the present debate over population policy, Muramatsu writes, "some maintain that an increase in the birthrate is imperative, while others wonder why Japan should raise its fertility rates in the face of an inevitable increase of tens of millions in the population expected during the several decades to come." Both groups, however, recognize that "the future fertility trends in this country will be largely determined by the people, not by interventions from outside."

Venera on Venus

The first spacecraft known to have made a soft landing on another planet was the Russian vehicle *Venera 7*, which reached the surface of Venus last December 15. At the time there was some doubt whether radio signals from *Venera 7* had continued until the actual landing and, if they had, whether they had persisted after the craft had come to rest. The Russians had failed in three earlier attempts to soft-land a spacecraft on Venus. *Venera 4*, *Venera 5* and *Venera 6* had all entered the planet's atmosphere successfully and had transmitted measurements through the first stages of a parachute descent, but in each case the messages had stopped when the landing capsules were between 17 and 22 kilometers above the surface. Details of the successful landing of *Venera 7* are now reported by a team of U.S.S.R. investigators in the *Journal of the Atmospheric Sciences*.

They describe *Venera 7*'s landing capsule as an oblong, aerodynamically shaped object that weighed about 100 kilograms more than the capsules carried by *Venera 5* and *Venera 6*. (They do not mention what the weight of those capsules was.) The extra weight was used to provide additional insulation to protect the capsule from the high temperatures that had presumably knocked out the transmitter on the earlier craft before they had reached the surface. The report notes that the capsule's parachute was fabricated of material able to withstand a temperature of about 530 degrees Celsius. The capsule was built to withstand the same temperature and a pressure of 2,700 pounds per square inch. Just before the capsule had entered the atmosphere of Venus the temperature inside it was about −7 degrees C. The capsule contained aneroid manometers able to measure pressures of between seven and 2,200 pounds per square inch and resistance thermometers with a range between 25 and 540 degrees C.

The descent through Venus' atmosphere took 35 minutes, or about 25 minutes less than the predicted time. The reason is not known. The parachute opened when the external temperature was about 200 degrees C. The capsule reached the surface moving at a velocity of 16.5 meters per second, or about 37 miles per hour. One evidence of the capsule's landing was an immediate reduction by a factor of 100 in the strength of the radio signals received on the earth, evidently caused by a sudden change in the alignment of the capsule's antenna. Intensive computer analysis was required to decode the signals that continued to arrive for 23 minutes.

The temperature on the surface of Venus was 474 degrees C., with a probable error of ±20 degrees. The highest temperature reported by earlier Venera capsules was about 325 degrees C., measured at an altitude of about 17 kilometers. The pressure at the surface, now measured for the first time, is 90 ± 15 kilograms per square centimeter, or about 1,300 pounds per square inch. An analysis of the temperature distribution with altitude leads the Russians to conclude that there is strong convective mixing, thus making it unlikely that solar radiation is trapped near the surface by the "greenhouse" mechanism. This refers to the hypothesis that the abundance of carbon dioxide in Venus' atmosphere would allow solar light to reach the planet's surface but would block the outflow of radiation at longer wavelengths. The final observation in the report is that "it seems fairly certain that the surface of Venus is hard enough to bring a spacecraft to an abrupt stop."

Rising Fever

The grave tick-borne infection known as Rocky Mountain spotted fever, now relatively rare in the Western states where it was first observed in 1896, appears to be on the rise in the U.S. as a whole. Fewer than 300 cases were reported nationwide in 1968, but in 1969 (the most recent year for which statistics are available) the number had risen to nearly 500. The state with the worst record was Virginia, which reported a total of 91 cases and four fatalities. Of the next five states on the list only one (Oklahoma) is west of the Mississippi; three of the five (Maryland, North Carolina and Tennessee) border on Virginia. None of the 49 continental states except Vermont was free of the disease in 1969. The symptoms of the disease are high fever, a hemorrhagic skin rash, headache, muscle pains and sometimes delirium and coma.

Reporting on the increase in *The Journal of the American Medical Association*, Allen H. Peters of the Virginia State Health Department attributes it to two interacting factors. One is the reversion of farmland to woodland that is well suited to the three species of ticks that are known vectors of the disease organism (*Rickettsia rickettsii*). The ticks feed on mice, rabbits and other wild animals and on domesticated dogs in particular. The other factor is the expansion of suburban populations into the new woodlands.

Sarah Language ? Learn

Can a chimpanzee learn a language? The answer is clouded by other questions: What constitutes proper training for language? Indeed, what is language? Without finally answering these fundamental questions David Premack of the University of California at Santa Barbara has reported in *Science* what appears to be some success in teaching Sarah, an African-born female chimpanzee, to use a written language. The words of the language are pieces of plas-

tic of various shapes and colors, backed with metal so that they adhere to a magnetized slate. Sentences are written vertically by placing one piece of plastic below the other.

Premack began the training by establishing and then "mapping" a social exchange between Sarah and a trainer. The trainer put a piece of fruit on the table; Sarah ate it. After a while the trainer put a plastic "word" alongside the fruit. Then the fruit was moved out of reach; Sarah was induced to place the word on the slate, after which she was given the fruit. Different fruits were offered along with different words (pieces of plastic). Sarah learned to associate the correct word with each fruit. Then a second perceptual class and a second word were involved; depending on which of several trainers was present, Sarah had to write "Mary apple" or "Randy apple" and so on in order to get the fruit. Again, the animal learned to use the correct trainer's name when several names were available but only one trainer was present. Sarah was taught to write such sentences as "Mary give apple Sarah."

After teaching Sarah the concepts "same" and "different" and the words for them, Premack and his colleagues introduced questions—first in the form "*X* is the same as what?" and then "Is *X* the same as *Y*?" with pieces of plastic standing for the interrogation mark, for "Yes" and for "No." Then Sarah went on to metalinguistics, or the use of language to teach language. She learned, for example, to place "name of" between the word "apple" and an apple, "not name of" between "apple" and a banana. She apparently understood the concept "name of," because she learned new words correctly after seeing an object accompanied by "name of" and one new plastic word and by "not name of" and another new word; it could not, then, have been merely the physical relation between object and word that led to the association.

Just as "name of" was introduced after some objects had been named, so the class concepts of color, shape and size were introduced after the establishment of such properties as "red" and "yellow," "round" and "square," "large" and "small." In time Sarah was able to manipulate the properties, using their words in syntactic contexts other than the ones they were learned in. For example, having learned "Mary give Sarah red" and "Brown color of chocolate," she could follow the instruction "Sarah insert brown [in] red dish," choosing a brown object from among several objects and choosing the dish of the right color.

Sarah learned the preposition "on," compound sentences, a symbol for plurals and such words as "all," "none," "one" and "several" and then went on to the conditional relation. Here, in order to avoid the discontinuous construction "if . . . then," Premack used a piece of plastic standing for the symbolic-logic particle $\supset$. In time Sarah could perform correctly in response to such instructions as "Mary take red $\supset$ Sarah take apple" and "Mary take green $\supset$ Sarah take banana."

Attack by Pheromone

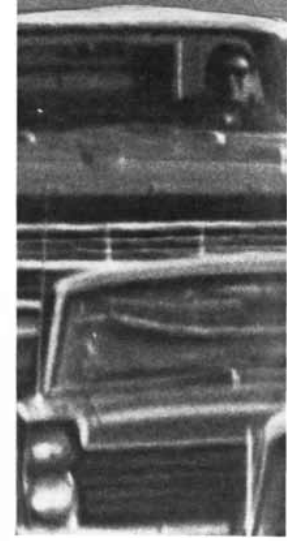
Certain ants conduct "slave raids" on other ants: they remove the pupae of the other ants, and when the pupa metamorphoses into an adult, it takes up the role of a worker in the raiders' nest. In their raids ants of the *Formica sanguinea* group lay a guiding trail of pheromone, a substance usually defined as something that is secreted by an animal that influences the behavior of other animals of the same species. It now appears that the pheromones of the *F. sanguinea* group also alarm and scatter the ants being raided. These pheromones are called "propaganda substances" by F. E. Regnier of Purdue University and E. O. Wilson of Harvard University.

Writing in *Science,* Regnier and Wilson report that the pheromone-secreting gland in the abdomen of ants in the *F. sanguinea* group is hyperdeveloped: it accounts for as much as a tenth of the insect's total weight. The secretions are hydrocarbons (some 60 percent decyl acetate and the rest dodecyl and tetradecyl acetates) that are unusually slow to evaporate. Regnier and Wilson observe that the French naturalist Pierre Huber, who discovered ant slavery in 1810, wrote: "One of the principal features [of raiding] seems to consist in exciting fear, and this effect is so strong that [the attacked ants] never return to their besieged nest, even when the oppressors have retired." Speculating that the panic observed by Huber and others might be caused by the raiders' pheromones, Regnier and Wilson performed chromatographic analyses of three kinds of extract from whole ants: extracts from raider ants, extracts from ants that had been raided and extracts from undisturbed ants of the raided species. None of the three chromatographic peaks associated with the acetates in the extracts from raiders were detected in the extracts from undisturbed ants. All three peaks showed up in the extracts from raided ants. The raided ants have their own pheromones to spread alarm when they are attacked, but Regnier and Wilson calculate that the raiders spray them with so much additional pheromone that they are alarmed even more than usual and retreat in confusion.

Ticklish Question

Try to tickle yourself. You cannot. Darwin suggested that the reason is that "the precise point to be touched must not be known" if a tickle is truly to tickle, but according to three investigators in England most children can be tickled even if they know where and when the stimulus is to be applied. L. Weiskrantz and C. Darlington of the University of Oxford and J. Elliott of the University of Sheffield undertook to find out specifically what knowledge is necessary for the "cancellation" of the ticklish sensation. They built a foot-sole tickler: a box with a slotted top through which a plastic pointer, counterweighted to maintain a constant upward pressure, protruded. The subject placed his bare foot on the box. The pointer could be moved the length of the slot, and thus about four inches along the sole of the foot, by means of a handle. The handle was operated either by the experimenter, by the subject himself or by the experimenter with the subject's hand on the handle so that his arm followed the movement passively. The 30 subjects (undergraduates) were divided into three groups of 10; the members of each group were tickled under two of the three conditions (externally administered, self-administered and passive) and the subjects were asked to compare the ticklishness of the two situations.

The investigators report in *Nature* that all subjects in the relevant groups found that being tickled by the experimenter was more ticklish than tickling themselves; all but one found being tickled by the experimenter more ticklish than the passive situation; most of the subjects found the passive condition more ticklish than tickling themselves. Two types of knowledge are involved in a self-administered tickle: the "command" to tickle and the feedback information provided by arm movement during the operation. Clearly cancellation of ticklishness is not due solely to the command information. The authors propose further experiments to determine how much ability to predict the fine timing of the tickle is imparted by the passive arm movement, and to ascertain the effect of such prediction. The whole question is not necessarily trivial, they point out. Perhaps pain is subject to similar forms of cancellation.



Some things are changing for the better.

Updating the ancient art of surveying with a beam of infrared light.

The surveyor's way of measuring distance—a method that had not really changed since the days of the Pharaohs—is finally changing for the better.

The painstakingly slow and centuries-old method of using an engineer's chain or surveyor's tape as a ruler is no longer the only accurate way to do the job. A new HP instrument called the 3800A Distance Meter now makes it possible for the surveyor to be free of his ancient chain.

As a result, a surveyor can precisely measure distances up to two miles in one-tenth the time, with a crew of two rather than three men. The advantages of electronic distance measuring are obvious: it takes fewer people to do more work, and they do it with first order accuracy.

Basically, the HP Distance Meter works by comparing the unknown distance being measured to the precisely known wavelength of a modulated beam of infrared light. The 3800A measures the length of time it takes the light beam to travel to a reflector and back, converts this elapsed time into distance, and produces the measurement—in feet or meters—on a digital display.

The 3800A's unique method of modulating the light beam with four different frequencies provides unambiguous measurements, internally corrected for air temperature and pressure, with an accuracy of 0.01 feet. Since the 3800A automatically calibrates distance 30 times a second, readings are not affected by momentary interruptions caused by traffic or pedestrians moving through the light beam.

While designed primarily for surveying and photogrammetric applications, the 3800A also can be used to detect the sway of buildings, the movements of glaciers and other tasks requiring accurate distance measurement.

The new instrument weighs only 34 pounds with its battery power pack, and costs $4,110.

Affordable radar—for the family car, and many other uses.

A new technology within a technology has made radar feasible for the family car and a host of brand-new uses in such diverse fields as industry, safety, navigation and security.

The HP 35200A Doppler Radar Module, about the size of a deck of cards and one-third the cost of previous radar units, can be produced in volume for less than $100. This makes radar inexpensive enough to perform motion-detecting and velocity-measuring tasks that formerly were impractical because of size and cost limitations.

Auto safety engineers, for example, can seriously contemplate designing a radar system to detect approaching vehicles and automatically avoid collisions. The six-ounce HP radar module has a range of some 2,000 feet and is capable of measuring speed within one-hundredth of a percent. That's the equivalent of clocking a car speeding 1,000 miles an hour with an accuracy of 1 mph.

But automobile radar is just one of many possible applications of the HP device. Others include navigational radar for boats, automatic landing systems for small aircraft, intrusion alarms, speedometers for rapid transit railroad trains, and systems for controlling the flow of traffic. Industrial uses range from regulating the speed of assembly lines to controlling the flow of steel through rolling mills.

These innovations are possible because of a successful marriage of two electronic technologies. HP combined the low cost and high reliability of integrated circuits (IC's) with the performance advantages of microwave frequencies. The resulting products are called "thin film hybrid IC's."

The microwave component capability which led to the radar module has been a reality at HP for more than two years. Details on either our thin film capability or the 35200A Doppler Radar Module are yours for the asking.

Around-the-clock analysis through automation: instantly, or a step at a time.

Gas chromatography, a process whereby a vaporized organic material is separated into its constituent compounds, is the fastest and most accurate way to reveal both the identity and exact quantity of the separated parts of a mixture. Therefore, a gas chromatograph (GC) is the most important single analytical tool at the chemist's command. But, since most laboratories operate on an eight-hour day, the typical GC sits idle the majority of the time.

That was yesterday. Now a modular series of HP instruments can enable a GC lab to operate 24 hours a day, without increasing its staff. And this can be done a step at a time, as the flow of work and budgetary limitations allow.

The first step is the HP 7670A Automatic Sampler. By injecting up to 36 consecutive samples, it can keep a GC going around the clock. So a lab can make well over 200 runs a week, and significantly increase its price/performance ratio.

Manual GC tests take 30 minutes, and cost about $5.00 for the operator's time. Adding the next building block—a 3370A Integrator—automatically measures, computes, and prints out a chromatograph report in 20 minutes at an operator cost of some $3.33.

The third and most decisive step toward full automation is a 3360A GC Data Processing System. This uses an HP computer, fully programmed for GC, which can simultaneously process data from eight GC's and immediately produce a complete typewritten report of each analysis. A full GC test report can be written in one minute, at the dramatically low cost of only 17 cents.

The system boosts a GC lab's capability to 6,000 samples a month. Even if you use it for as few as 1,000 samples per month, it saves more than 300 hours for your staff and pays for itself in just over six months.

The ability to make these analyses faster, and at less cost, is a significant contribution to the fields of pollution control, quality assurance, process control, and analyses in the fields of chemistry, agriculture, food, medicine, and drugs. If you'd like more information on the economics of HP's step-by-step automation of GC labs, write for the Fall 1970 issue of Analytical Advances.

For information on the technological advances or products mentioned in this ad, address your inquiry to Hewlett-Packard, 1505 Page Mill Road, Palo Alto, California 94304; Europe: 1217 Meyrin-Geneva, Switzerland.

00112

HEWLETT hp PACKARD

Measurement, Analysis, Computation

PATHWAYS IN THE BRAIN

New techniques combining various staining methods with electron and light microscopy make it possible to investigate in detail the connections among nerve cells and the circuitry of the brain

by Lennart Heimer

The functioning of the brain plainly depends on a system of physical organization of its billions of nerve cells—a wiring plan, so to speak. The cells and the brain's various functional centers are linked together by an elaborate circuitry of pathways and interconnections forming a network of communication. In order to arrive at a detailed understanding of how the brain works we need a clear knowledge of this wiring diagram. Obviously the diagram itself could not explain the workings of the human mind, but a meaningful picture of the wiring system is a prerequisite for such understanding, and for more than a century investigators have been seeking to map the nervous pathways by studying sections of nerve tissue. Within the past few years a combination of new techniques, involving the combined use of the light microscope and the electron microscope, has made it possible to explore the circuitry of the brain in detail.

Communication in the nervous system consists in the transmission of electric signals that are relayed from cell to cell. Consider first the structural apparatus of the cell itself. It receives information through synapses (points of contact) on its cell body and on dendrites (twiglike branches) located close to the cell body. The nerve cell integrates this information and then fires impulses along its axon (a single nerve fiber), which ramifies into branches ending in boutons (synaptic bulbs) that pass the signals on across synapses to other nerve cells [*see top illustration on page 51*]. Nerve cells are generally clustered or bundled together in groups, and the arrangement that particularly interests us here is one in which the axons are aligned side by side in a bundle that forms a pathway. Some pathways are very short, connecting only the nerve cells clustered in a group known as a nucleus. Others are somewhat longer, connecting cell groups several millimeters apart. The pathways extending from the brain to the lower spinal cord are some two feet long, and from the spinal cord other long pathways go out to the body's sensory and motor organs.

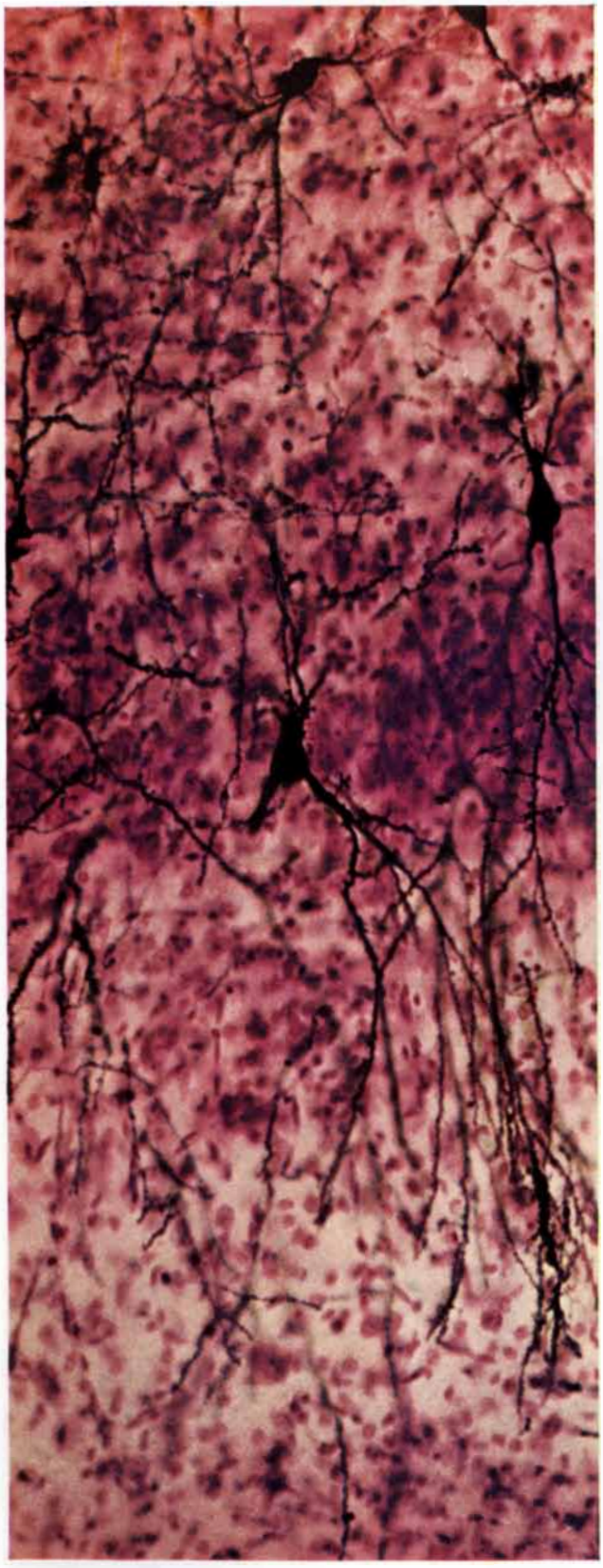
The existence of structural pathways in the brain was recognized many years ago. Some of them, massive bundles that show up as glistening white fibers because of the myelin sheath coating the axons, are visible even without a microscope. The simple outward appearance of the pathways clothes a stupendous complexity. It is now known that the nerve cells and their interconnections constitute a mass of structures so complicated that if it were pictured in full detail it would look like a dense jungle whose components would be impossible to disentangle.

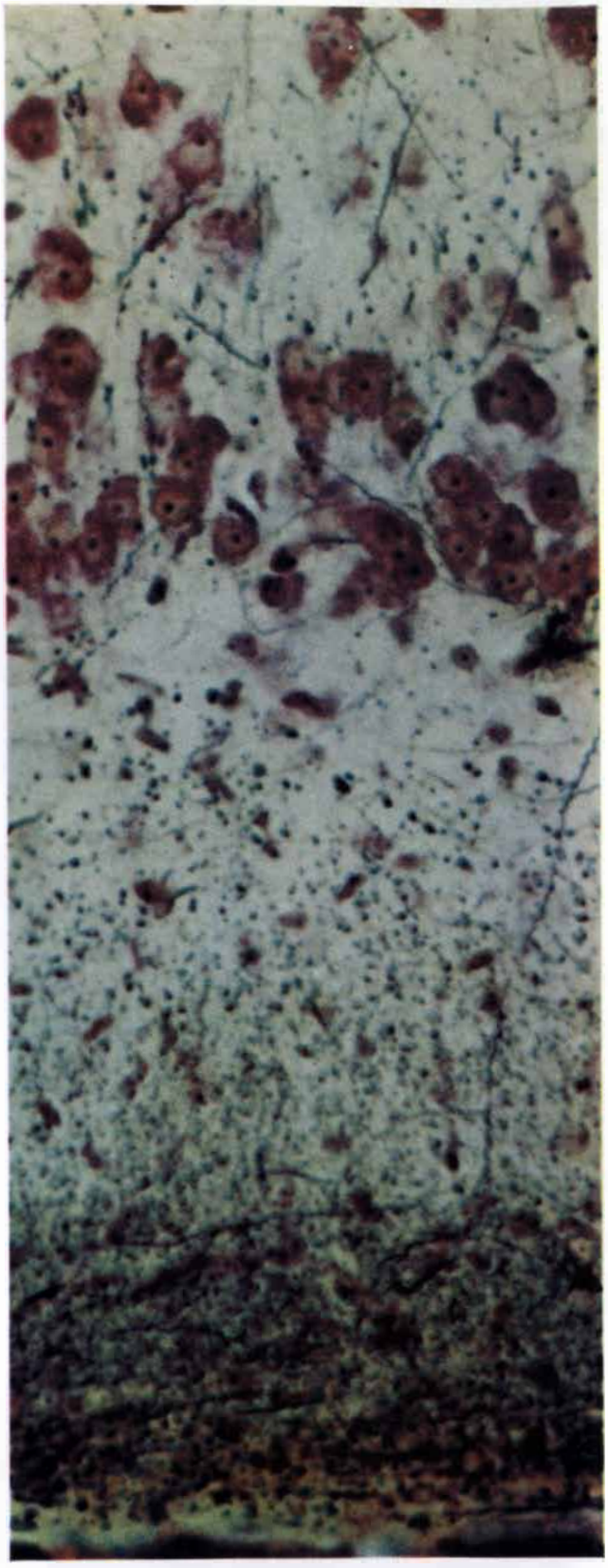
The Classic Stains

The disentanglement of this jungle began when investigators found selective stains that picked out individual features of the nerve cell for viewing under the microscope. In 1884 Franz Nissl of Germany discovered that the cytoplasm of the nerve-cell body has an affinity for methylene blue, so that this stain (and other alkaline dyes employed since) selectively shows the cell body. The Nissl method made it possible to outline and identify well-defined regions in the brain. At about the same time Camillo Golgi of Italy developed his method for the selective staining of certain nerve cells by the use of silver salts.

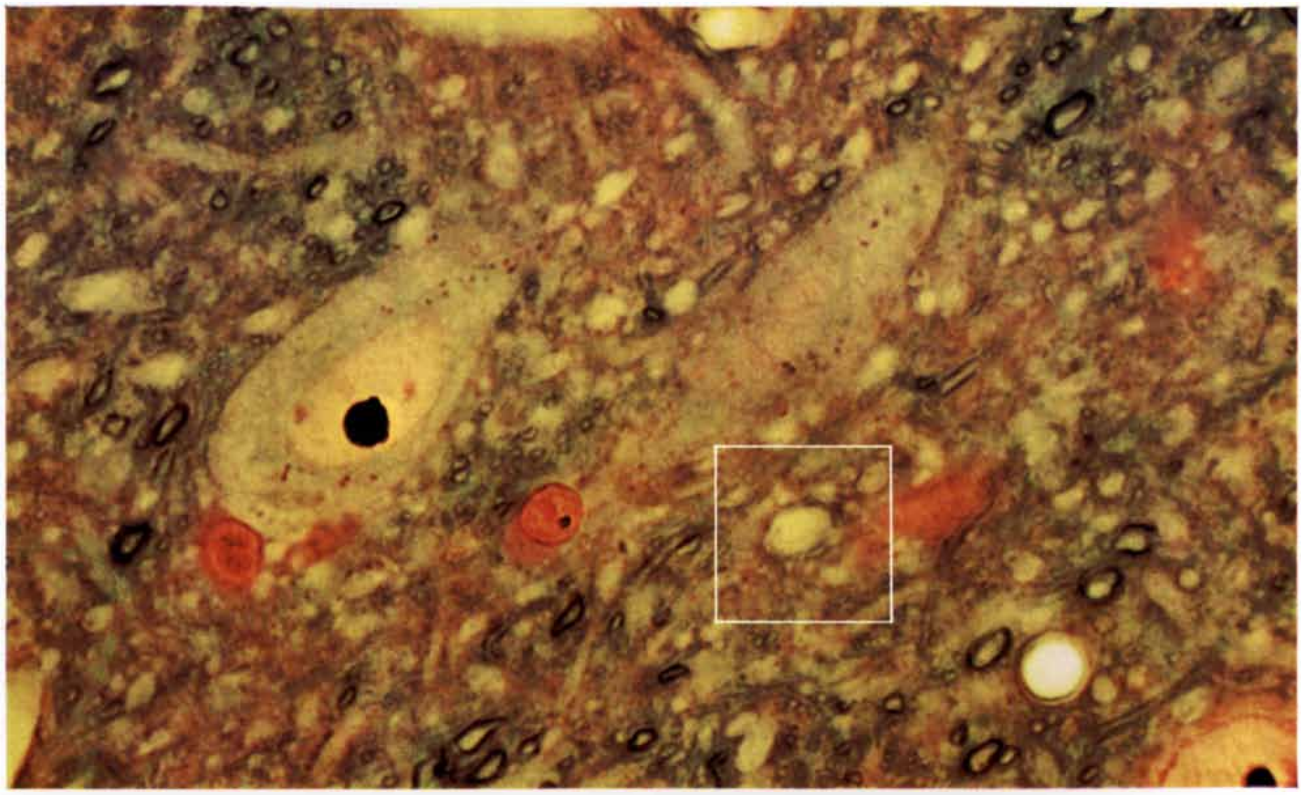
Golgi's staining method shows most of a nerve cell's details, including the dendrites and the axon. If the stain impregnated all the cells in a section of appreciable thickness, the details would be lost in a dense tangle of fibers. Because the Golgi method stains only a small percentage of the cells, however, it is possible to resolve the details of these cells' structures and connections in a comparatively thick section of brain tissue, as thick as a few tenths of a millimeter. (With other staining techniques one is limited to thinner slices—a few hundredths of a millimeter—for meaningful examination of the cell processes and connections.) Thanks to the thickness of the Golgi section, a nerve cell's processes can be traced for a considerable distance, often the entire length of a short pathway connecting two clusters of nerve cells. The unique advantages of the Golgi method were soon exploited by many investigators, most notably by Santiago Ramón y Cajal of Spain. His host of drawings of nerve cells and their interconnections in various regions of the brain still provide a structural base for many of our functional concepts. Ramón y Cajal shared the 1906 Nobel prize in physiology and medicine with Golgi.

At best, a picture of a stained section in the optical microscope shows only bits and pieces of the nerve tissue. A great step forward in defining the structures that make up the pathways of the brain came when the electron microscope was brought to bear on this problem. That instrument can display, with a high degree of magnification and resolution, the details of the individual elements that constitute the complex nerve cell and the interconnections among cells. The electron microscope, limited to a slice of tissue less than a tenth of a micron thick, gives only an extremely small field of view, so small that it would be futile to try to trace even the shortest pathway by putting together a series of electron micrographs. The instrument is very effective, however, in delineating the fine structure of objects within this

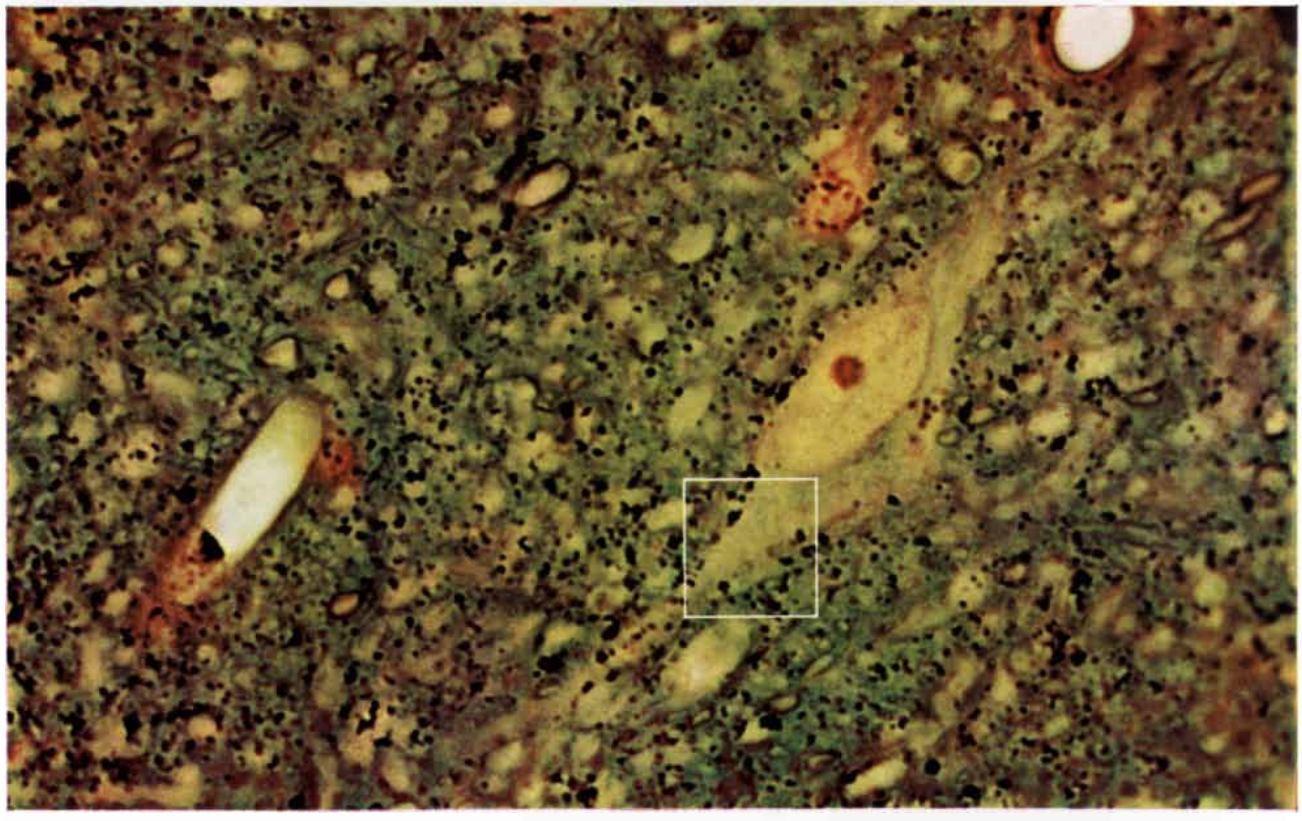




OLFACTORY CORTEX of the brain is seen in photomicrographs of sections stained by the Golgi method (*left*) and the reduced-silver method (*right*) and counterstained with cresylechtviolet. The Golgi stain (*black*) allows one to trace a few cells and their processes through a section of tissue. The Golgi section, prepared by Enrique Ramón-Moliner, shows that some of the dendrites (impulse-receiving processes) of large cortical cells (*center*) extend toward the lower surface of the brain (*bottom*). The reduced-silver method visualizes axons, or nerve fibers (*dark brown*), that have degenerated because their cell bodies were destroyed by an experimental lesion in the olfactory bulb. Five days after the lesion was made the animal was sacrificed and the brain section prepared. Degenerating axons cross near the surface (*bottom*). Above them is a zone of terminal degeneration. Large cortical cells are near top.



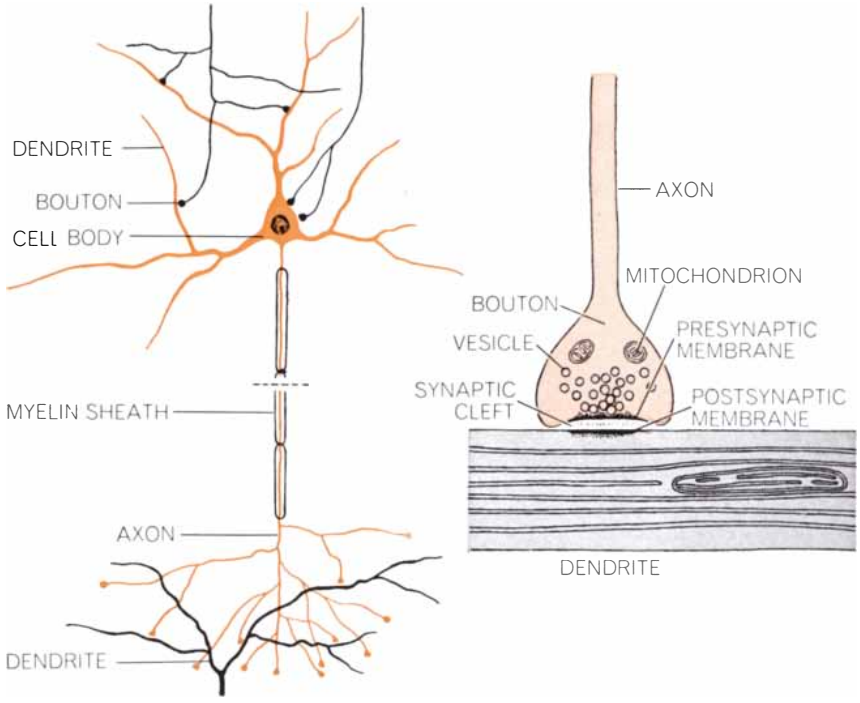
NERVE CELLS AND PROCESSES are yellow in a photomicrograph of a silver-stained section of the substantia innominata, an area where axons from the olfactory tubercle terminate. The rectangle marks a sectioned dendrite (*see illustrations on page 57*).



SIMILAR SECTION from the substantia innominata was prepared (after a survival time of two days) following a lesion in the olfactory tubercle, part of the olfactory cortex. Large silver particles (*black*) indicate degenerating boutons (*axon terminals*) whose cell bodies were destroyed by the lesion. The electron micrograph on page 60 is from a region similar to the one in the white rectangle.

field of view. For example, it can be employed to analyze the elements of the synaptic contact where information flows from one cell to the next.

The synapse is a focal point for the investigation of the functioning of the brain. It is here that the interactions and modifications of nervous impulses that are responsible for determining an animal's behavior take place. Analyses of the synaptic region with the electron microscope have already cast some light on the structure of this transmission apparatus. They have shown that in the nerve cells of the brain and the spinal cord the boutons from which signals are passed across the synapse usually contain a number of clear vesicles (fluid-filled sacs) between 300 and 600 angstroms in diameter. The vesicles are believed to contain a substance that serves as a chemical transmitter of the signal. Apparently when a nerve impulse arrives at the bouton it releases a small quantity of this substance, which then crosses a cleft 200 to 300 angstroms wide to the next cell and thereby delivers a signal that excites or inhibits that cell [see "The Synapse," by Sir John Eccles; SCIENTIFIC AMERICAN, January, 1965].

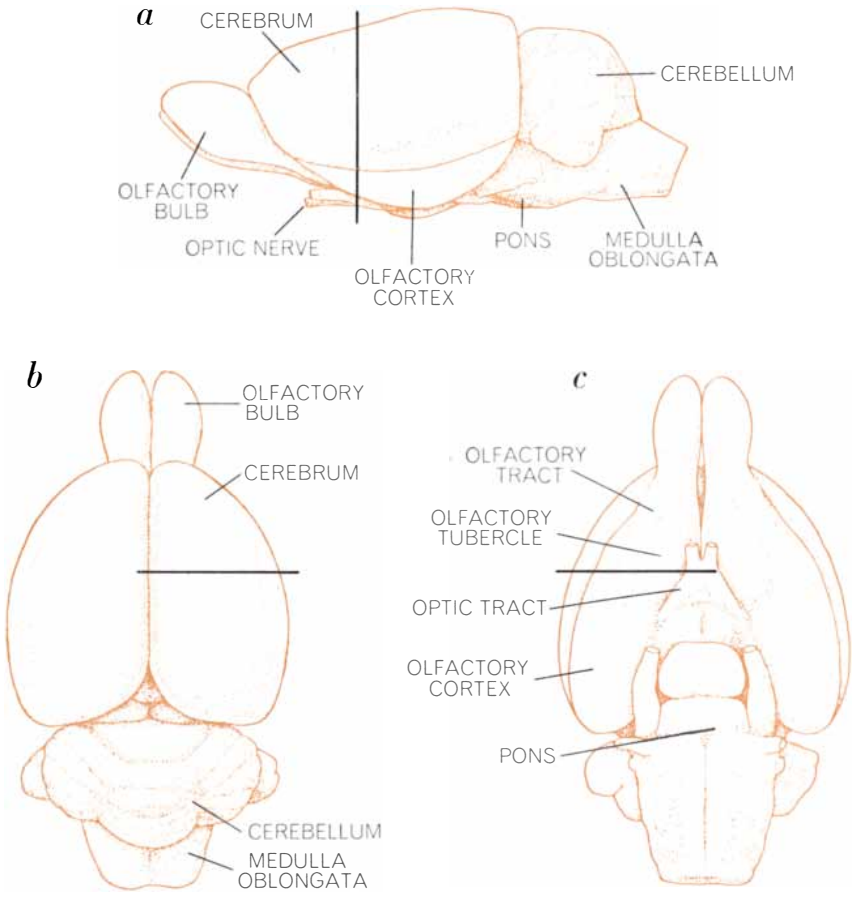


NERVE CELL and its processes (*color*) and parts of three other cells (*black*) are diagrammed at the left. A nerve impulse travels along an axon, which ramifies into a number of terminal branches that end in boutons that make contact primarily with the bodies or dendrites of other cells. Points of contact are synapses; one is diagrammed at the right.

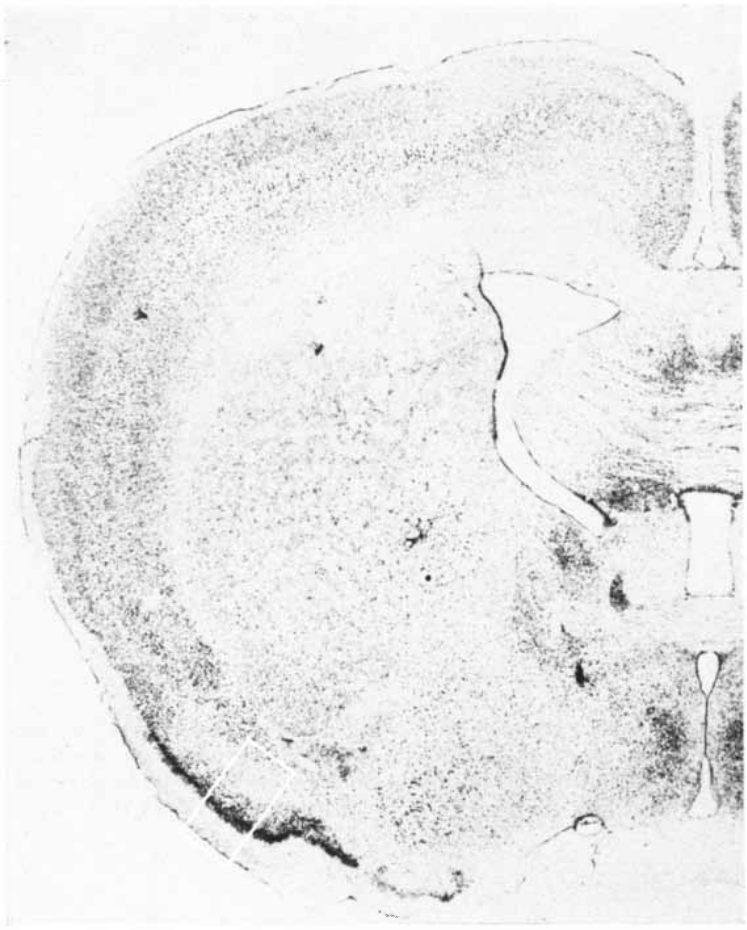
Tracing Pathways

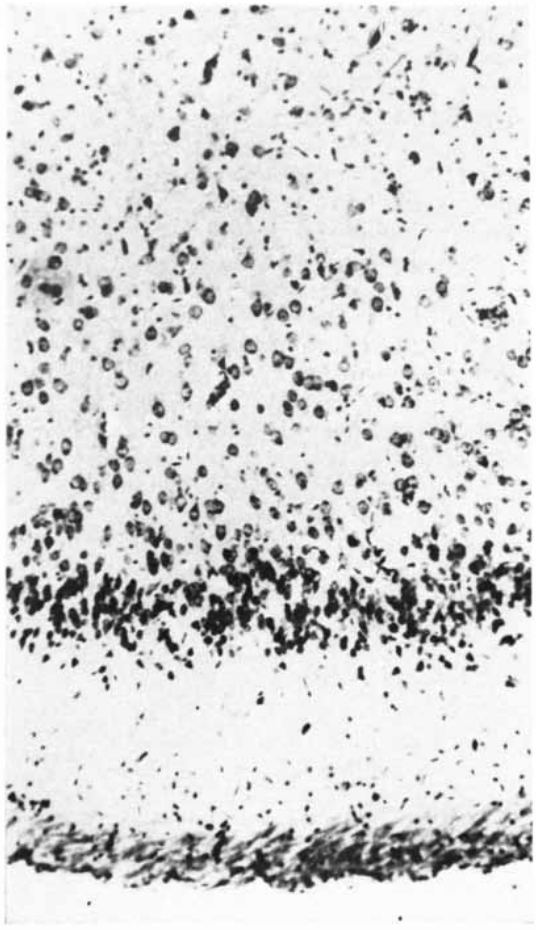
Together the electron microscope and the optical microscope, with the aid of special staining techniques, now provide powerful tools for plotting the nerve pathways. Serial pictures of silver-stained bundles of nerve fibers made with the optical microscope outline the routes of pathways, and the electron microscope reveals their details. Even with these tools, however, it is seldom possible to trace an entire pathway from its beginning to its end in normal tissue. Fortunately a discovery made more than a century ago provides a means of conducting such an exploration in experimental preparations.

In 1850 Augustus Volney Waller of England noticed in experiments on nerve tissue that when a nerve cell was damaged, the fibrous structure ahead of the lesion degenerated, breaking down into fragments. Then in the 1880's Vittorio Marchi and Giovanni Algeri of Italy found a way to selectively stain fibers that are undergoing "Wallerian degeneration." Their technique stains the myelin coat of such fibers; the myelin is first treated with a solution of potassium dichromate and then stained with osmium tetroxide. This discovery made it possible to trace a pathway by following the route of fibers whose myelin is break-

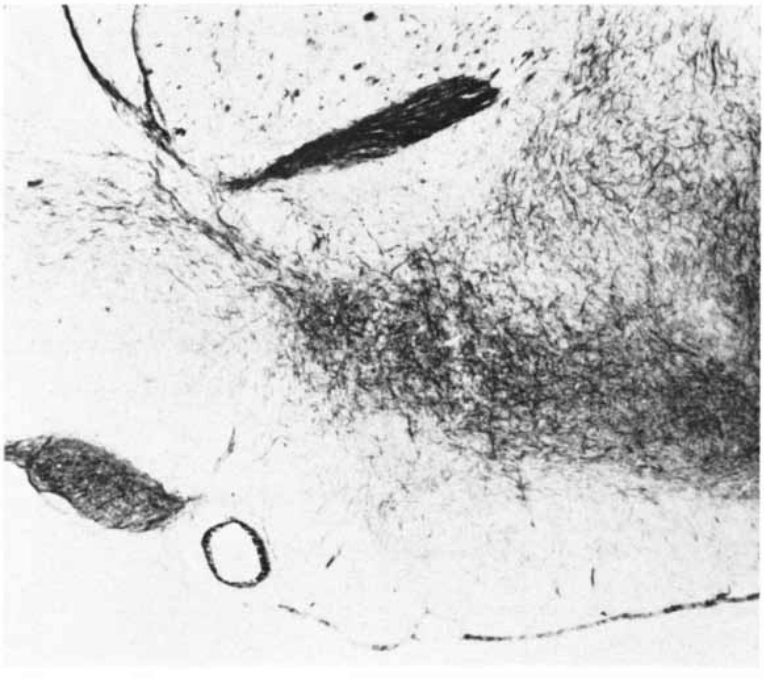


RAT BRAIN is seen from the side (*a*), the top (*b*) and the bottom (*c*). In each drawing the heavy black line shows the plane of the transverse sections in subsequent illustrations.





NISSL STAIN visualizes cell bodies, identifying cell "territories" in the brain. A low-power photomicrograph (*left*) of a transverse section of the rat brain, made by Walle J. H. Nauta and Harvey J. Karten, shows the layered arrangement of cells in the cortex, near the brain's surface. The packed cells in the olfactory cortex (*rectangle*) are enlarged 100 diameters in a photomicrograph (*right*).



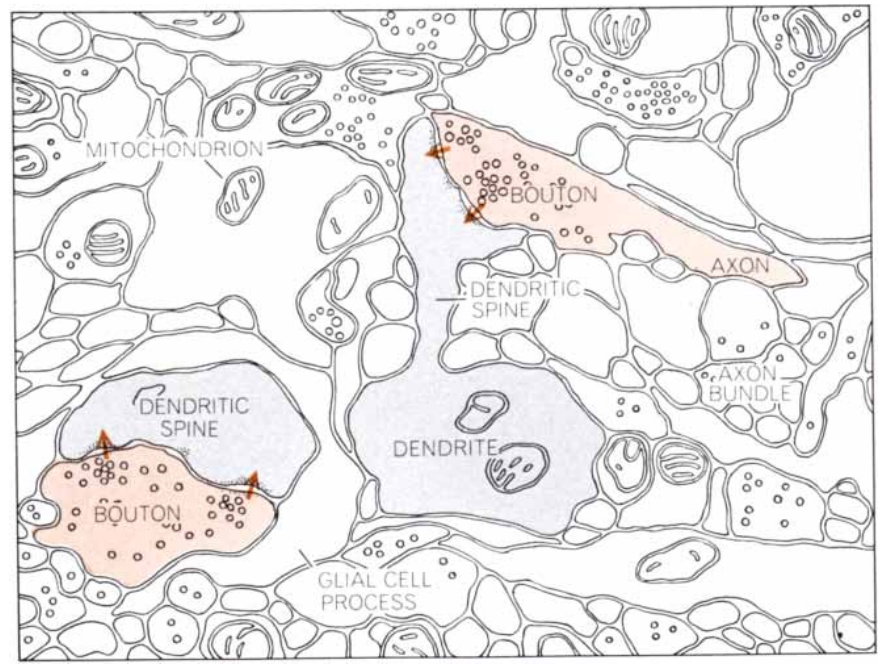
BIELSCHOWSKY STAIN is taken up mainly by large nerve fibers and is therefore effective in tracing massive pathways across a low-power micrograph (*left*). It stains all such fibers, however, so that individual fibers cannot easily be followed. When a portion of the basal part of the brain (*rectangle*) is enlarged 40 diameters, for example, complex network of fibers cannot be disentangled (*right*).

ing down. The technique still left important gaps in the picture because it failed to show those parts of the pathway that were not myelinated. This includes not only the terminal axon branches and their boutons but also many of the stem axons that are extremely thin and have no myelin sheath.

An important breakthrough came in the 1940's and 1950's, when refinements in the staining of nerve fibers with silver made it possible to mark and identify degenerating material in the interior of fibers. "Reduced silver" techniques for staining fibers had been developed at the turn of the century by Ramón y Cajal and by Max Bielschowsky in Germany, but for many years these methods were applied only to the study of normal tissues. Application of reduced-silver stains to degenerating fibers opened the way to the mapping of thinly myelinated and unmyelinated axons and boutons. The most widely employed technique was developed by Walle J. H. Nauta in collaboration with Paul A. Gygax, working first in Switzerland and later at the Walter Reed Army Institute for Research in Washington. The section is pretreated first with potassium permanganate and then with uranyl nitrate before being impregnated with a solution of silver nitrate. Thereafter it is transferred to an ammoniacal silver solution, and finally it is reduced in a dilute formalin–citric acid–alcohol solution. The technique not only highlights degenerating fibers but also suppresses the affinity of normal fibers for silver; thus it becomes easy to identify and trace fibers that are breaking down. The development of this convenient method for tracing nerve pathways through animal experiments launched a veritable boom in the exploration of the anatomy of the nervous system within the past decade, and the technique itself has been refined by a number of investigators, including the author in collaboration with Robert Fink.

Olfactory Pathways

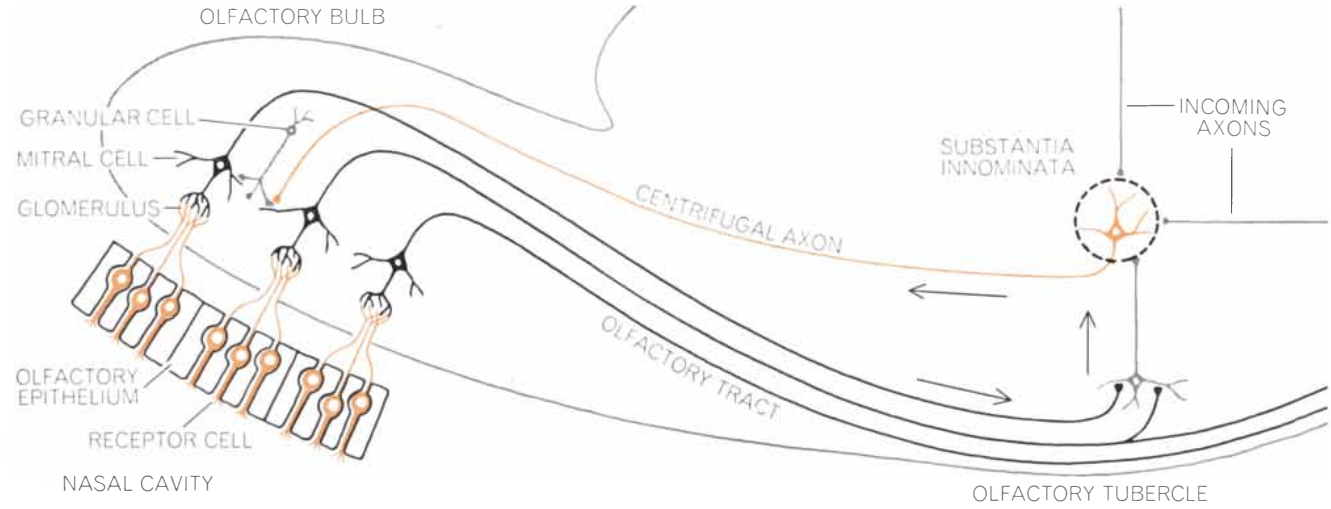
To illustrate the tactics and the nature of the findings in these explorations, I shall describe my own work on the brain's olfactory system. In all vertebrates, including man, the olfactory system seems to be built on essentially the same basic plan. Odorous substances in the air are detected by receptor cells situated in the upper part of the nasal cavity. These cells (of which there are several million) apparently are the only peripheral sensory receptors in vertebrates that are actually nerve cells. On



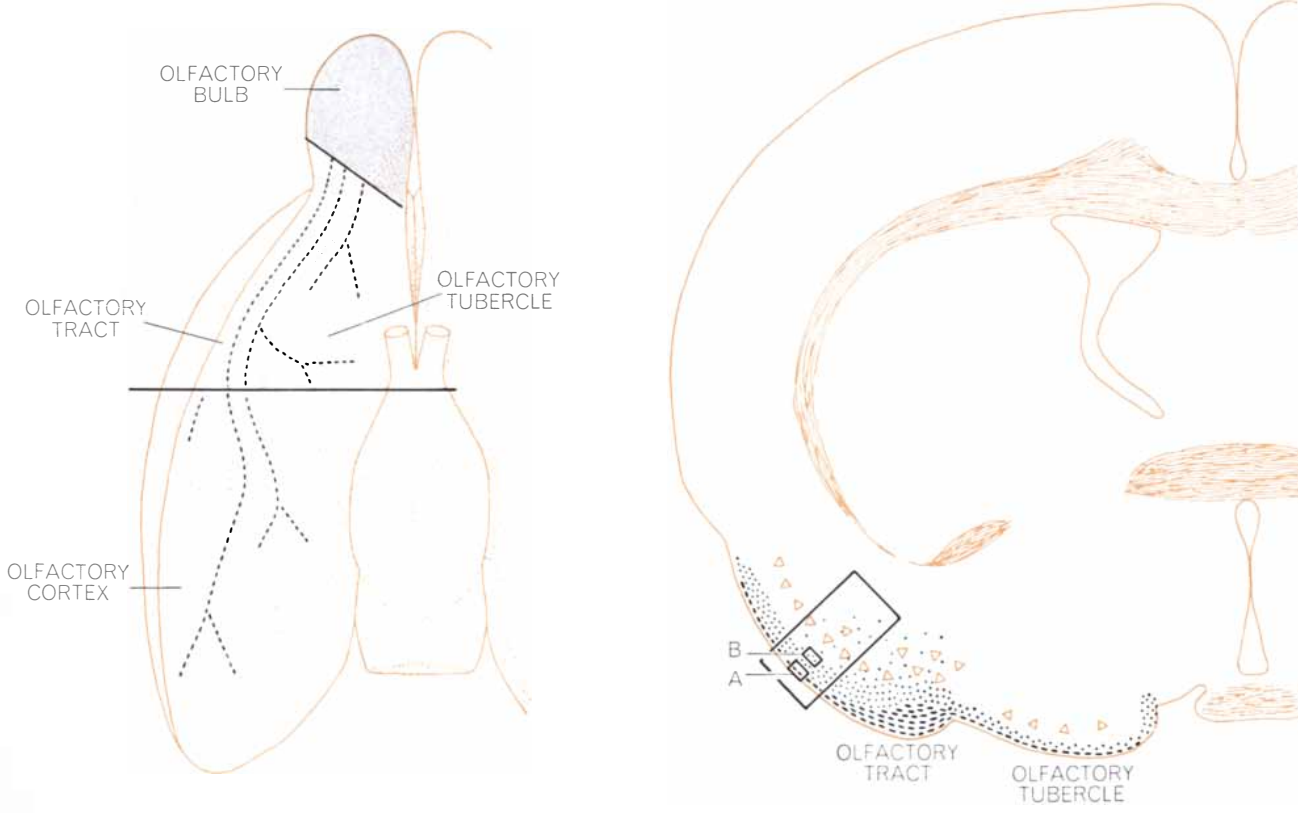
ELECTRON MICROGRAPH enlarges an olfactory cortex sample 24,000 diameters. The map identifies some components in the portion of the micrograph outlined in white. Arrows show the direction of transmission at synapses. Glial cells provide support for nerve cells.

receiving the stimulus, the cells' axons convey the resulting nerve impulses to the first way station, the brain's olfactory bulb. This is a complex structure from which, on the outgoing side, several thousand mitral cells (shaped like a bishop's miter) send messages on to olfactory centers in other parts of the brain. Additional cell types provide for intricate circuits within the bulb, and it is conceivable that the message sensed by the olfactory receptors is greatly modified in the bulb before being transmitted to other regions of the brain.

Let us for the moment pass over the anatomical arrangements within the bulb and concentrate on the pathway stemming from the mitral cells. Our chief interest in studying the olfactory system lies in finding out where its pathways terminate. We first undertook a de-



OLFACTORY SYSTEM of the rat has receptor cells in the nasal cavity that are stimulated by odorous substances and transmit messages to mitral cells in the olfactory bulb. Mitral-cell axons go to the olfactory cortex, which includes the olfactory tubercle. The author has found that olfactory tubercle cells communicate with cells in the substantia innominata. These may be the same cells that have been found to send axons from the substantia innominata to the granular cells of the olfactory bulb; the broken circle indicates the uncertainty. The possibility of such a connection suggests the existence of a feedback loop (*arrows*) that modifies sensory inputs to the olfactory system. Other incoming pathways converging on the substantia may also modify olfactory responses.



OLFACTORY PATHWAY is traced by the reduced-silver technique, which preferentially stains degenerating axons of mitral cells destroyed by a lesion in the olfactory bulb (*left*). The axons spread over the bottom surface of the cerebral hemisphere (*broken lines*) and appear to terminate just below the surface (*right*). The large rectangle shows the approximate site of the color photomicrograph at the right on page 49; the small rectangles (*A, B*) show the sites of the two photomicrographs on the opposite page.

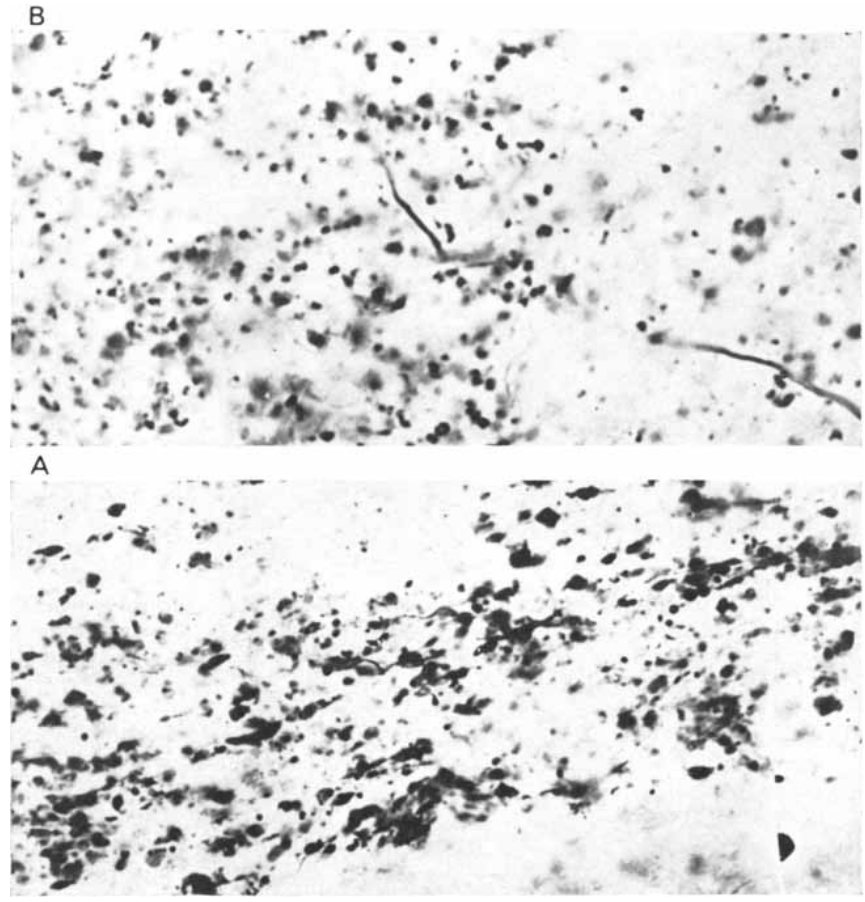
tailed study of the terminus of the mitral pathway, which has been the subject of a certain amount of controversy.

In rats we destroyed the mitral cells in the olfactory bulb by surgical intervention and, after a survival time of from three to five days, sacrificed the animal to conduct a microscopic examination of the fibers leading from these cells. Impregnated with silver by the Nauta-Gygax technique, the fibers showed a track of degeneration, marked by a line of droplike fragments connected by thin bridges, such as would be expected in a disintegrating fiber. Looking at a series of sections made consecutively through the brain, one can follow the degenerating mitral-cell axons, forming a pathway (the "olfactory tract") that is visible to the unaided eye as it curves around the olfactory tubercle. The pathway branches out and spreads its fan of disintegrating fibers over most of the cortical mantle of the basal part of the brain.

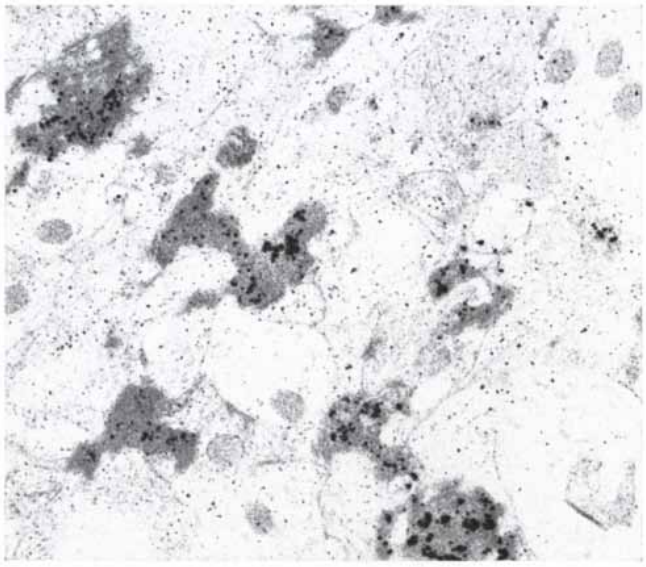
In this region, where the pathway terminates, one finds black, silver-bearing particles scattered about apparently at random [*see upper micrograph at right*]. Do they represent degenerating boutons? The particles do not show identifying features under the optical microscope. In collaboration with Alan Peters of Boston University we therefore examined silver-impregnated sections of this tissue with the electron microscope. The micrographs disclosed that many of the comparatively large, more or less spherical particles were indeed degenerating boutons. They could be identified as degenerating boutons by the density of their material, their shrunken and deformed shape and their content of vesicles or their association with a specialized membrane across a synapse. In the optical microscope the degenerating bouton shows up only as a black particle because of the microscope's lesser resolution, but in the electron microscope, where the thickness of the section is only a fifteenth of a micron, we are looking at just a small slice of the bouton—about a fifteenth to a thirtieth of its total diameter. There we can see that the degenerating bouton is only partly filled with silver, and we can make out details of its identifying structure [*see top illustration on next page*]. The electron micrographs also indicate that smaller dark particles found in the region where the mitral-cell pathway ends are fragments of the terminal branches of degenerating axons.

Let us now follow a pathway from the olfactory tubercle. Using the experimental procedure of creating superficial lesions in the tubercle with a silver plate heated to 70 degrees Celsius, we have traced the pathway of degenerating fibers to restricted regions in the thalamus and the hypothalamus. In addition we found that it terminates in the region called the substantia innominata, which is near (and apparently related to) the hypothalamus. A recent discovery by Joseph L. Price and Thomas P. S. Powell of the University of Oxford lends particular interest to this finding. They learned that the axons of certain large nerve cells in the substantia innominata are directed toward the olfactory bulb, where they communicate with what are called granular cells. Since the olfactory tubercle is known to be linked directly to the olfactory bulb by a pathway leading from the bulb, it seems entirely possible that return messages may be sent from the tubercle terminus to the bulb through a junction in the substantia innominata. If so, we can suppose that there is a feedback loop in which inputs originating from the bulb itself and traveling around through the tubercle and substantia innominata modify the bulb's activity. In order to decide this question we must find out whether or not the tubercle fibers "talk to" (communicate with) the dendrites of the same substantia innominata cells that send messages to the bulb. Whatever the exact organization of these connections may be, it seems certain that the brain can modulate the transmission of olfactory information as early as the level of the first synapse of the olfactory pathway.

The substantia innominata is rich in large dendrites, which traverse the region in various directions. Analysis with the electron microscope shows that these dendrites are covered with boutons of several different types, varying in the size of their vesicles, in the character of their synaptic membranes and in other respects. Studies of cells in other regions of the brain have indicated that a given type of bouton represents a specific pathway. This suggests that the presence of more than one type of bouton on a dendrite signifies the receipt of messages from more than one source or pathway. In short, the dendrites in the substantia



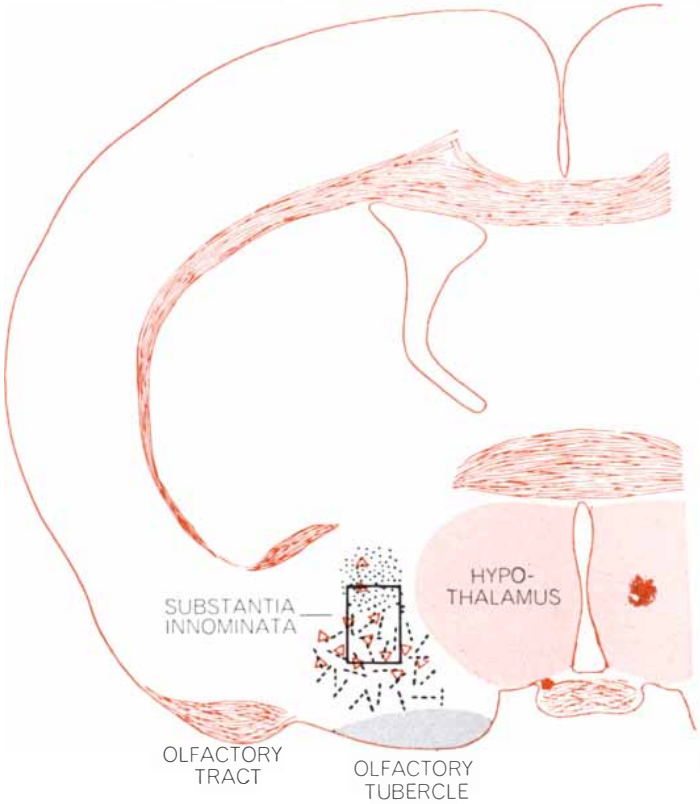
EVIDENCE for the pathway traced in the illustration at the bottom of the opposite page is in these photomicrographs of sections from the brain of an animal sacrificed five days after an olfactory-bulb lesion. Silver-stained degenerating axons are revealed as droplike fragments arranged in linear fashion (*A, bottom*). More spherical silver-stained particles in a deeper layer (*B, top*) were identified as degenerating axon branches and boutons.

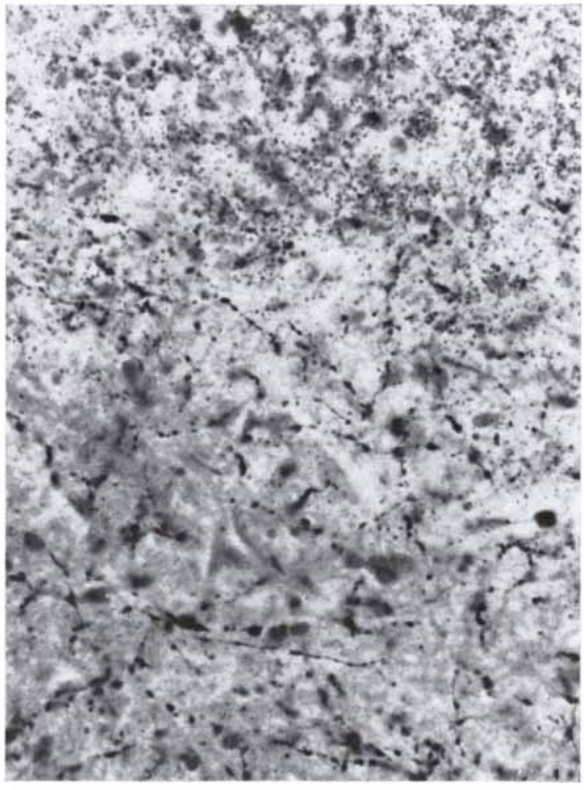


DEGENERATING BOUTON in the cortex is enlarged 25,000 diameters in an electron micrograph (*left*). Increased density and deformation are signs of degeneration. Similar boutons in a silver-stained section (three-day survival time) contain heavy accumulations of silver (*right*), proving that many spherical particles in photomicrographs are degenerating boutons of mitral-cell axons.

innominata may receive information not only from the olfactory tubercle but also from several other regions of the brain.

Our first step in investigating the question of the existence of a feedback loop was to try to identify the type of bouton representing the pathway from the olfactory tubercle. For this inquiry we needed a combination of information from the optical and the electron microscopes, and to avoid ambiguities the examinations had to be performed on one and the same specimen of tissue. This meant that we first had to cut a semithin section and stain it for observation in the optical microscope and then use part of the same block to cut sections thin enough for the electron microscope. The usual method of preparing thin sections of brain tissue, which is extremely soft, is to harden the tissue with a fixative





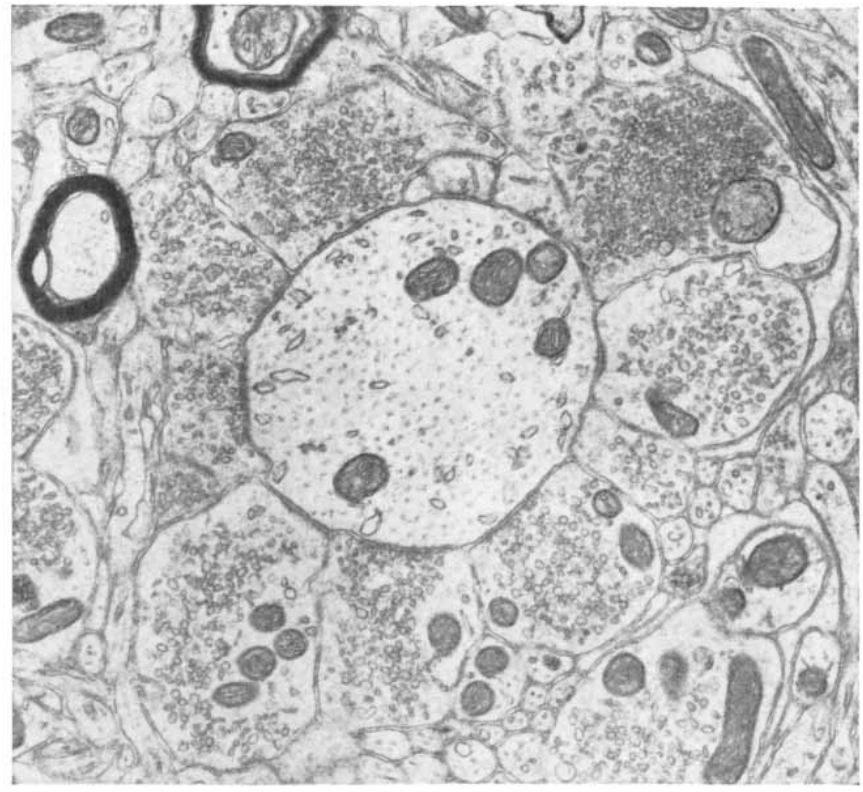
PATHWAY WAS TRACED FARTHER by making a lesion in the olfactory tubercle (*left*), causing degeneration of the axons of cells with which the mitral-cell axons synapse. The tubercle-cell fibers could be followed to the substantia innominata, where they terminate in a localized area (*rectangle*), as shown by boundary between region of degenerating fibers and terminal degeneration (*right*).

such as formalin and alcohol (which preserves the tissue's natural structure) and then freeze it or embed it in a solid material so that thin slices can be cut with a microtome. I had found that the silver-impregnation technique for identifying degenerating fibers and boutons could be applied to tissue embedded in plastic, which is essential for cutting ultrathin sections of the tissue. After examining a semithin section about five microns thick in the optical microscope to obtain a well-defined picture of the degenerating-fiber structure we had under study, we trimmed down the block from which the section had been cut so that the block was limited to the area of interest. From this block we then cut sections about a fifteenth of a micron thick for close analysis in the electron microscope.

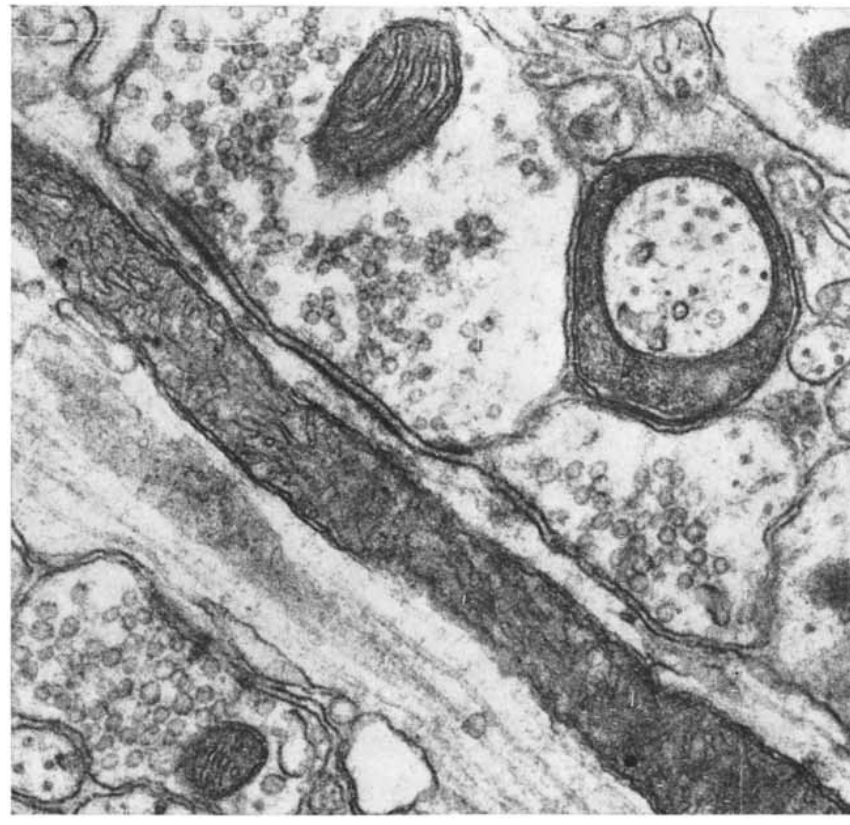
This analysis identified the type of bouton involved in the olfactory tubercle pathway. It indicated that the bouton probably was of a type that makes a "symmetric" synaptic contact, its membrane matching in appearance the membrane of the receiving cell structure facing it across the synaptic cleft.

Identifying the Receiving Cells

We are still left with the problem of identifying the substantia innominata cells that receive the messages coming from the olfactory tubercle. Are the receiving cells the ones that send their axons to the olfactory bulb, thus completing a feedback loop? It is very difficult to identify the cell itself, because of the electron microscope's narrow field of view. In the electron micrograph we are looking at only part of a dendrite; the picture is not broad enough to show the cell body to which the dendrite is attached. Fortunately a technique that promises to solve the problem is now being developed. It is based on use of the Golgi staining method, which shows the various processes of a nerve cell, dendrites as well as the cell body and the axon. William Stell of the National Institutes of Health and Theodor Blackstad of the University of Aarhus in Denmark independently found that after an entire nerve cell has been impregnated by the Golgi method to delineate its processes in silver, the block containing the cell can be trimmed down to a fragment showing only a part of the cell, and the fragment can then be sectioned for viewing in the electron microscope. Thus the stained material seen there (say a portion of dendrite) is known to belong to the cell that was stained and identified to begin with. There are still some diffi-



DENDRITE in substantia innominata (*see top illustration on page 50*), enlarged 25,000 diameters in an electron micrograph, is surrounded by a sheath of different types of boutons.



TWO DIFFERENT TYPES of boutons in the substantia innominata are enlarged 45,000 diameters in this electron micrograph. The large bouton in the upper part of the micrograph contains small synaptic vesicles; the other two boutons have large vesicles. The question is: Does either one of the bouton types belong to olfactory-tubercle cells?

(an informative advertisement)

An up-to-date report on Air Bags. The good news. The bad news.

We listen. In recent months, one aspect of the auto safety issue has caused a great deal of confusion. Namely, air bags.

Air bags are large, balloon-like devices that inflate, on impact, between the passenger and the instrument panel.

When they work correctly, they work beautifully. When they don't work correctly, they can be dangerous.

By August, 1973, an air bag system, or an equivalent, will be required by Federal law in the front compartment of every new car in this country.

Therein lies the problem.

Ford Motor Company is concerned that air bags will not have been sufficiently tested and proven for all car lines by then.

We believe more time is necessary to ensure reliability and reduce costs. And right now, the cost is high; reliability is low.

What the situation will be two years or two weeks from now, we can't predict. And frankly, the picture changes by the minute.

All this report hopes to do is bring you up to date. It concerns important pros and cons, and estimated prices at this moment. At the end, it asks you to make several choices. Send them to us.

The air bag may be a lifesaver—but it may be too much, too soon.

READ ON:

First of all, Ford Motor Company is not against auto safety. Fifteen years ago, the first auto safety package ever offered—including seat belts, padded dash and concavesteering wheel—came from Ford. And right now, we're urgently working to meet the Federal deadline.

Nor does it mean Ford Motor Company is against air bags—we're not. We're for any practical restraint system that will save lives.

For that very reason, Ford engineers are working on a number of restraint systems that may prove to be a better value than air bags.

But honestly, at this point, we don't know.

A FEW BASICS:

There are two kinds of auto restraint systems: active and passive.

The Active System:
Can be either voluntary or involuntary. An example of the voluntary active system is the belt-and-harness arrangement on all new cars today. An example of the involuntary concept is the belt-and-harness arrangement combined with either the Flasher or Buckle-Start systems (more about these later). In both the voluntary and involuntary systems, when the belt-and-harness arrangement is fastened, they are reliable and effective.

The Passive System:
Intended to operate independently of the will of the occupants. Required on all new cars by 1973. Air bags typify this system.

THE GOOD NEWS:

Unquestionably the appeal of any passive restraint concept is the idea of automatic protection. Potentially, it would be present at all times, no matter what the disposition of the occupants.

In theory, air bags are better than today's safety belts—not because they're

safer overall—but because as much as 60 percent of the time the belts are worthless for lack of use. And it's a fact only about 2 people in 5 use them.

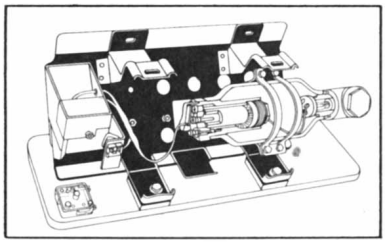
In theory, air bags can save thousands of lives a year. Ford Motor Company agrees. And, as one engineer puts it, "They're a good alternative to being killed."

But so much for theory. In the last two years, Ford has devoted intensive study to air bags. Several unanswered questions have arisen.

THE BAD NEWS:

1. *What about reliability?*—Tests show that in order to envelop a passenger in the event of a crash, an air bag must inflate to three times the size of a bedroom pillow, at a speed of 100 mph, in 40 thousandths of a second. To accomplish this requires a sudden blast of air or gas, triggered by a delicate sensor mechanism. Such a mechanism must be extremely reliable. In a recent test of 40 early production sensors, only three performed according to Ford Motor Company's reliability requirements.

2. *What about accidental discharge?*—The danger of an air bag being triggered accidentally is of serious concern. Imagine driving along at 60 mph and suddenly having an enormous pillow thrust in your face. The possibility also necessitates a high degree of reliability.



The air bag is inflated by this complex mechanism. Its reliability is questionable, thus far.

3. *What about multiple impact?*—The air bag begins deflating immediately after deployment. In the event the car hits other objects, following initial impact, the bag would offer no protection. Ford engineers are working on a reinflation device, but right now it's completely experimental.

4. *Will it injure people?*—When triggered, the air bag inflates with a sound like the blast of a shotgun. This may constitute a danger to people with heart conditions. Another major concern is that children who might be standing near the bag at the moment of deployment could also be severely injured. We're working on the problem, but we haven't solved it yet.

5. *Is the testing adequate?*—Thus far most air bag testing has been done with dummies or animals. Only a limited number of people have been used. How people in general will react is unknown.

6. *Who services it?*—What maintenance checks will be necessary to make certain of the air bag's effectiveness? Who does the checking? (The dealer? State inspectors? The corner gas station?) At this point, no one knows.

WHAT PRICE?

Present estimates show that the suggested retail price of a car equipped with air bag systems to meet the 1973 deadline may increase in the range of 100 dollars per car. By 1975, a passive restraint system will be required for back seat passengers. If that system is an air bag, it could add another $100 per car. Stay with us; the reading gets better.

OTHER POSSIBILITIES:

Is the air bag the only realistic approach to occupant protection?

Not necessarily.

The truth of the matter is that no one has the answer to the problem.

However, Ford is working on a number of promising possibilities for improving occupant protection. Here are some of them:

Buckle-Start System:
With it, your car won't start unless the seat belts are fastened by the driver and right-hand front seat passenger. Annoying perhaps, but a real lifesaver if it makes people wear belts. The Buckle-Start would cost roughly 1/3 of the price of an air bag system and would be simple to install.

Flasher System:
A red light goes on outside the car if your seat belt is not fastened. With the cooperation of Federal and State authorities in requiring people to wear their safety belts, the threat of getting a ticket could be a real incentive to conform. This system would be relatively simple and inexpensive. However, enforcement could be a problem, and the system does invite "tampering."

Static System:
Another approach under consideration involves a large, stationary padded "roll" running across the front of the car to "catch" and cushion an occupant upon impact. A system like this would be reliable and uncomplicated. However, it is too soon for Ford to determine whether the idea is feasible or not.

WHERE WE COME OUT:

At the outset certain conclusions are inevitable:

Indeed, air bag restraint systems are potentially beneficial. But more developmental time is needed: (a) to ensure reliability through exhaustive testing and (b) to prevent costs from becoming an excessive burden to the consumer.

We believe in cars that are engineered, not legislated. And as surely as we offer you a choice of options, accessories and colors in the car you buy, we tend to think you should be able to choose the kind of safety equipment you want.

Send us your thoughts. We listen. And we listen better.

Ford

...has a better idea (we listen better)

WHERE DO YOU COME OUT?

Clip this out and mail it to:

Ford Motor Company Listens
Department SA-AB
The American Road
Dearborn, Michigan 48121

Please answer the following:

1. Do you buckle your seat belts in your personal car?

☐ Always ☐ On long trips
☐ Most of the time ☐ Never

2. Do you fasten your shoulder strap?

☐ Regularly ☐ Seldom ☐ Never

3. Do you favor the Federal Government's requirement of a passive restraint system in 1974 model cars?

☐ Yes ☐ No ☐ Not sure

4. Which restraint system idea do you favor?

☐ Air bags ☐ Buckle-Start system
☐ Flasher system ☐ Static system

5. How much would you be willing to pay for such a restraint system?

☐ $25 ☐ $75 ☐ $100

6. Would you pay an additional $100 for a rear seat system?

☐ Yes ☐ No

Please give us any other comments you have on restraint systems. We would like to hear from you.

POSTLESION SECTION like the one in the photomicrograph at the bottom of page 50 was made 36 hours after a lesion in the tubercle. The electron micrograph shows a dendrite surrounded by a dark, degenerating bouton and several normal boutons. The degenerating bouton must belong to an axon coming from the damaged olfactory tubercle. Preliminary studies of this kind indicate that the olfactory-tubercle pathway is represented by large-vesicle boutons that make "symmetric" contact with dendrites of the substantia innominata cells, that is, the presynaptic and the postsynaptic membranes are of about equal density.

culties in applying this technique, but it will eventually help to answer the question of which cells in the substantia innominata are linked to the pathway from the olfactory tubercle.

It is a commonplace in science that, as Ramón y Cajal liked to remark, "discoveries are a function of the methods used." Certainly this truism is being demonstrated dramatically in the current closing in on the pathways of the brain. In addition to the Golgi technique, refined silver methods and the use of the electron microscope, which continue to be the mainstay of this work, other new methods for tracing pathways are available. There are ways of making the nerve cells in certain systems fluoresce. Autoradiography also shows great promise as a tract-tracing method. And there are other chemical techniques that will have increasingly important roles in the elucidation of neuronal systems.

No one expects that it will ever be possible to map in complete detail the entire nerve plexus of the brain, with its uncountable multitude of fibers and synaptic contacts. The prospects are now good, however, for obtaining the ground plan of many of the functionally significant circuits and finding out which of these pathways converge to modify behavior. To take one specific example, it will be very interesting to identify the network of pathways, originating in various regions of the brain and converging on centrifugal systems, that feed information into the olfactory bulb. The pattern of inputs from these various sources no doubt influences the output of impulses from the bulb, with consequent effects on those behaviors that are influenced by odor—a diversity of kinds of behavior that range from eating to mating. The behavioral effect of the odor of a spicy dish of meatballs, for instance, obviously depends a great deal on whether or not one is hungry.

Now, the oil with EDP

Over ten years ago, Texaco introduced 'Stop Loss'—an organized lubrication control system with limitless application.

Taking a lubrication system and adding EDP should be simple. But Texaco had to do more. You see, every company's system is unique. Computerize a paper mill's lubrication system and you've done one. Now you have trucking firms, pharmaceutical houses and cotton mills. And each system is different.

What Texaco did was to put our lubrication engineers and computer experts together to come up with a system which would easily adapt to any company, any plant, any process. And work exceptionally well.

And they did. We call it Texaco COMPUTEX. It took two years to complete.

Texaco COMPUTEX is a computerized lubrication control program that easily adapts to every industry.

Written in COBOL for the IBM System 360, the program matrix can be used with any COBOL compiler. Scheduling under the program can be done for all machinery by either date, mileage, time usage or throughput.

Texaco's unique knowledge about industries it serves now creates an interface of lubrication and computer technology. One more step to better service.

At Texaco it happens every day.

TEXACO

We're working to keep your trust.

How Polaroid went to unusual lengths to give

A simple, inexpensive way to make studio-quality portraits of people.

As long as there have been cameras, there has been photographic portraiture. Because nothing interests people more than people.

But until now, there has never been a portrait camera that gives the amateur what he needs: portability, simplicity of operation, high picture quality, and low cost.

Polaroid Corporation has directed its efforts for some time toward filling this gap. And has succeeded.

Our approach was to resist the tendency toward complexity; we strove to maintain singleness of purpose. And to achieve that purpose, we were as willing to use or modify existing principles as to design new features.

The result is that the new Polaroid Portrait Land Camera is the world's first and only portable, simple-to-use, low-cost camera that produces studio-quality portraits.

At the core of the problem was the dilemma of focal length vs. perspective distortion.

Pictures from conventional cameras, taken at close range, show pronounced distortion. Increasing the distance to the subject lessens distortion, but makes the image too small for a desirable portrait.

Techniques have been developed to try to overcome this problem. The telephoto lens is one; but it is expensive, and loses picture quality and corner illumination.

The most satisfactory solution has been the "studio" portrait camera. By using relatively simple lenses with a long focal length, and keeping them at full distance from the film, quality is maintained. But it is usually a heavy, expensive camera requiring careful adjustment.

With this background, Polaroid chose a 220mm, single-element lens, placed at full distance from the film plane. And to maintain lightness in the resulting elongated body, we molded it of a rigid though lightweight plastic. So the entire camera weighs under two pounds.

We added a shutter with a constant speed of 1/60 sec., and an aperture with a normal setting of f29.

We thus not only minimized perspective distortion,

people what they wanted but could never get.

but added enough depth of field so that precise subject distance is not required. And the shutter speed is optimal for conventional posed portraits.

Long focal length gives the same beauty as in the eye of the beholder.

Since the subject distance is fairly constant at 39 inches, we were able to design a unique combination viewfinder and split-image rangefinder with no moving parts. The user simply walks toward the subject until the two images coincide, and snaps the picture.

The entire mechanism was built into the camera handle, allowing a 5-inch base length. This compares with a maximum base of 2½ inches in most other cameras. And the greater the base length, the greater the accuracy.

Since one image must travel a greater distance than the other to reach the eyepiece, a compensating lens system, similar to a Galilean telescope, was built in.

And to maintain this high level of accuracy, adjusting set-screws allow each rangefinder to be individually calibrated at the factory for framing and ranging, and for avoiding other coincident image errors.

The rangefinder needs only 2 moving parts.

Duplicating the features of a studio camera was not enough. We needed studio lighting.

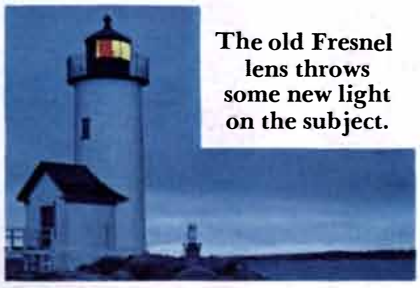
So we added a receptacle for a Magicube. It requires no battery. It yields four uniform flash shots without reloading. It will not advance when the fourth shot is finished, and the shutter locks to prevent taking another shot until the cube is replaced.

To utilize this burst of light to fullest advantage, we found a new way to apply an old principle.

In the 19th century, the Fresnel lens was developed to magnify the beam of a lighthouse without the great bulk of a conventionally-shaped lens. The theoretical surface of the lens was simply divided into small segments, all mounted in a single, flat plane.

Our modification of the Fresnel lens concentrates the light on the subject and transforms a small source of light (a tiny flashbulb) into a large area source.

And despite the concentration of the light, the effect is of controlled softness.



The old Fresnel lens throws some new light on the subject.

The camera has other interesting features. Such as a built-in timer to tell you when the picture is developed. A unique shutter mechanism, designed with a minimal number of moving parts, for higher reliability. And there are no exposure settings.

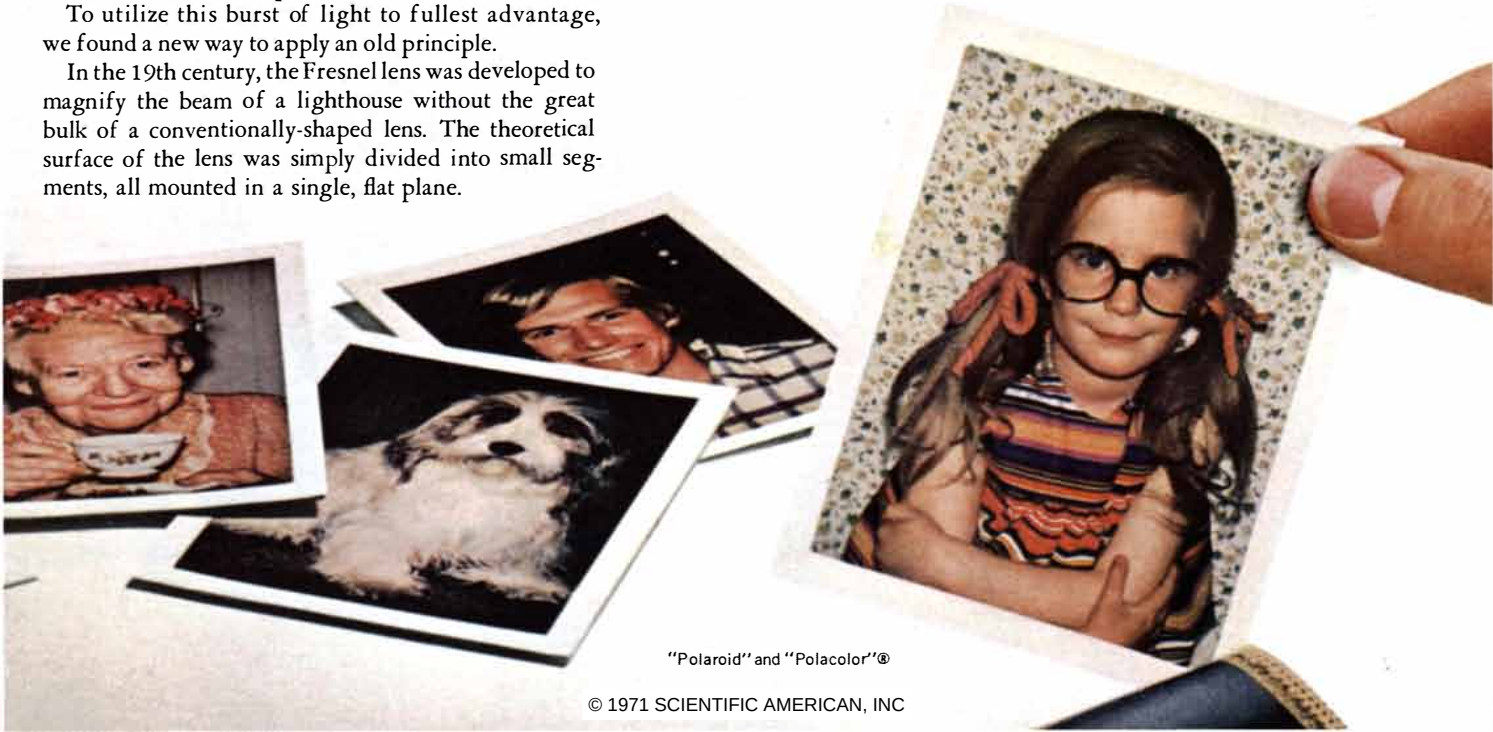
And, of course, it gives you a finished color portrait in a minute.

It uses only Polaroid Polacolor Land film (Type 108) in the big 3⅜" x 4¼" format that gives a big close-up portrait. Which is why we're marketing the camera under the trade name of Big Shot.

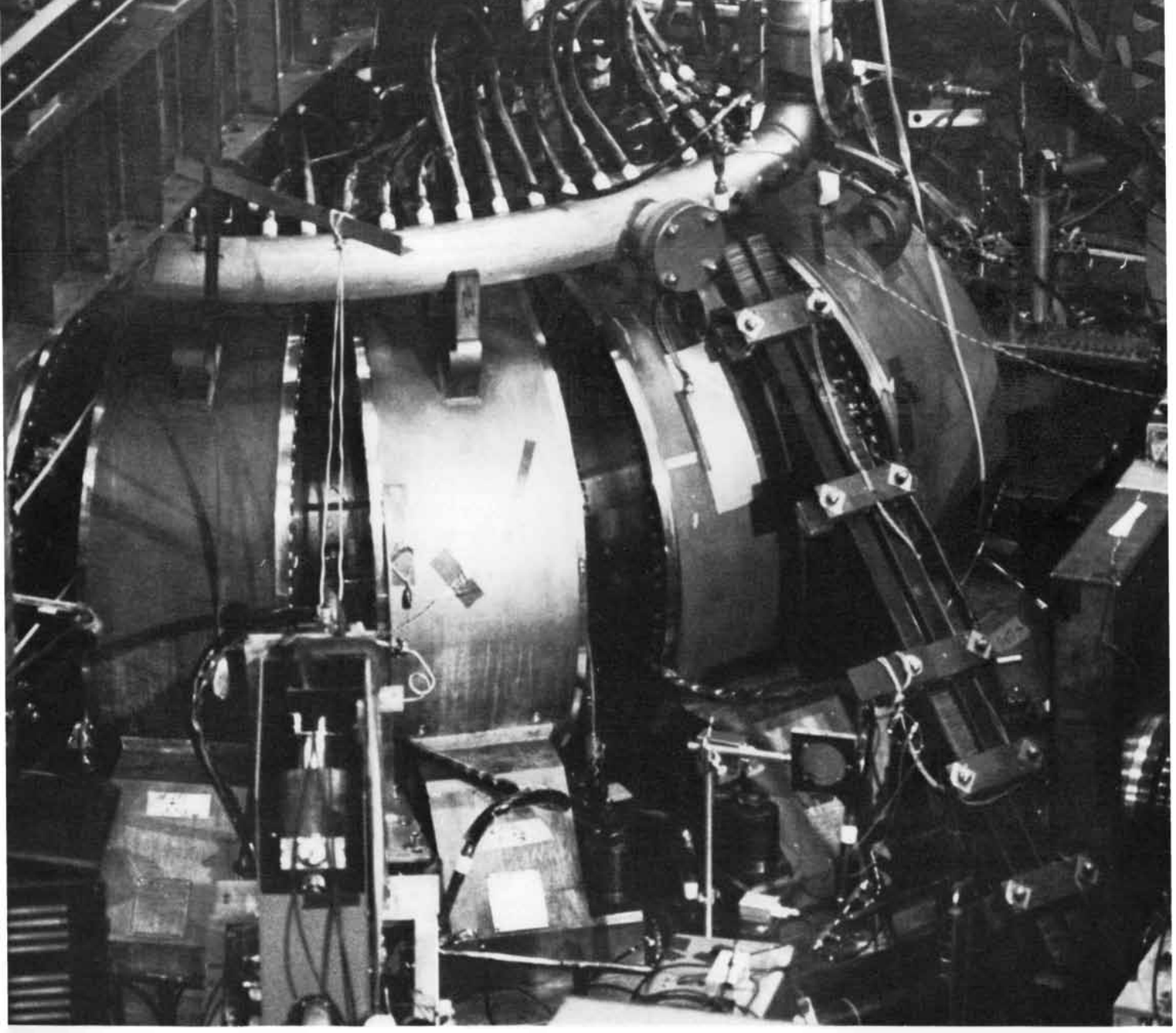
But what we're most proud of is the concept. The totality. Being first to fill a common need, and to fulfill the goals we set for ourselves.

Which puts what we do in proper perspective.

The Polaroid Portrait Land Camera. Under $20.



"Polaroid" and "Polacolor"®



Experimental fusion device, Princeton University.

Photographed by PHILLIP TRAGER

SCIENTIFIC AMERICAN

Will devote the entire September 1971 issue to

ENERGY and POWER

The Sacred Source of the Seine

The spring that gives rise to the river was once the sanctuary of Sequana, a Celtic goddess of healing. A unique accumulation of 2,000-year-old wood sculpture has been uncovered at the site

by Simone-Antoinette Deyts

For many cultures of ancient times springs were sacred places, perhaps because the phenomenon of water issuing from the earth without any apparent source seems magical. The Greeks and the Romans worshiped at springs and also prescribed bathing in them for the sick who were seeking divine help. France at the beginning of the Christian Era was the scene of similar practices. One spring in particular, located near the modern city of Dijon, was deemed sacred to the Celtic goddess Sequana, and pilgrims by the hundreds went there in search of restored health. The same spring is the source of the river Seine, which flows through Paris to the sea.

A century of archaeological investigation at the source of the Seine has revealed the remains of a temple complex, an abundance of offerings and a charm-

SCULPTURES IN WOOD portray a young man (*left*) and a young woman (*right*) who lived in Gaul around the time of Christ. The works are among 400 wood sculptures found in recent years at the sanctuary of Sequana. They were preserved by burial in wet soil.

ing bronze sculpture of Sequana herself. Within the past decade excavation of a marshy area near the temple complex has uncovered the largest single accumulation of Celtic sculpture in wood ever found: no fewer than 400 well-preserved effigy figures, large and small. Most are votive offerings: objects dedicated to Sequana. Among the offerings are larger-than-life representations of individual pilgrims hewn out of the trunks of oak trees. The discovery of these wood sculptures adds much to knowledge of the arts in ancient Gaul and to an understanding of Gaulish religious practices; it is the most important event of recent decades in the realm of Celtic archaeology.

A visitor leaving Paris in search of the sacred source of the Seine is well advised to avoid the main route to Dijon and to go by way of Troyes. This route will allow him to stop a little beyond Troyes at Châtillon-sur-Seine, where he will have an opportunity to rid himself of a few widespread misconceptions of ancient Gaul. Here he can see the Treasure of Vix, which includes a skillfully chiseled gold diadem and a five-foot bronze vase decorated with a handsome frieze of warriors.

The two objects are among several discovered in 1952 in the nearby grave of a Gaulish princess. Of Greek rather than native workmanship, they were made during the sixth century B.C. Their presence in the royal grave is evidence of two little-realized facts about Gaul before the Roman conquest. First, imports from Greece show that the Gauls maintained trade contacts with the Mediterranean world for centuries before the Romans came. Second, the high quality of the pieces attests to the sophistication of Gaulish taste in works of art.

Evidence of this kind discredits a common view of the pre-Roman Gauls as barbarians clad in skins and armed with frightful bludgeons. This schoolroom image of the Gaulish warrior is a legacy from the Romans, who had much trouble breaking Gaulish resistance at the end of the first century B.C. and thoroughly resented it. The falsity of the Roman cliché will become even more apparent as the visitor continues a few miles south on the road to Dijon and reaches the turnoff to the source of the Seine. Here he will find that the Gauls' supposedly barbaric heritage includes not only a taste for art but also a ritualized life of the spirit.

Do not expect an imposing spectacle. What meets the eye is a vale enclosing a small, well-kept park surrounded by firs and cypresses. Off to the right is a pool and a man-made grotto. Here stands a monument raised in 1867 by the citizens of Paris in honor of the stream that has contributed so much to the city's life. This pool is not, however, the sacred spring of Gaulish times. To find the spring one must cross the bottom of the park on the left until one meets a low, rocky cliff that forms one side of the vale. A few yards farther along flagstones mark the spot. The water rises between the flags, runs along a stone-lined channel to fill an ancient catchment basin—a carefully hollowed stone slab—and overflows to join the streamlet descending from the present-day pool.

"Clear water," writes the French philosopher Gaston Bachelard, "provides a temptingly simple symbol for purity. Man needs neither instruction nor tradition in order to find this a natural image." The symbol appears to hold true for all times and all places and it is a short step from the image of purity to the concept of a presiding deity. By a happy chance of preservation we are able to see exactly how the Gauls who worshiped here visualized Sequana, the divine personification of the Seine. Shortly before A.D. 300, probably when some invader was ravaging the countryside, the custodians of Sequana's temple removed the bronze image of the goddess, together with another sculpture, and buried them for safekeeping in a narrow pit at the foot of the cliff. The ancient hiding place was uncovered during a routine archaeological investigation in 1933; the find was promptly named the Treasure of the Seine.

The Gauls portrayed Sequana standing upright in a boat, her arms outstretched. The goddess's features are soft and yet filled with majesty; her openhanded gesture seems to offer welcome and help to all who approach her. The boat that carries her is a charming conceit: its prow is a duck's head, the bill holding a fruit, and its stern is an upswept duck's tail [*see bottom illustration on page 73*]. It is a bark eminently suited to carrying a river goddess. The sculpture appears to have been made during the second century.

Thanks to the work of generations of archaeologists since the middle of the 19th century, it is now possible to recreate the appearance of Sequana's shrine in its heyday and to trace the development of the temple complex from its beginning. Sometime around A.D. 50 the masonry channel leading

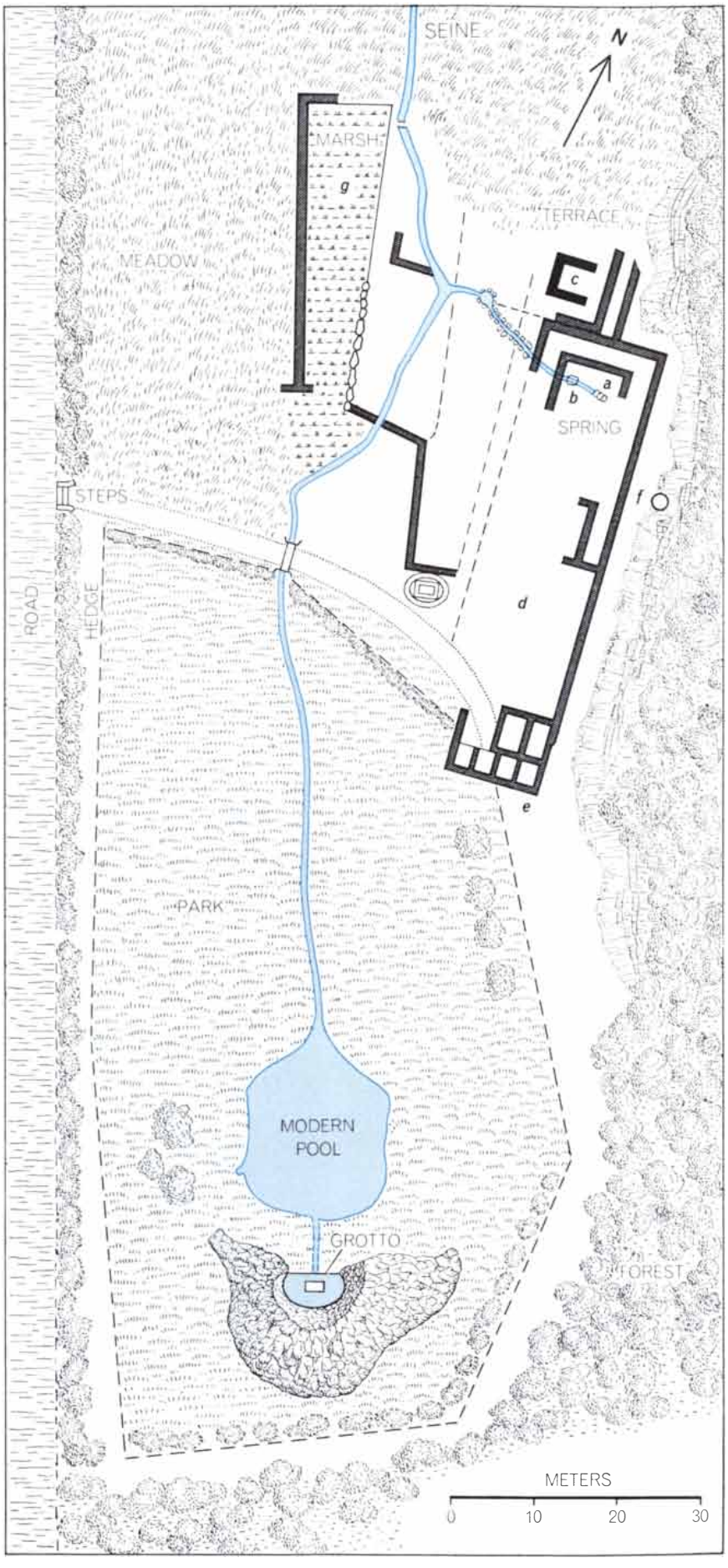
SOURCE OF THE SEINE lies north of Dijon on the Plateau de Langres, a highland area of Burgundy. The name of the river is derived from Sequana, who is its presiding deity.

from the spring was built and a small temple was erected some 30 feet farther north. The temple architecture was typically Gaulish; it owed nothing to Roman influences. It was a square chamber, presumably housing a figure of the goddess, surrounded by an open porch where worshipers could assemble. Archaeological evidence suggests that some of the healing ritual was conducted on lower ground to the west of the temple in a marshy area near the bank of the present stream. It is even possible that the marsh was a ritual site many years before the temple was raised.

Next a series of buildings was added, probably under Roman influence, on the higher ground between the catchment basin and the foot of the cliff. The most southerly of these buildings consisted of seven chambers: a row of four square rooms, each roughly 15 by 15 feet, and three larger oblong rooms. The four smaller chambers may have housed the priests who attended the sanctuary, and the three larger ones may have served for their dealings with the pilgrims—perhaps consultation or perhaps actual medical attention.

Farther to the north a monumental entryway of four pillars led the pilgrims into a long courtyard. The eastern side of the courtyard was occupied by a covered colonnade that sheltered a porch. The floor of the porch was plain masonry but the paving of the courtyard was a mosaic with a geometric design. At the northern end of the courtyard three steps led down to the spring, the channel from the spring and the catchment basin.

This grand plan, executed in a space some 300 feet long, came to light only gradually over a period of decades. At first almost nothing could be seen of it; most of the masonry, even the individual drums of stone that once made up the pillars of the colonnade, had been carted away over the centuries for local construction. It was not until 1932 that work in the marshy area west of the tem-



SHRINE OF SEQUANA, built around the source of the Seine (*a*), borders on a modern monument to the river (*bottom*) raised by the people of Paris. The shrine at first consisted of a catchment basin (*b*) below the spring and a simple temple nearby (*c*). Later, probably under Roman influence, a courtyard (*d*) was added, with a suite of priests' rooms at one end (*e*) and a covered porch along one side. The "Treasure of the Seine," which includes a sculpture of Sequana, was hidden beside a cliff just east of the site (*f*). Wood objects had been dumped in a marshy area (*g*) that was later filled in.

ple complex revealed three walls outlining a rectangular area dubbed "the pool" by the investigators. Later exploration revealed that although the area is a marsh now and was evidently a marsh in the earliest days of the sanctuary, it had been filled with earth sometime after A.D. 100 to form a second courtyard. It was from the bottom of the earth-filled "pool" that digging in 1963 and in 1966–1967 retrieved the unique assemblage of wood votive figures.

These remarkable sculptures are not the only votive offerings to be found at the sacred source of the Seine. In one of the priests' rooms, for example, early excavations uncovered a large earthenware pot bearing the Latin inscription "*Deae Sequana(e) Rufus donavit*" ("Rufus gives this to the goddess Sequana"). Inside the pot was a smaller one containing more than 800 bronze coins of the third and fourth century. Around the smaller pot was a pile of 120 diminutive bronze votive plaques. It is unlikely that the Rufus of the inscription had anything to do with what the pot contained at the time of its discovery; he would have had to make 120 offerings. Moreover, many of the plaques portray female ailments. The two pots and their contents must surely have been a stolen hoard of temple treasure, buried for temporary safekeeping by looters during a time of trouble.

TREE LIMB was made into the figure of a woman by confining the carving to the two extremities and leaving the rest untouched. The entire figure appears at left; the head, enlarged and seen in near-profile, is at right.

The formal Latin inscription on votive offerings is VSLM, an abbreviation of "*Votum solvit libens merito*" ("Willingly offers deserved homage" or, more idiomatically, "So-and-so fulfills his vow"). This inscription is missing from Rufus' pot, but it appears on several other objects found at the sanctuary of Sequana. "*De Sequana*," we read on one offering. "*Sienulla Vectii filia VSLM*" ("Sienulla, daughter of Vectius, fulfills her vow to Sequana").

The Gauls evidently felt that it honored the goddess more to address her in the language of their conquerors. Eleven Latin inscriptions have been found at the source of the Seine, although some of the Latin is erratic. In other inscriptions Latin characters were used but only to transliterate the native language. One five-line inscription appears on the front of a votive offering in Latin characters and on the back in Greek characters; the words on both sides are the same pure Gaulish.

So far some 1,000 stone or bronze votive offerings have been recovered at the sanctuary of Sequana. The stone offerings, carved from a locally available soft limestone, are the most numerous. They usually represent the afflicted part of the offerer's body, so that among them are effigies of arms, legs, hands, feet, heads, breasts and sexual organs, and even a few attempts to portray internal anatomy. Usually no effort was made to indicate the nature of the affliction, although some medical scholars assert that certain effigies portray goiter, hernia or even mental retardation.

In addition to these effigies of a part of the body there are stone busts and full figures, some as much as three feet tall. The individual represented, who was either seeking health or simply worshiping Sequana, is generally portrayed holding an offering of some kind: a dog, a rabbit, a piece of fruit or a purse heavy with money [*see illustration on page 72*].

The bronze plaques are quite a different story. They are thin pieces of metal, most of them less than two inches wide. They were evidently mass-produced by

being cut out of a larger sheet as they were needed; an image of a part of the body was then punched into the plaque. If a pilgrim's eyes were to be represented, round or oval outlines were punched; in some cases pupils and eyelashes were added. The eyes on a few plaques are simply holes, perhaps indicating that the donor was blind [*see top illustration on page 73*].

Sexual organs and breasts are the other parts that appear most frequently on the plaques. They too are stylized. The pelvis is a triangle; the male pelvis was marked with a phallus and the female pelvis was left unmarked. All representations of breasts were the same, although the donors sometimes added details.

When excavation of the "pool" revealed wood objects preserved in the waterlogged soil, we were fortunate in being able to draw on the experience of archaeologists who had developed various means of preserving such finds. If

VOTIVE OFFERINGS of three types are, from left to right, an arm, a portrayal of the body interior that emphasizes the stomach, and a series of heads. The maker of votive heads probably sold one at a time to pilgrims whose disability called for such an offering.

MAN IN A COWLED CAPE is one of several wood sculptures that portray full figures. The pigeon-toed stance may represent a physical handicap suffered by the donor.

water-soaked ancient wood is not protected, it crumbles to powder as it dries. The two principal methods of preservation are freeze-drying and water replacement; we chose water replacement for its convenience in dealing with a large number of specimens in a variety of sizes. As the objects were uncovered they were immersed in a water solution of a soluble wax (polyethylene glycol). When the solution has the proper concentration, it gradually replaces the original water in every cell cavity of the specimen. Thereafter the wood can be safely exposed to the air. We received invaluable aid and advice in this preservative technique from the Institut Royal du Patrimoine Artistique in Brussels.

Some of the largest wood sculptures, like the large stone ones, were portraits of pilgrims. There is, however, a sharp distinction between the two. The stone sculpture is fluid and curvilinear; the wood figures, although in no way aesthetically inferior, are stiff and angular. Many of them are carved with a cowled cape. The drape of the garment conceals arms held close to the body; the viewer receives an impression of power and quiet persistence. One figure of a man with his cowl over his head moves the viewer with its awkwardness and its faithful portrayal of pigeon-toed feet.

The cowled cape is a native garment, the *bardocucullus* at which the Latin poet Martial poked fun. When the cowl was not over the head, it was pushed back; the cape itself hung straight down with scarcely a fold. It has been suggested that this was a ceremonial garment, first made popular by the bards of Gaul, that eventually came to have religious connotations.

Other kinds of garment also adorn the wood sculptures. One is a simple straight coat, lacking detail at the collar, that falls to the ankles and is held in at the waist with a broad, heavy belt. This kind of coat appears several times. Tunics of various shapes and lengths are also in evidence; one is remarkably reminiscent of a modern short jacket.

Not enough wood sculptures of female pilgrims have been found to allow much generalization about the garb of Gaulish women. Nonetheless, to judge from one quite detailed example, a common feminine costume may have been a full-length tunic with a square neck. The tunic was worn under a cloak, open at the breast, that fell from the shoulder to the knee in front and continued in a sloping line to the back.

The most impressive sculpture of a woman reveals little or nothing about costume. The sculptor's material was a tree limb; he confined his chiseling to its two extremities, adding only a few details to the rough length of wood. A short curve suggests shoulders; another curve, almost imperceptible, at the bottom of the figure hints at a garment. Below it appear the feet, impossibly tiny.

It is the woman's face that places this work among the most beautiful in wood. The hair, parted in the middle, falls sleekly along the neck. The forehead is high, the line of the eyebrows clean-cut, the nose bold, the upper lip short and the lower one pouting. The oval face and its parts harmonize to present a forceful image that is both intense and self-contained [*see illustration on page 68*].

Among the wood votive offerings are odd sculptures consisting of two, three or even four heads carved one above another on a single log. One is immediately reminded of the Indian totem poles of the American Northwest. The execution is rough and geometrical. A straight line suggests the brow and the eyes, a small slit the mouth and a pyramid the nose.

What was intended by the repetition of the images? Examination reveals that there are two kinds of repetition. In some of these sculptures the wood separating one head from another is deeply scored, suggesting that the sculptor had intended to cut the heads apart. In others, however, the heads have virtually no neck, leaving no room for their being cut apart. Heads of the first kind thus seem to be analogous to the mass-produced bronze plaques. They were probably individual offerings prepared in series with a minimum of effort and waste; each head could be cut off on demand. Heads of the second kind appear to have had a different purpose. Perhaps they are analogous to repetitive prayers. If Sequana was properly invoked, she would give solicitous attention to the pilgrim whose head was represented by a votive gift. If the head was repeated, would not her ministrations be multiplied accordingly?

Such a practice, which is common in primitive religions today, is suggested by other evidence at the site. Among the stone votive offerings are some portraying as many as three heads or six legs. Curiously, in the stone sculptures the repetition runs horizontally rather than vertically. Presumably the horizontal system was as economical of stone as the vertical system was of wood. Material affairs, no less than spiritual ones, obey certain laws, and the Gauls were a notably practical people.

Two wood portrait heads that are anything but mass-produced tell us some-

thing about Roman artistic influence in Gaul at this period. One represents a young man and the other a young woman [*see illustration on page 65*]. The male head is large; its circumference is evidently as great as the dimensions of the log permitted. A continuous curve runs from the crown to the hairline at the nape of the neck; the hair falls across the forehead in a fringe. A jutting brow shelters eyes with well-defined heavy lids. The nose is large; deep lines beside the nostrils run down to the corners of a small, thick-lipped mouth above a rounded chin. The detail of the modeling is amazing. The lively chiseling of the mouth and chin contrasts with the smoothness of other facial areas to enhance the forcefulness of the features.

The face of the young woman is long and thin. The forehead is low, the eyes are heavy-lidded, the nose is straight and strong, the mouth thin and tight-lipped. Under the lower lip a round chin loses its definition farther down, where the neck shows a certain fleshiness. The hair, in heavy plaits, falls from the crown of the head and reaches just to the neck, forming a frame for the face; the plaiting is modeled in low relief. The sculptor was restricted in his handling of the subject by the narrow dimensions of the log, but this has not made his portrait, with its sense of balance and serenity, any the less interesting.

The two works rank with the finest of Gaulish—more properly Gaulish-Roman—sculptures. What is their position historically? The pottery and coins unearthed along with the wood votary offerings suggest that the offerings were buried sometime around A.D. 50 or 60. I should emphasize that this dating is preliminary; the excavation of the marsh is not yet finished and future findings could change the estimate. Nonetheless, if one may assume that a date around the middle of the first century after Christ is correct, it follows that the sculptures were executed earlier, possibly even before the Romans reached this part of Gaul late in the first century B.C. The unsophisticated "native" character of many of the objects from the marsh tends to support this conclusion.

At the same time the two portrait heads are proof that Roman influence, if not the Roman presence, was felt at an early date in Gaul. Roman portrait heads very much like these but cast in bronze have been found in both France and Switzerland. The resemblance indicates that the sculptors at the shrine of Sequana must have seen examples of such

LIVELY DISTORTION resulted from the artist's need to portray the figure of a sturdy bull within the narrow dimensions of the raw material: a section of trunk from an oak tree.

STONE FIGURE of a pilgrim is representative of another material much used by the makers of votive offerings. Objects carved in stone outnumber all other kinds of offerings to Sequana. Pilgrims portrayed in stone bear gifts; in this instance it is a purseful of money.

workmanship. It seems to have taken almost no time for the Gaulish sculptors to have absorbed the lesson from abroad. They went on to apply their own genius to the execution of lifelike portraits not in bronze but in wood, a material they had wholly mastered.

One may speculate about the appearance of the shrine of Sequana in the days when it was a center of Gaulish pilgrimage. It is worth recalling in this connection a famous center of healing in Greece that was visited by the Greek geographer Pausanias in the second century of the Christian Era: the Temple of Aesculapius at Epidaurus, across the Saronic Gulf from Athens. Pausanias' account of the Greek shrine suggests some interesting parallels with the Gaulish one. Pilgrims who sought the aid of Aesculapius were expected to spend a night within the temple precincts. Prior purification was required, and the presence of water conduits adjacent to the porch of the temple at Epidaurus suggests that the ritual involved ablution. In any case the pilgrims spent the night under the temple porch; to some of them Aesculapius would appear in a dream, and the fortunate dreamers could count themselves cured. It was then customary to pay for the carving of a temple inscription, stating the pilgrim's name, the illness that had inspired his visit and the nature of his cure. Pausanias saw six monumental pillars covered with inscriptions of this kind during his visit to Epidaurus.

In terms of parallels between Epidaurus and the shrine of Sequana, there was clearly no lack of water for ceremonial ablutions at the source of the Seine. The catchment basin below the spring had almost certainly been a part of the shrine even before the colonnade and porch were built. It seems likely that once this part of the shrine was constructed, purification at the basin was followed by an overnight vigil on the porch. The priests in the rooms nearby would have been ready to serve as interpreters if Sequana should reveal herself in the pilgrims' dreams.

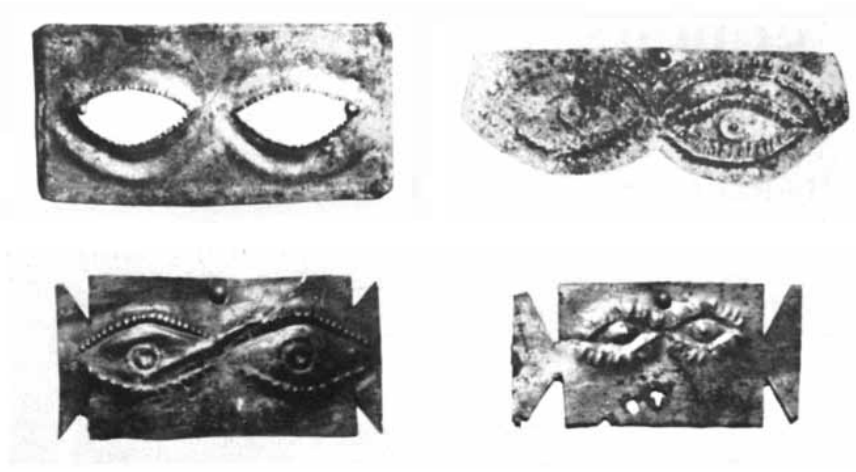
That the Gaulish pilgrims had their own characteristic way of memorializing a successful visit to Sequana's shrine is abundantly indicated by the number of their votive offerings. The custom survived elsewhere in France as late as the sixth century, as is made plain by an outraged description of a local temple that we owe to a church notable, Gregory of Tours. "The neighboring barbarians," Gregory wrote, "came to make sacrifices and to gorge themselves with meat and wine to the point of vomiting; there they adored idols as divinities and de-

posited wood sculptures representing such-and-such part of their body that had been attacked by some illness." The old ways seem to have died hard. Early church councils also fulminated against these pagan practices. "Destroy the springs and the groves they dare call sacred," orders a directive of the sixth century. "Prohibit the images they hang at crossroads and, should you discover any, burn them. Be certain that you will be cured by no other means than by invoking the cross of Christ."

Where were the pilgrims' votive offerings obtained? The presence of mass-produced items at the shrine of Sequana suggests that a bustling street of shops dealing in these objects stood somewhere nearby. Commercialism of this kind would help to account for an odd phenomenon: the uneven artistic quality of the votive offerings. For every exquisite sculpture one finds a dozen shamefully inept ones. This was doubtless how the shopkeepers were able to cater to pilgrims of very different means.

In the mind's eye we can readily populate Sequana's vale. There will be a crowd of people; not just the sick but also their friends and relatives. Perhaps the sick will make a show of their piety by wearing the plain cowled cape of ceremony. The dress of their companions, however, can be as colorful as the imagination desires. There will be throngs at the shops bargaining for votive offerings. In the mosaic-paved courtyard there will be offertory tables displaying the pilgrims' purchases. Within the small temple building there will be crowded shelves where the testimony of past cures is exhibited. Here, we may expect, the richer votive offerings will occupy a prominent position and the humbler ones will be relegated to an obscure corner.

As the day wears on the number of people will diminish. Pilgrims not yet interviewed in the priests' quarters will drift off to one of the neighboring inns to await the morning. So will the companions of the day's chosen pilgrims, as the pilgrims wash themselves before going for the night to Sequana's porch. The blind are led, the lame assisted, perhaps a part of the porch is set aside for the women in the group. Finally, night falls and the hours arrive when Sequana will make good the promise of her outstretched hands. If we now let reality take the place of imagination, we see that it is the simple remains of the sanctuary before us, and the glimpse they offer of the mind of ancient Gaul, that are the true Treasure of the Seine.



BRONZE PLAQUES are the third kind of votive offering found at the sanctuary of Sequana. Presumably the plaques shown here were offered by pilgrims with eye trouble. The one showing only empty sockets (*top left*) may have been meant to convey blindness.

BRONZE SCULPTURE OF SEQUANA shows the goddess standing in a boat with a duck's-head prow and a duck's-tail stern. The figure is rendered in the traditional Roman style.

SUPERNOVA REMNANTS

Radiation emitted by the debris left over from catastrophic stellar explosions may yield clues to many astronomical puzzles

by Paul Gorenstein and Wallace Tucker

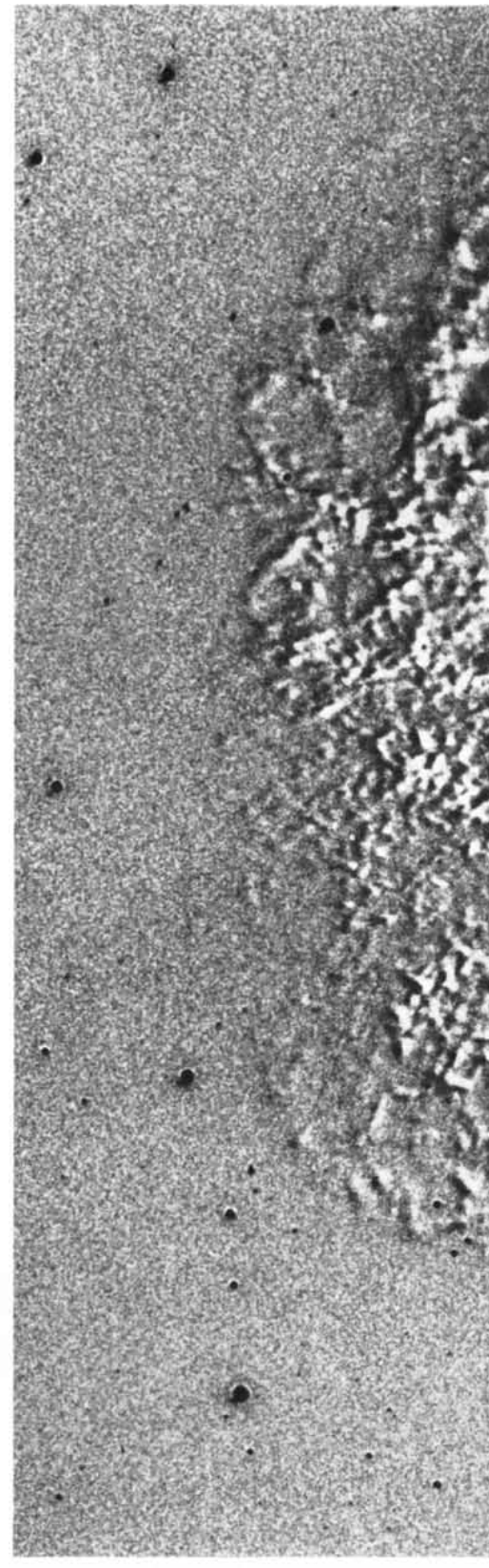
"The stars, like measles, fade at last." These words by the poet Samuel Hoffenstein provide a graphic, if flippant, description of the way in which most stars come to the end of their evolutionary trail. A few stars, however, refuse to make such a quiet exit when their nuclear fires grow cold. About one star in 100 explodes catastrophically. For almost two weeks the exploding star—a supernova—radiates more energy than a billion suns and ejects matter at close to the velocity of light. Cosmic rays (the nuclei of hydrogen and other light elements) are probably produced in abundance, and elements heavier than iron are thought to be synthesized during the first few seconds of the explosion itself. These heavy elements, along with those manufactured earlier by the star during its lifetime, are ejected into interstellar space to become part of a cosmic reservoir that supplies the raw material for new stars and planets. Quite likely most of the matter of which we are made was originally part of the debris of supernovas.

The expanding shell of debris creates a nebula that for hundreds or even thousands of years radiates vigorously in both the X-ray and radio regions of the spectrum. In addition the cataclysm may leave behind a small, extremely dense core. The recent discovery of pulsars suggests that this core, rotating at high speed, can be a pulsating source of radiation. Thus the "death" of a star is sometimes followed by a "transfiguration" in which the star begins a new life of physical conditions entirely different from those encountered in normal stars: superdense matter, enormously strong magnetic fields and rapid rotation. For the first 1,000 years or so of this second life the total power emitted by the star is equal to the power of 1,000 suns.

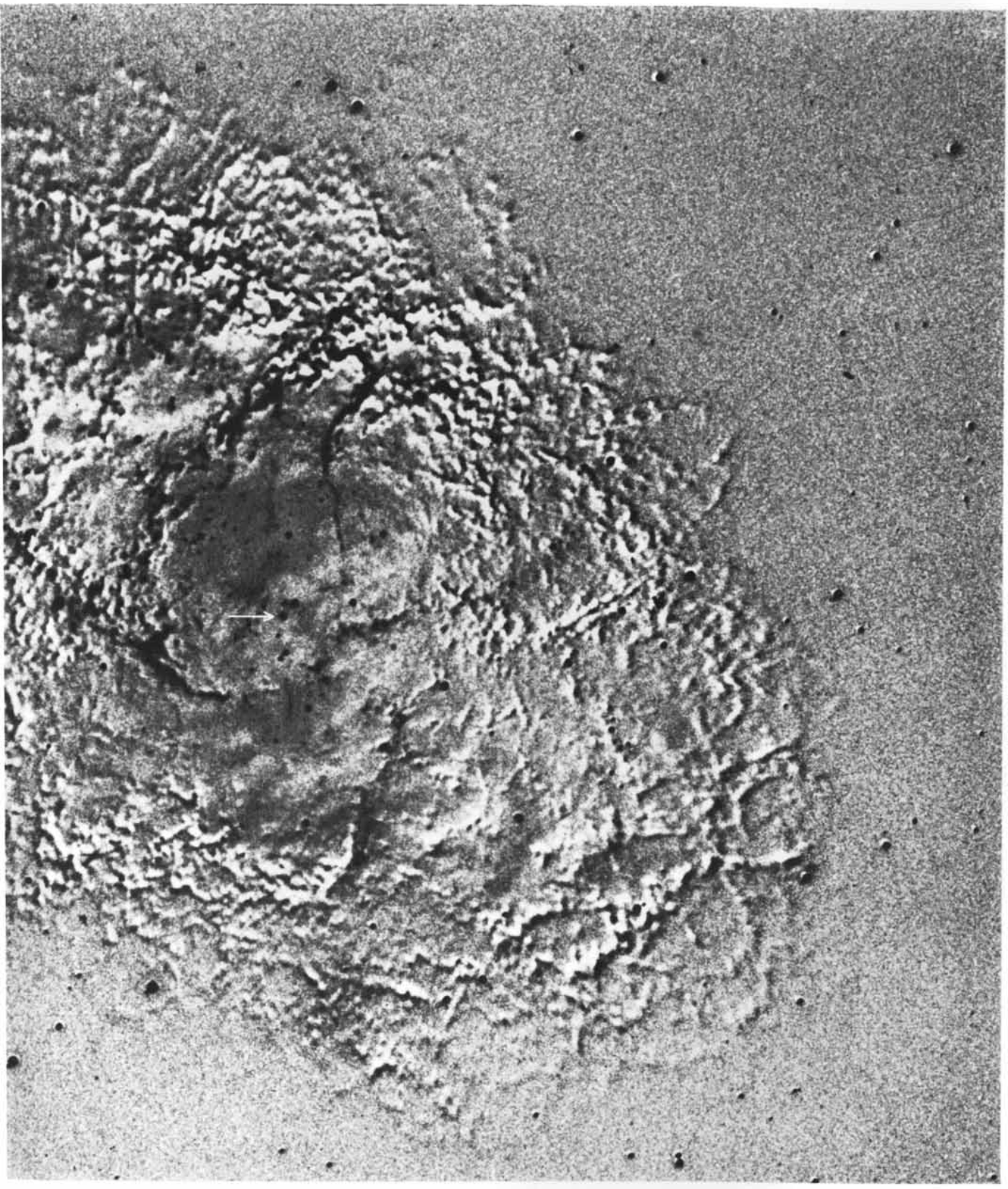
In our galaxy of some 100 billion stars a supernova explosion occurs about once every 100 years. Since the nebula and the transfigured star survive for a long time after the outburst, a sizable number of supernova remnants should always be present in our galaxy and should therefore be accessible to detailed observation. At least two dozen such remnants have now been identified.

Four of the supernova remnants have also been identified with novas whose sudden appearance in the sky can be found in historical records. The four are the objects seen in A.D. 1006, in 1054 (the Crab Nebula), in 1572 (Tycho's nova) and in 1604 (Kepler's nova)—all before the invention of the telescope. Although the earlier two outbursts were probably brighter than the later two, there is no evidence that they made the slightest impression on Europeans, who were convinced of the immutability of the heavens. The supernova of 1572, however, was immediately sighted by Tycho Brahe, who observed it regularly until it disappeared from view. Tycho published his observations in 1573 in *De Stella Nova*. "Initially," he wrote, "the nova was brighter than any other fixed star, including Sirius and Vega. It was even brighter than Jupiter.... It maintained approximately its luminosity for almost the whole of November. On a clear day it could be seen...even at noon.... It slowly faded and finally disappeared completely in March of 1574."

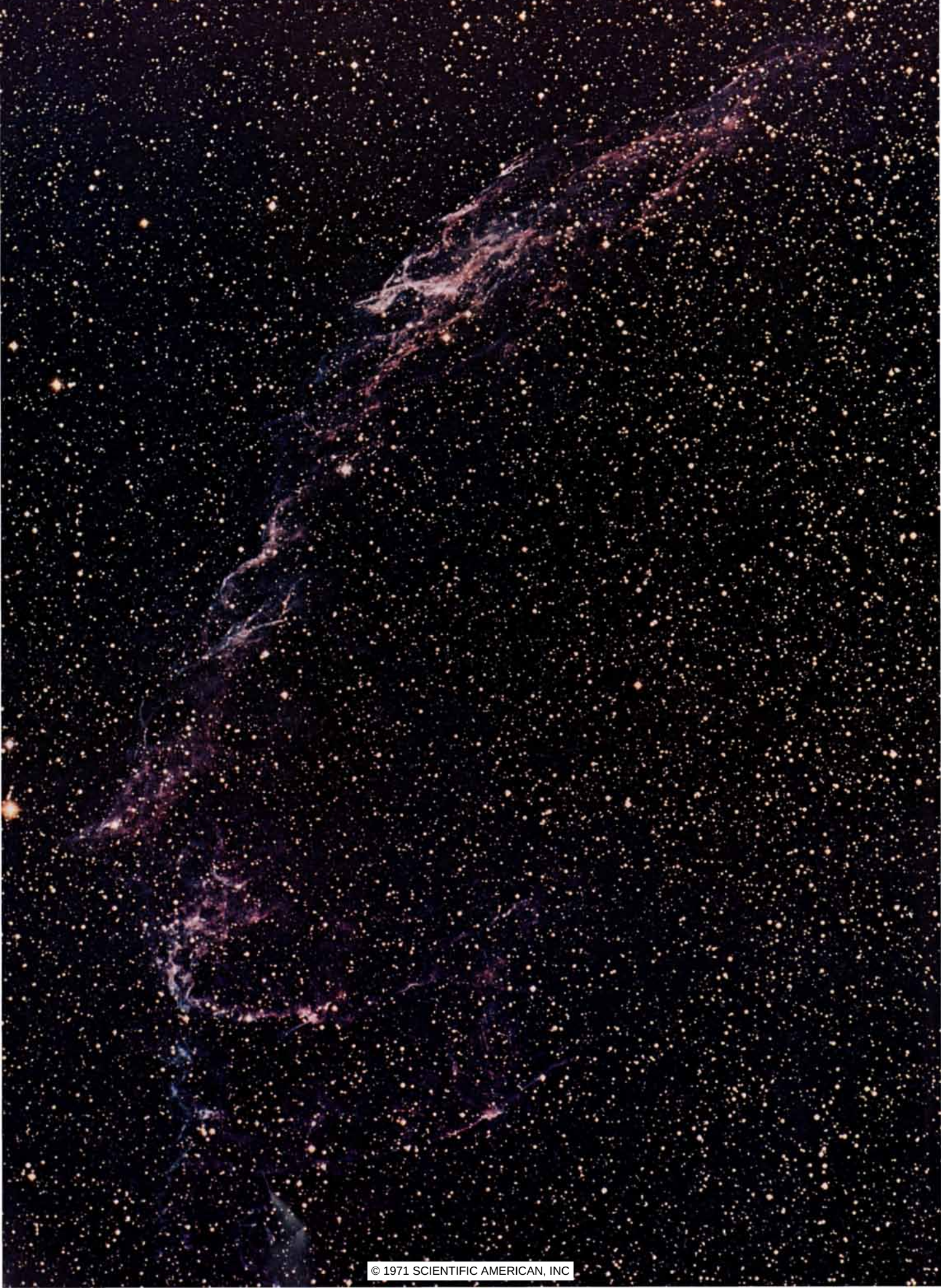
The search for the remnant of this event was unsuccessful until 1952, when R. Hanbury Brown and Cyril Hazard of the University of Cambridge discovered a strong radio source near but not quite at the position given by Tycho. With the help of this information the optical remnant was found by Rudolph Minkowski with the 200-inch Hale telescope. Recently X-ray observations with rocket-



EXPANSION OF CRAB NEBULA is vividly demonstrated in this composite photograph, obtained by superimposing a positive image and a negative image made 14 years apart. Both original plates were made with the 200-inch Hale telescope on Palomar Mountain; the positive, in which the radiating gaseous filaments of the nebula appear



light, was printed from a plate made by Walter Baade in 1950; the negative, in which the same structures appear dark, was printed from a plate made by Guido Münch in 1964. Münch also made the composite. The slight relative displacements of the light and dark images resulting from the nebula's expansion during the interval between the two exposures is responsible for the almost tactile, bas-relief quality of the composite. The Crab Nebula is a remnant of a supernova explosion observed in A.D. 1054 by Chinese astronomers. It is some 5,000 light-years away and at present its visible portion measures about six light-years across. In addition to its visible luminosity, the Crab is also a powerful source of radio waves and X rays. This prodigious energy output has been attributed to the conversion of the kinetic energy of a central rotating star (*arrow*) to electromagnetic energy. The star, another remnant of the original supernova explosion, was recently identified as a pulsar: an extremely dense neutron star that emits radiation in pulses.



borne instruments have shown that the remnant of Tycho's nova is also a strong source of X rays with energies between 1,000 and 10,000 electron volts.

The supernova of 1604 was studied by Johannes Kepler; it attained a brightness equal to or somewhat greater than that of Jupiter. It was visible by day for about a month and at night for about a year. In 1941 Walter Baade of the Mount Wilson Observatory, who had analyzed the old records, discovered an optical nebula near the point where the supernova had appeared in 1604. The nebula consists of bright filaments that cover an area subtending about a minute of arc. Some 10 years later Hanbury Brown and Hazard discovered that it too was a radio source.

The supernovas of 1006 and 1054 occurred during a time when astronomy was flourishing in the Orient. In those days the appearance of a new bright object in the sky was a portentous event for individuals and nations alike; meticulous astronomical records were kept for interpretation by court astrologers. The use of these chronicles to identify the supernovas of 1006 and 1054 and to obtain information about them is an outstanding example of interdisciplinary research involving cooperation between Orientalists and astronomers.

In one of the chronicles of the Sun Dynasty, translated by J. J. L. Duyvendak of the University of Leiden, the supernova of 1054 is mentioned several times. Shortly after the appearance of this object the head astronomer of the emperor's court filed a report of his observations and an interpretation of the significance of the event: "Prostrating myself, I have observed the appearance of a guest star; on the star there was a slightly iridescent yellow color. Respectfully, according to the dispositions for Emperors, I have prognosticated, and the result said: The guest star does not infringe upon Aldebaran; this shows that a Plentiful One is Lord, and that the country has a Great Worth. I request that this be given to the Bureau of Historiography to be preserved." In the late 1920's Edwin P. Hubble of Mount Wilson suggested that the Crab Nebula might be the remnant of the nova of 1054. He noted that the nebula's size and rate of expansion indicated that it had originated about 900 years earlier. Detailed studies by K. Lundmark, Jan H. Oort, Nicholas U. Mayall and Duyvendak subsequently confirmed Hubble's conjecture.

The supernova of 1006 was conclusively identified primarily as a result of a study by Bernard Goldstein of Old Arabic, Syriac, Latin, Japanese and Chinese chronicles. From them it was learned that at the end of April in 1006 a supernova occurred in the constellation Lupus. Initially the star was as bright as Venus, and it remained visible at night for much longer than a year. Today it is difficult to detect the remnant of this event in the visible part of the spectrum. Radio emission from the remnant was discovered, however, in 1965.

Not all the "guest stars" reported in the Chinese chronicles or new stars reported by European astronomers qualify as supernovas. Some were comets and some were ordinary novas: explosions involving the outer layers of a star that occur much more frequently than supernovas and are less spectacular. Comets, of course, move across the sky and thus can be easily distinguished from novas and supernovas, which remain fixed. It is not so easy to discriminate between novas and supernovas, even though supernovas are intrinsically much brighter. A nova, if it is nearby, can look as bright as or brighter than a supernova that is much farther away. Moreover, the light from a distant object is often absorbed or scattered by dust particles in the interstellar medium. For this reason supernovas were not recognized as a distinct class of exploding stars, intrinsically more luminous than novas, until the 1920's.

The distinction was made after it was recognized that many objects called nebulas and thought to be nearby were actually distant galaxies. In the period from 1885 to 1920 about 10 novas were discovered in such "nebulas." In some cases the brightness of the nova at its maximum was equal to or greater than the brightness of the associated nebula. When Hubble demonstrated that most nebulas are independent star systems, comparable to our own galaxy and more than a million light-years away, it became apparent that a nova observed in such a system must be more than 10,000 times more luminous than the ordinary novas that flare up in our galaxy about 20 times a year. Fritz Zwicky of Mount Wilson and Baade suggested that the more catastrophic outbursts be called "supernovas."

Since a supernova explosion is a rare event within any one galaxy, the only way its initial phases can be studied is by monitoring many galaxies, thereby increasing the odds that a supernova will be sighted. In 1933 Zwicky began a systematic search for supernova explosions in other galaxies. Today his pioneering work is being carried on by astronomers all over the world; they discover 10 to 20 supernova outbursts per year. Taking into account the number of galaxies observed, one can arrive at the estimate that a supernova explosion occurs once per 100 years per galaxy.

The characteristic signature of a supernova remnant is the emission of radio waves whose distribution of energy with wavelength is nonthermal, meaning a distribution of a type not associated with an ordinary hot object. If this radiation also appears to originate with a shell-like structure or to emanate from a region showing optically visible filaments, its source is almost certainly the remnant of a supernova. Probably all the remnants identified on this basis are less than 100,000 years old. The younger members of the group, with 1,000 times the energy output of the sun, are among the most luminous objects in our galaxy. Not all the energy is emitted in the form of visible light or radio waves. In some remnants much if not most of the emitted energy is in the form of energetic photons: X rays with energies ranging from a few hundred electron volts to as much as a million. Since supernova remnants emit energy over such a broad frequency range, from X rays to radio waves, they have a unique role in astronomy and are of special importance for the understanding of astrophysical processes.

At the time Hubble proposed that the Crab Nebula might be the remnant of the supernova of 1054, astronomers had few tools for studying anything but the visible spectrum. The earth's atmosphere blocks out radiation with wavelengths shorter than about 3,100 angstroms, which lies in the near-ultraviolet part of the spectrum. Although some radio wavelengths penetrate the atmosphere freely, radio astronomy was limited by inadequate technology before World War II and had few practitioners. Wartime advances in radio technology dramatically widened astronomy's horizons. They introduced sensitive techniques for detecting radio waves to a new generation of astronomers who were keen to apply them to observing the cosmos. At the other end of the spectrum postwar advances in

PART OF OLD REMNANT known as the Cygnus loop or the Veil Nebula appears in the color photograph on the opposite page. The photograph was made with the 48-inch Schmidt telescope of the Hale Observatories. Entire nebula is shown on next page.

ENTIRE CYGNUS LOOP was photographed in red light with the 48-inch Schmidt telescope. The expansion of this shell-like structure has been going on for perhaps as long as 60,000 years; it currently has a diameter of about 100 light-years. This supernova remnant is also a strong source of both radio waves and X rays. No pulsating component has yet been discovered in it, however.

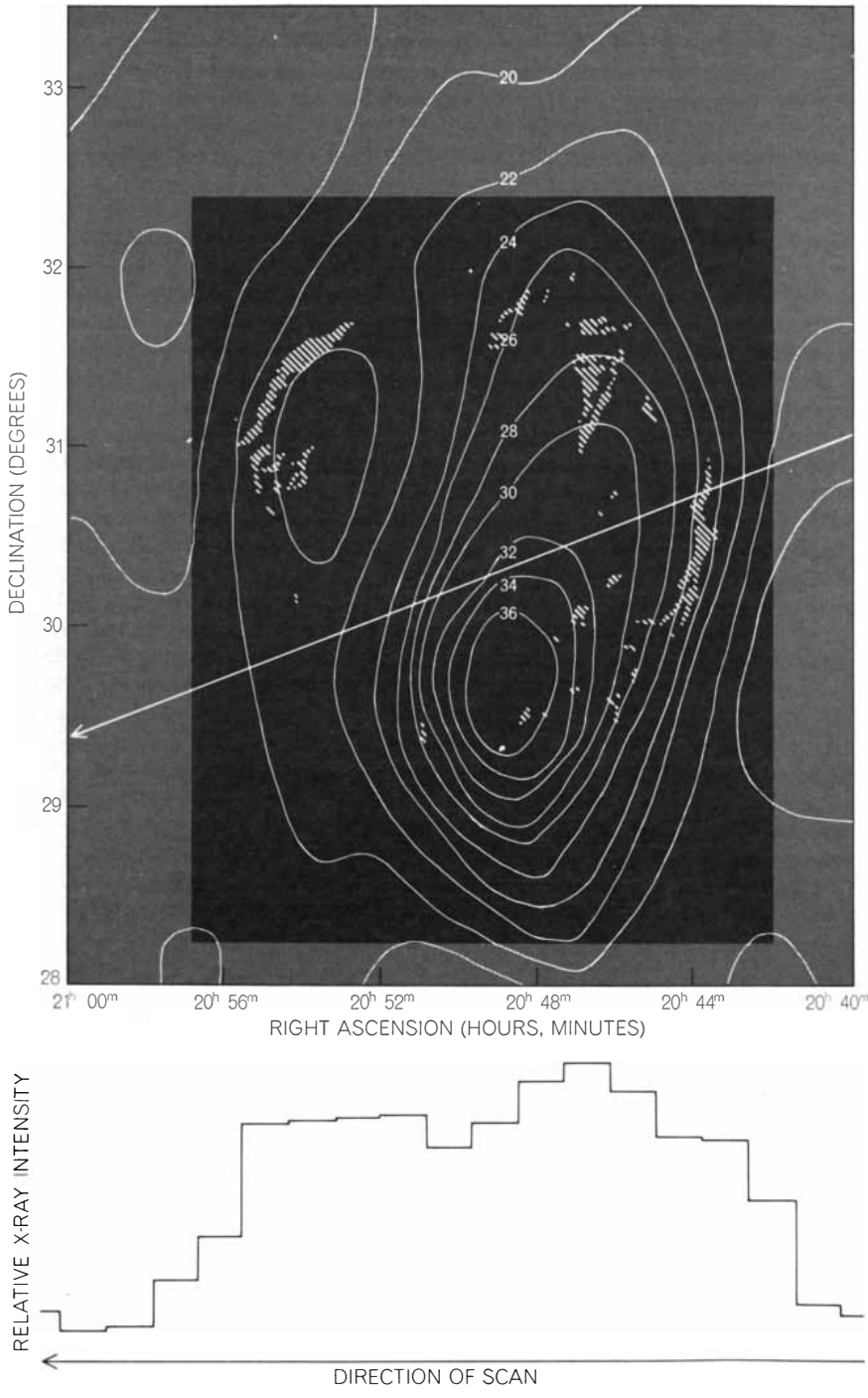
rocketry made it feasible for the first time to lift instruments above the opaque blanket of the earth's atmosphere.

The Crab Nebula turned out to be one of the strongest sources of radio waves in the sky and was the first radio source to be identified with a visible object. It was hoped, therefore, that radio observations might explain some of the nebula's enigmatic aspects. For example, precise optical observations had shown that the observed rate of expansion could not lead to the observed size unless there had been some kind of acceleration after the 1054 explosion. Furthermore, variable light features, or "ripples," had been discovered near the center of the nebula. The ripples appeared at irregular intervals, moved rapidly outward at a velocity of up to a tenth the velocity of light and vanished after about a month.

The discovery of powerful radio emissions from the Crab Nebula at first only deepened the mystery. Astronomers could not understand how such large fluxes of radio emission could be produced by an astronomical object. Then Hannes Alfvén and A. Herlofson of Sweden, K. O. Kiepenheuer of Germany and V. L. Ginzburg of the U.S.S.R. showed that nonthermal radio emission could be expected if the object comprised two ingredients: electrons of very high energy and a magnetic field with which they could interact. If the electrons were trapped by the field, they would travel in orbits around lines of magnetic force and emit radiation in a narrow beam tangent to their orbits in the direction of motion. Commonly called synchrotron radiation because it was studied in connection with the synchrotron electron accelerators developed in the 1940's, the radiation is polarized with its electric vector perpendicular both to the magnetic field and to the direction of radiation [*see illustration on next page*]. The intensity of this radiation and the peak frequency increase rapidly with energy.

In 1953 I. S. Shklovsky of the U.S.S.R. explained how the synchrotron mechanism could also account for the continuous optical spectrum of the Crab Nebula. The subsequent observation that the optical and radio emission are polarized led to the general acceptance of the synchrotron hypothesis [*see illustration on page 81*]. Thus it became clear that the Crab Nebula is a cosmic synchrotron, permeated by electrons with energies of 1,000 billion electron volts or even higher, and by magnetic fields more than 10 times stronger than the field in interstellar space. The problem of understanding the conditions in the Crab Nebula and other supernova remnants then became one of understanding the origin of the magnetic field and the vast fluxes of high-energy electrons.

A new element was added to the story in 1964, when instruments flown aboard sounding rockets detected X-ray emission from the Crab Nebula. Although X rays can also be produced by the synchrotron process, this interpretation presented difficulties because synchro-



OTHER VIEWS of the Cygnus loop are represented in this illustration. At top is a contour map of the radio emission from the region of the nebula made by S. Kenderdine of the University of Cambridge; numbers on contours indicate radio intensity in multiples of 900 degrees Kelvin. The black rectangular area corresponds to the field of the photograph on the opposite page; hatching indicates main visible features of the nebula. The long white arrow represents the direction scanned by an X-ray detector during a rocket flight in which X-ray emission from the region of the nebula was observed. The graph at bottom shows the recorded X-ray intensity integrated over intervals along the direction of the scan.

tron electrons energetic enough to produce X rays would quickly lose their energy by radiation. Hence the question arose of how electrons of sufficient energy were continuously supplied some 900 years after the initial explosion.

Then late in 1967 pulsars were discovered. These remarkable objects emit short pulses of radio waves at extremely regular intervals, ranging from about a quarter of a second to somewhat more than a second. Their identity was clarified the following year, when pulsars with much shorter periods were found in the Crab Nebula and in another supernova remnant: Vela X. Thus an intimate connection between pulsars and supernova remnants was established. Furthermore, pulsations from the Crab Nebula were observed in both the visible and the X-ray regions of the spectrum. The source of these pulsations is a faint object that in astronomical photographs looks like an ordinary star. Recently it has been found that at X-ray energies above a few million electron volts the energy output of the pulsating component is comparable to the energy contained in the steady-state radiation from the entire nebula.

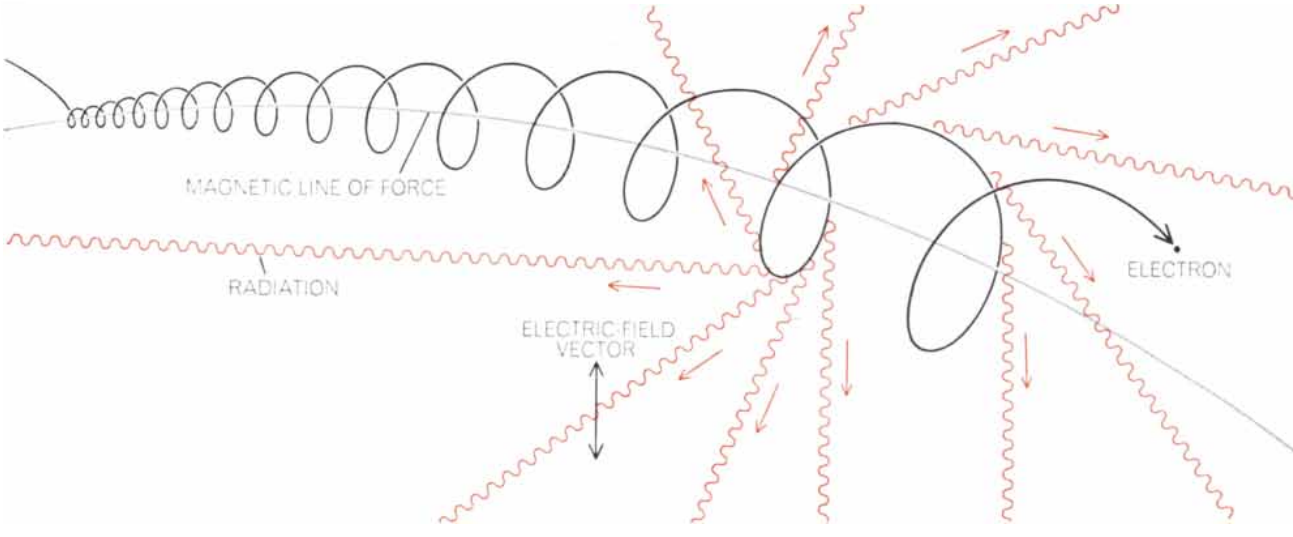
The most recent supernova explosion in our galaxy took place in about 1700 in the constellation Cassiopeia not far from Tycho's supernova. It may seem surprising that this event was not observed even though a rather widespread interest in astronomy had developed by then. The explanation appears to be that the light from that region of the sky is obscured by a large amount of dust; moreover, the supernova itself may have been less brilliant than the supernovas of 1572 and 1604. The first hint that there was a supernova remnant in the region was provided in 1948 by Martin Ryle and F. G. Smith, who with the aid of one of the first radio telescopes at the University of Cambridge discovered in Cassiopeia an intense radio source that they named Cassiopeia A. With the 200-inch Hale telescope Baade and Minkowski then succeeded in obtaining a photograph of a faint fragmentary nebula that could be identified with the radio source. From studies of the nebula the time of the event was determined.

Within the past few years rocket-borne instruments have detected intense X-ray emission from Cassiopeia A. The structure of the remnant, best revealed in a map of its radio emission, consists of an expanding shell whose diameter is about four minutes of arc [*see illustrations on pages 82 and 83*]. The shell structure is suggestive of matter ejected at high velocities. So far the search for a central object or pulsating source of radiation has been unsuccessful.

An example of a much older supernova remnant is the Cygnus loop [*see illustration on page 78*]. Compared with the Crab Nebula and Cassiopeia A it is a huge object stretching across three degrees of arc in the sky, six times the apparent diameter of the moon. Like Cassiopeia A it is an expanding shell, but the expansion has been going on for a much longer time, possibly 60,000 years. The Cygnus loop is also a strong source of nonthermal radio waves, and last year it was discovered to be an intense source of X rays. The X rays from the Cygnus loop have a lower average energy, however, than those emitted by the Crab Nebula. The search for a central star or pulsar in the Cygnus loop has not been successful.

The cause of a supernova outburst is still the subject of intensive investigation and considerable controversy, but it is generally agreed that the onset of the explosion is ultimately related to instabilities in the structure of the star that arise when the supply of nuclear fuel in the central parts of the star is exhausted. These instabilities occur only in stars whose mass is greater than about one and a half times that of the sun. Less massive stars, including the sun, begin to contract when their nuclear fuel is consumed. Ultimately the pull of gravity is balanced by the pressure of "degenerate" electrons: an incompressible electron fluid that finally emerges because no two electrons can occupy the same energy state. When this stable configuration is reached, the star has become a white dwarf and slowly dies "not with a bang but a whimper." In the case of stars whose mass exceeds about 1.5 solar masses, the density and temperature in the central core exceed the critical values beyond which stability is possible. Then the star collapses under the influence of gravity and an explosion ensues.

In a supernova explosion the outer shells of the collapsing star are ejected at high velocity; in some, if not all, cases a dense relict is left behind. The idea that superdense bodies could be formed as end products of supernova explosions



SYNCHROTRON RADIATION is emitted by electrons spiraling at relativistic speeds (that is, speeds near the speed of light) along a line of force in a magnetic field. The radiation is emitted in a narrow beam along the direction of the electron's motion; the predominant wavelength depends on the velocity of the electron and the strength of the magnetic field. Radiation produced by this mechanism is highly polarized, with its electric vector perpendicular to both the line of force and the direction of the radiation.

is not new. Hypothetical superdense stars composed almost entirely of neutrons were conceived some 40 years ago by the Russian physicist L. D. Landau and independently by J. Robert Oppenheimer. In 1938 Zwicky suggested that such stars could be formed in supernova explosions. The gravitational collapse that takes place before the supernova explosion compresses the matter in the core of the star to a radius of about 10 kilometers. Since the residual core still has a mass comparable to that of the sun, its density approximately equals the density of matter in the nucleus of an atom: on the order of 100,000 tons per cubic millimeter.

Since the discovery of pulsars the existence of neutron stars has been widely accepted. In addition theorists have speculated that if a very massive star were to explode as a supernova, the core might collapse gravitationally without limit to form what has been called a "black hole." The gravitational field of a black hole would be so strong that nothing—not even light—could escape its grip. If black holes exist, they can be detected only through gravitational effects. For example, a black hole might perturb the motion of a nearby visible companion or interact with its radiation in a way that might be observable.

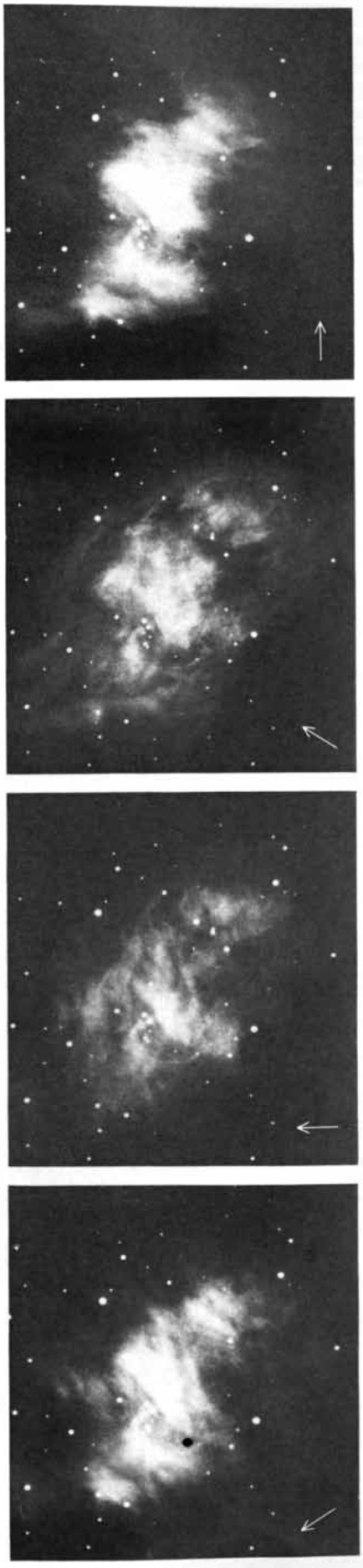
The gravitational collapse that creates a neutron star necessarily leads to an increase in the strength of the star's magnetic field and in the rate of its rotation. The lines of magnetic flux that pass through highly conducting material, such as the matter inside a star, can be regarded as "frozen into" the material; when the conductor moves, the magnetic field moves with it. Hence when the core of a star collapses, the magnetic field is compressed along with the matter. As a result the magnetic field on the surface of a neutron star can be more than 1,000 billion times stronger than the average magnetic field on the sun.

The rapid rotation of a neutron star also follows naturally from the way it is formed. Almost all stars rotate; the period of rotation ranges from a few hours to about a month. Conservation of angular momentum implies that as a rotating star collapses it must rotate faster, unless angular momentum is carried away by radiation fields and expanding debris. Since the radius of the core shrinks by a factor of 10,000 or more when a star collapses to the configuration of a neutron star, a rotation period as short as a millisecond could result. Given the neutron star's dimensions and rotation period and assuming it to be a rigid body, it is a straightforward matter to estimate the energy stored in rotation. For a 10-millisecond period the rotational energy of a neutron star is of the order of 10^{50} ergs, a quantity comparable to the total energy radiated by a star during its entire lifetime.

Thus we see that a supernova explosion releases gravitational energy in several forms. First there is the radiant energy emitted directly in the early phases of the explosion. The matter simultaneously ejected at high velocity carries away translational kinetic energy. Finally the neutron star that survives is endowed with an enormous amount of rotational kinetic energy. Although our understanding is by no means complete, it appears that the observed characteristics of supernova remnants can be accounted for by the transformation of rotational and translational kinetic energy into electromagnetic radiation over a relatively long period of time.

Let us see how these ideas can be applied to the remnant of the guest star of 1054, the Crab Nebula. The discovery of a pulsar in the Crab Nebula with a period of only 30 milliseconds led to the general acceptance of the idea that pulsars are rotating neutron stars. It became apparent that there is an energetic coupling between the pulsar and the nebula when a group of radio astronomers at the Arecibo Ionospheric Observatory in Puerto Rico found that the pulsar's rate of rotation is decreasing at the rate of one part in 2,400 per year. This slowing down, although seemingly slight, implies that the Crab pulsar is losing rotational energy at a rate adequate to sustain all the radiation from the entire nebula. Thus the long-sought source of energy for maintaining the Crab Nebula has finally been discovered. The problem now becomes one of understanding how rotational energy can be converted into high-energy particles. There is general agreement that the conversion is effected by processes that take place in a rapidly rotating intense magnetic field.

Rapid rotation and intense magnetic

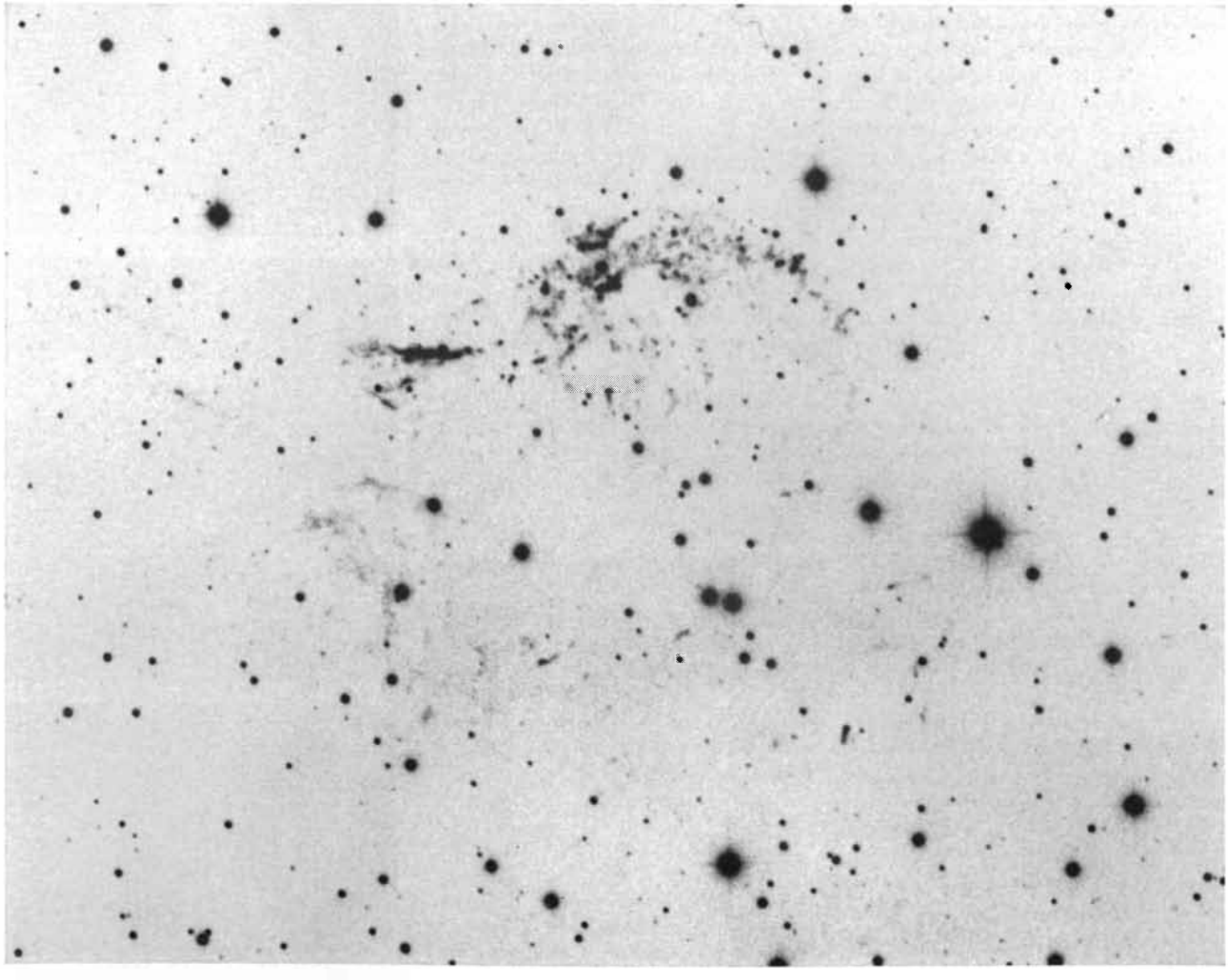


POLARIZATION OF LIGHT emitted by the Crab Nebula is evident in this series of photographs made through a polarized filter in four different orientations. The arrow in the lower right-hand corner of each photograph indicates the direction of the electric vector of the light that passes the filter. Such photographs are taken as evidence that the light from the Crab Nebula is synchrotron radiation emitted by high-energy electrons.

fields combine to create large-scale electric fields that accelerate charged particles to high energy. The particles produce the observed pulses of radio and optical radiation and X rays. As the particles stream out into the nebula they also give rise to the continuous and extended source of radio, optical and X-ray emission. Together with the large-scale electromagnetic waves generated by the rotating magnet, the accelerated particles carry angular momentum and energy away from the neutron star, slowing it down.

This general picture probably applies to the activity observed in the other young supernova remnants. We cannot be so certain as with the Crab Nebula, since a pulsar has been observed to date in only one other remnant: Vela X. It is possible that pulsars exist in other supernova remnants but that they are not favorably oriented for us to detect them. According to one explanation, high-energy particles stream out along a narrow cone (or cones) from one or more spots on the rotating neutron star, so that the pulsar resembles a lighthouse in operation. If this view is correct, one would expect to observe the pulsar directly only if the beam happened to include the solar system in its sweep. Otherwise one would observe only the indirect effects of the high-energy particles interacting in the nebula. If the width of the pulsar beam emanating from the Crab Nebula is typical (it is less than 12 degrees), one can readily figure that on the average only one pulsar in 30 should be detectable. This expectation is consistent with observing pulsars in only two of the known supernova remnants. Another indirect argument for the existence of pulsars in all supernova remnants is a need for continuous injection of high-energy particles into the nebulas for at least a few years after the initial explosion to account for radio emission of the strength observed.

With the passage of time the rotational velocity of the neutron star slows to such an extent that X-ray emission from the source substantially ceases. Radio emission can continue for much longer, however, because it can be sustained with less electron energy and a lower magnetic field. When the central source no longer supplies high-energy electrons in abundance, the nebula also loses its ability to generate X rays by the synchrotron process. The radio emission lasts considerably longer since it is produced by electrons of lower energy that

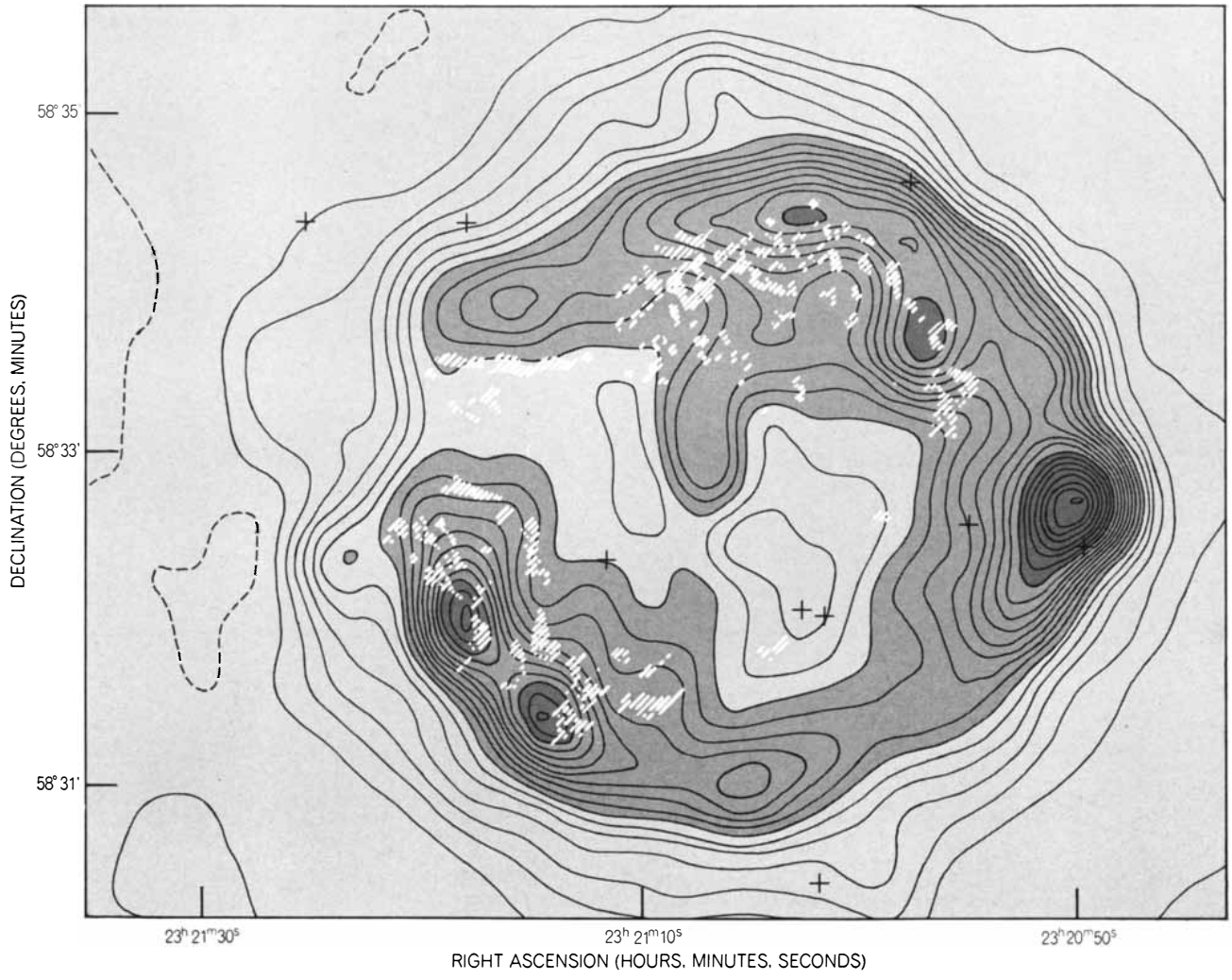


CASSIOPEIA A, first identified as an intense radio source by Martin Ryle and F. G. Smith of the University of Cambridge in 1948, was subsequently shown to correspond to a faint, fragmentary nebula by Baade and Rudolph Minkowski, who used the 200-inch Hale telescope to obtain photographs such as this negative made in red light by Minkowski. The supernova explosion that gave rise to the Cassiopeia A nebula is estimated to have taken place about 1700, making this nebula the newest supernova remnant in our galaxy.

have a much longer radiative lifetime than the electrons that produce X rays.

It seems probable that all the 50-odd known pulsars are the spinning cores left behind by supernova explosions, but we have been able to detect the nebular remnants of only two of them. This raises a related question: Does every supernova explosion leave behind a pulsar? There is no general agreement on the answer. The total number of pulsars that have been found agrees at least in order of magnitude with an estimate based on the rate of occurrence of supernova explosions, the lifetime of a pulsar and the idea that we should detect only the fraction that radiates in our direction. Considering the uncertainties in these quantities, however, one cannot exclude the possibility that only some of the explosions produce a pulsar; others may leave a black hole or no compact object at all.

Let us now try to trace what happens to the matter ejected into the interstellar medium at supersonic velocities by the explosion. The general picture is as follows. The expanding debris from the explosion acts like a piston driving into the interstellar medium, producing a shock wave that travels slightly faster than the ejecta. Thus the kinetic energy of the ejected matter is gradually transformed into the energy of a shock wave in the interstellar medium. For the first few thousand years or even tens of thousands of years the temperature of the matter behind the shock front is extremely hot, on the order of one million to 100 million degrees Kelvin. A gas at this temperature emits most of its energy in the form of X rays, with the intensity of the emission depending on the amount of mass in the hot shell. If the supernova remnant is, say, 2,000 light-years away, a detectable flux of X rays will result when about 25 solar masses of interstellar matter have been swept up; this takes about 10,000 years. As the shock wave expands it slows down and the postshock temperature decreases. As a consequence the radiation is emitted at longer and longer wavelengths. The total power of the radiation increases with time, however, because of the steady increase in the mass of interstellar matter swept up by the shock wave. The process continues until most of the kinetic energy carried by the original ejecta and transformed into a shock wave has been radiated away, after which the emission fades rapidly. This may take anywhere from 30,000 to a few

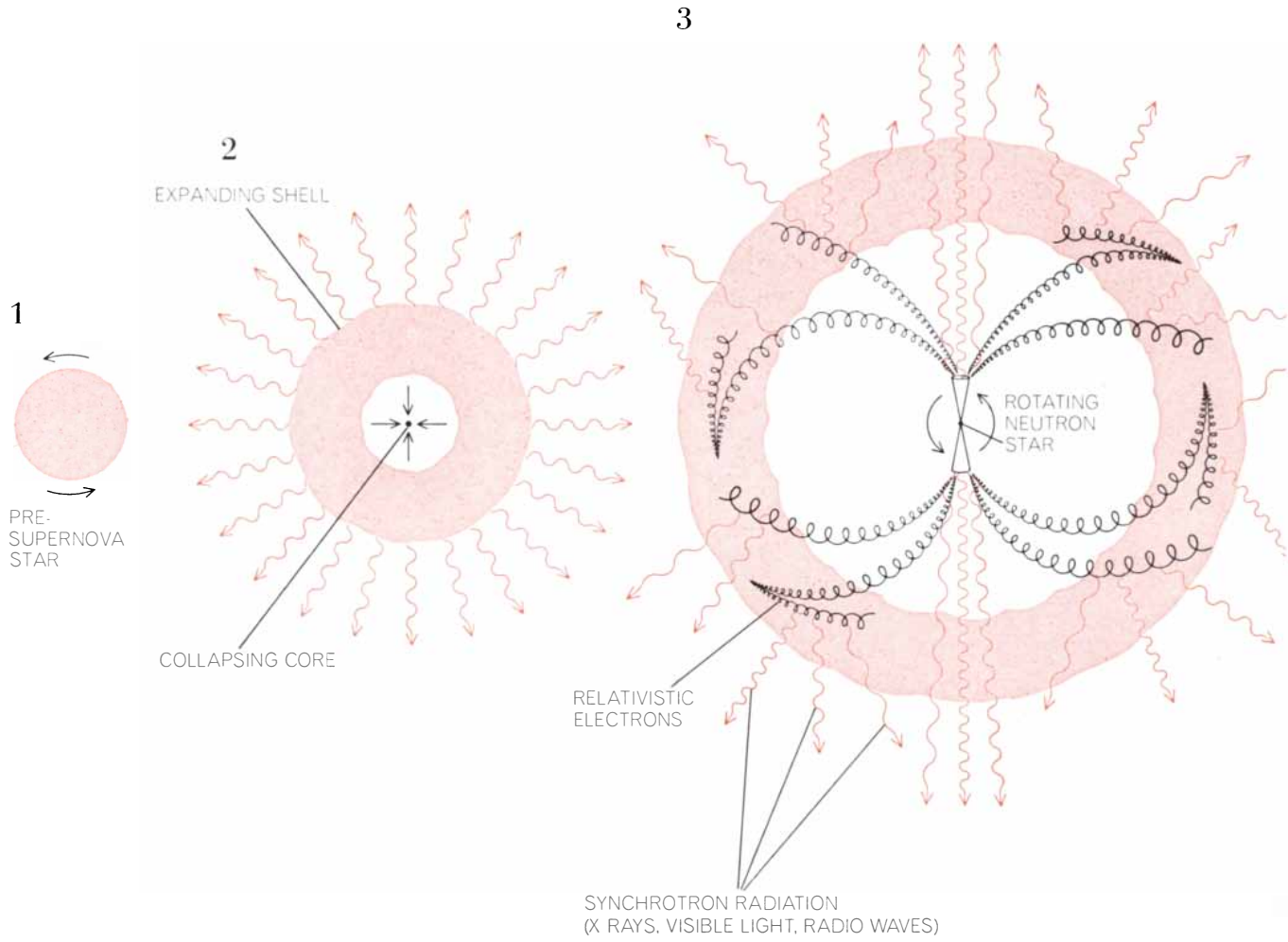


RADIO MAP OF CASSIOPEIA A was made by Ryle and his colleagues at Cambridge in 1965. The contour interval is 3,000 degrees Kelvin; progressively darker gray areas show higher-intensity peaks. Hatching again indicates main visible features. The field of view is the same as in the photograph on the opposite page; the positions of bright stars in the field are indicated by crosses. The map shows that the Cassiopeia A radio source has a circular structure that is presumably the projection of a spherical shell.

hundred thousand years, depending on the energy of the explosion and the density of the interstellar medium. During the final phase, prior to exhaustion, the shock wave will be radiating at a rate of about 100,000 suns, chiefly in the extreme ultraviolet part of the spectrum. Over a large volume of space the radiation will raise the temperature of the interstellar gas to the ionization point: the point where outer electrons are stripped from atoms and molecules.

A possible example of a blast wave produced by a supernova is the Cygnus loop. The loop is an extended source of X-ray, optical and radio emission. If the shock-wave model is applicable here, the X-ray emission is presumably from the hot plasma immediately behind the shock front and the optical emission comes from cooler, denser filaments. The radio emission is due to synchrotron radiation from high-velocity electrons as they encounter the regions of strong magnetic field produced by the shock wave. These electrons are most likely left over from an earlier period when the rotational energy of the neutron star, or pulsar, was transformed into particles moving at a large fraction of the velocity of light.

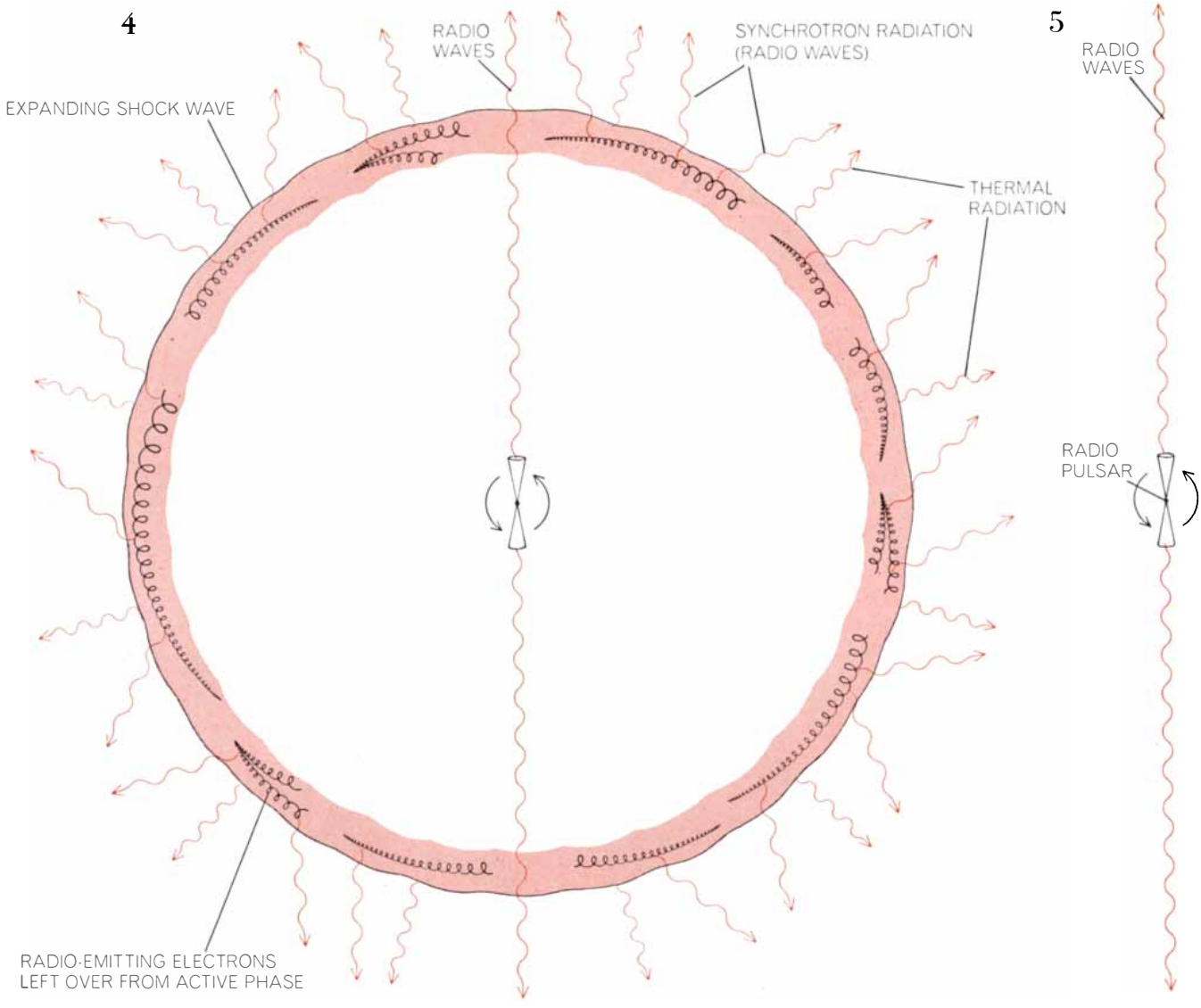
We can summarize by saying that if a pulsar is formed, during the first 1,000 years or so following the explosion the emission is dominated by the interaction of the pulsar's high-energy electrons with the magnetic field of the nebula, resulting in the emission of X rays, visible light and radio waves. When the remnant has attained an age of about 10,000 years, the expanding shock wave generated by the explosion will have swept up a large amount of interstellar matter and heated it to a very high temperature. During this phase there will be thermal emission of X rays and ultraviolet radiation. Thus for a few tens of thousands or perhaps hundreds of thousands of years the radio emission and the glow behind the shock wave will mark the position in space where a supernova explosion once occurred. After that, for observers situated in the right plane, the ever slower radio pulses from the neutron star may continue for millions of years, a reminder of one of the most cataclysmic events in all nature.



EVOLUTION OF A SUPERNOVA REMNANT is depicted in this sequence of drawings, beginning with the star in its presupernova state (*1*). When the star explodes (*2*), the outer shell is ejected into interstellar space, the core collapses to form a neutron star with a radius of about 10 kilometers, and enormous fluxes of radiation are emitted. In the active supernova stage (*3*) the highly magnetized, rapidly rotating neutron star at the center produces relativistic electrons, which spiral out into the nebula, emitting X rays, visible light and radio waves by the synchrotron mechanism. In addition, if the observer is in the right plane, pulsed radiation with a period equal to that of the rotating neutron star may also be observed. The Crab Nebula is an example of a supernova remnant in this stage.

Supernovas and their remnants are clearly very important on a stellar scale. One is prompted to ask if processes similar to those we have described can take place on the scale of galaxies and perhaps account for the tremendous power output observed in quasars, the starlike objects that seem to radiate as much energy as 50 to 100 galaxies. Supernova remnants and quasars are remarkably similar in a number of respects. Both, for example, are nonthermal sources of radio waves in which large fluxes of high-energy particles are being continuously generated, and in both the ultimate energy source is concentrated in a small volume. One proposal is that the activity of a quasar is nothing less than a chain of supernova explosions that produce a more or less steady source of radiation.

Another class of models is based on the idea that a rapidly rotating, highly magnetized object can be an efficient machine for producing high-energy electrons. If a neutron star can supply the energy for the Crab Nebula and possibly for other supernova remnants, then perhaps a much more massive rotating object in the core of a galaxy could account for the existence of quasars. In these models the rotator must have a mass of about a billion solar masses, a diameter similar to that of the solar system, a magnetic field between 1,000 and a million gauss and a rotation period ranging from a year down to a day, depending essentially on the radius of the rotator. The energy source is the kinetic energy of rotation; the rotation itself would result from gravitational collapse, as in the case of a neutron star. The rotational energy would be transformed into high-energy electrons and eventually into electromagnetic radiation by the large-scale, strong electric fields produced by the rotating magnetic field, in analogy with the pulsar in the Crab Nebula. Thus the solution to the riddle of the quasars may well lie in an understanding of supernova remnants. Whether or not this is so remains to be seen, but in any case supernova remnants will continue to serve as extraordinary laboratories where one can study fundamental processes of nature under extreme physical conditions.



In the next evolutionary stage (4) the energy of the expanding shell has gone into the production of a blast wave in the interstellar medium. The hot plasma behind the blast wave is a source of extreme-ultraviolet and X radiation. Relativistic electrons left over from the active stage produce radio synchrotron radiation when they encounter the high magnetic field in the region behind the shock wave. The central neutron star is no longer an important source of power for the nebula, although pulsed radio emission may still occur. The Cygnus loop is probably an example of a supernova remnant in this stage. In the final stage (5) the energy of the blast wave has been radiated away and all that remains is a radio pulsar, which may continue to emit ever slower radio pulses for millions of years.

A Grazing Ecosystem in the Serengeti

Large herds of zebras, wildebeests and Thomson's gazelles migrate in succession across the plains of Tanzania. These migrations are synchronized with the availability of specific tissues of grasses

by Richard H. V. Bell

The seasonal migrations of grazing animals over the African savanna are a remarkable natural spectacle. It is clear that these migrations, which often involve several thousand animals, are correlated with the transition from a wet season to a dry season and vice versa. Another influence that is not so clear has emerged from work over the past decade in the vast grasslands of the Serengeti region of northern Tanzania. It is that the grasses and the animals interact in such a way as to give rise to a succession wherein the grazing species follow one another in characteristic sequence during their seasonal movements. Studying the phenomenon in the Serengeti National Park, my colleagues and I have found that the principal migratory species (wildebeest, zebra and Thomson's gazelle) have physiological reasons for seeking different tissues of the grasses and herbs (the principal tissues being the leaves, the sheaths and the stems) and so tend to move on when the preferred tissue becomes scarce or hard to get at. Changes in the availability of the plant tissues determine the order of movement of the animals. In the Serengeti the zebras move first, followed successively by the wildebeests and the Thomson's gazelles. The semimigratory topi, which is also numerous in the region, tends to associate with the zebra.

The Serengeti National Park is an area of 5,600 square miles of grassland and open woodland. It forms the central part of the Serengeti Ecological Unit, an area of about 9,000 square miles. The National Park and the controlled hunting areas that comprise the Ecological Unit are agencies of the Tanzanian government. Ecological research in the area has been conducted for almost 14 years, beginning with a survey in 1957 by W. H. Pearsall of University College London and including pioneering aerial studies by the German zoologist Bernhard Grzimek and his son Michael. These studies were followed by the Serengeti Research Project and the establishment of its successor the Serengeti Research Institute, which have been operated by the Tanzanian National Parks with aid from several international organizations. From these studies the elements of the system underlying the complex mosaic of life in the Serengeti region have begun to emerge.

More than half a million wild animals live in the Serengeti Ecological Unit. Most of them are migratory; indeed, the ecological unit has been defined by R. M. Watson of the Serengeti

HERD OF WILDEBEESTS grazes on the Serengeti Plains in the Serengeti National Park. In its seasonal migration across the region the wildebeest follows the larger zebra and is

Research Project as the area regularly occupied by the migratory populations in the course of their seasonal movements. The most important relationship in the Serengeti ecosystem is the use of the herb layer—the grasses and herbs—by the grazing ungulates, which constitute more than 90 percent of the mammalian biomass. In order of body size the most numerous ungulate species are the buffalo (*Syncerus caffer*), the zebra (*Equus burchelli*), the wildebeest (*Connochaetes taurinus*), the topi (*Damaliscus korrigum*) and Thomson's gazelle (*Gazella thomsoni*). One can understand the pattern in the way these animals make use of the herb layer only by starting at the level of the plant cell.

Every plant cell is surrounded by a relatively inflexible wall, which is largely composed of the carbohydrate cellulose. The wall forms a structural box that contains the cytoplasm, which has proteins and soluble carbohydrates as its most important constituents. At all levels above the cell (that is, the levels of the tissue and the entire plant) the structure of the plant material depends on the cell wall. In young growing tissues, which have to expand and are not required to support weight, the cell wall is thin and fairly flexible. This condition is characteristic of the leaves and fruits of grasses and other flowering plants. In older tissues, particularly those that bear weight, the cell wall becomes thickened and hardened by other compounds, notably lignin, and the cytoplasm is reduced. This is the situation in the stem of a tall grass plant or a shrub. Finally the lignified cell wall comes to occupy most of the cell. The result is wood.

For the herbivores the division of their plant food into cell wall and cytoplasm is important because it has a profound influence on the availability of their two primary nutritional requirements: protein and energy. Mammalian herbivores do not have digestive enzymes of their own that can break down the cellulose cell walls of their plant food. The animals succeed in attacking the cell walls by harboring in their stomachs or intestines microorganisms (bacteria and protozoa) that do possess cellulolytic enzymes. Even with the help of these microorganisms, however, the breakdown of cellulose is slow. Moreover, it is greatly retarded by the presence of lignin. The cell wall therefore constitutes a physical obstacle to the extraction of the cell's contents, particularly in structural tissues where the wall is thick and lignified.

Once the cell wall has been broken down it is available to the herbivore as a source of energy. Additional energy can be obtained from the cytoplasm. The total amount of energy that the animal can get from the plant exceeds the amount of protein, which comes from only part of the cell and is in even shorter supply in structural plant tissues where the cell wall is thick.

For these reasons the herbivores must confront the problem of extracting enough protein from a food supply containing superabundant and obstructive carbohydrates. The animals cope with the problem at both the intake stage and the digestive stage of feeding. At both of these stages there is a basic difference in strategy between the rumi-

followed by the smaller Thomson's gazelle. The zebra tends to feed on the upper parts of grasses and herbs, the wildebeest on the middle parts and the Thomson's gazelle on the lower parts. Earlier animals open up food supply of later ones by grazing and trampling.

nant animals (cattle, sheep, antelope, deer) and the nonruminant ones (horse, zebra, elephant). The ruminant approach can be characterized as being selective and the nonruminant approach as being tolerant.

The strategy of the ruminant digestive system can be summarized as maximizing the efficiency of the use of protein at the expense of (indeed, in spite of) the superabundant supply of energy. The special characteristics that achieve this result are the system of pregastric fermentation (which entails having the site where cellulose is fermented ahead of the site where protein is absorbed) and the cycle of regurgitation and mastication. The mechanism operates as follows. Food ingested at the mouth is masticated and exposed to the salivary enzymes. When it is swallowed for the first time, its components are still fairly large and many plant cells are still intact. The rumen acts as a filter that operates on the basis of specific gravity: particles that are of large size and therefore of low specific gravity float on the surface of the rumen liquor, so that they are passed into the diverticulum and are then regurgitated. They repeat the cycle until the mechanical and chemical breakdown of the plant cells is far enough advanced for them to sink in the rumen liquor and be passed on to the intestine. This process ensures relatively complete extraction of the cell contents and accounts for the high digestibility of protein found in feeding trials with ruminants.

Clearly, however, the rate of passage of food through the gut is limited by the rate of breakdown of material in the rumen. The breakdown rate is largely dependent on the quantity and quality of cell wall in the food. Feeding trials show that when the diet contains much cell wall, particularly lignified wall, the rate of passage of food in ruminants is quite low. As a result, even though the degree of extraction of protein is high, the overall rate of assimilation of protein in these conditions is low.

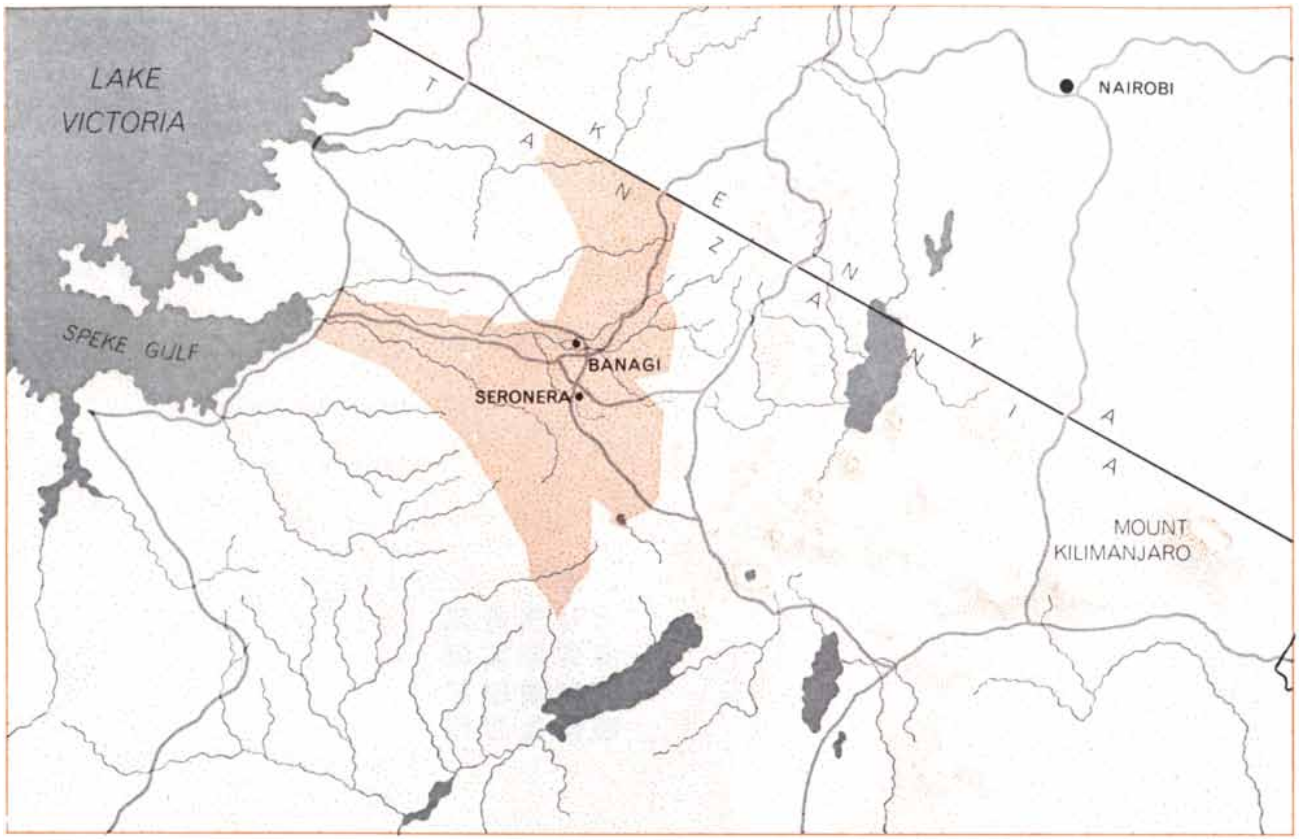
When cell-wall constituents reach a certain concentration in the food, the ruminant fails to assimilate enough protein to meet its maintenance requirement. In this situation the animal is forced to select components of the vegetation that have thin cell walls and a high concentration of protein, namely the leaves, fruits and shoots of grasses and other plants in the herb layer. At the same time, however, the ruminant's system of pregastric fermentation recycles and reconstitutes food constituents in such a way as to ensure that once the protein has been extracted it will be used to best advantage. The strategy of the ruminant, then, is based on high efficiency of extraction and utilization of protein at the expense of a high rate of intake and processing of food, with the consequent emphasis on selecting for high-protein plant components.

Much less information is available on the nutrition of nonruminants. What is known comes mainly from tests with the domestic horse, of which the zebra is a close relative. The horse differs from the ruminants in that the sites of fermentation of cellulose are the enlarged colon and the large intestine, whereas the principal site for extracting and assimilating protein is the relatively small and simple stomach. Therefore the animal is extracting protein without either the mechanical maceration or the cellulolytic attack that in the ruminant breaks down the cell walls. In the nonruminant the extraction of protein is depressed even more by an increase in the quantity of cell wall than it is in the ruminant. Moreover, once the protein has been extracted in the horse's stomach it is immediately assimilated as amino acids and is not subjected to the recycling and reconstitution that increase the efficiency of protein utilization in the ruminant.

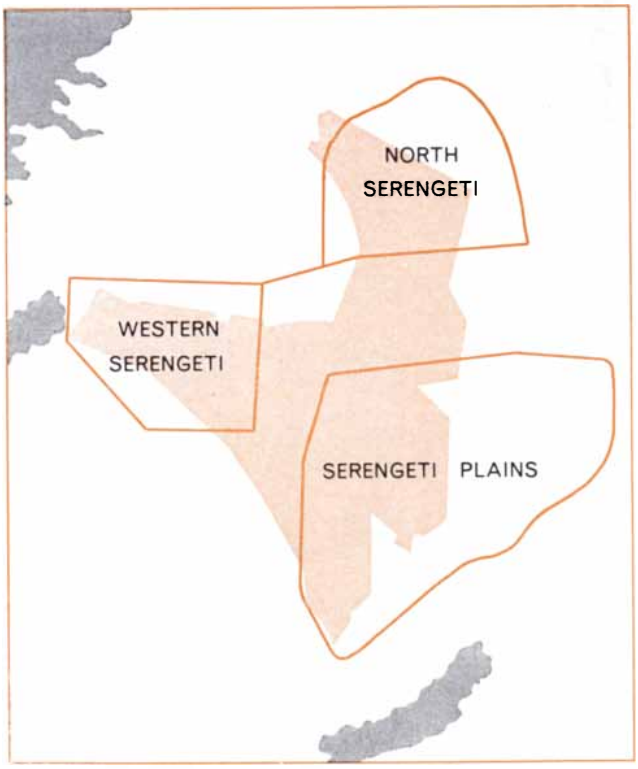
For these reasons the horse is less efficient than the ruminant in utilizing the protein in its food. When food is limited, the ruminants will survive longer than the horse. How, then, do the horses maintain a place? The answer lies in the different digestive system of the horse. The system has no mechanism that imposes a limit on the rate of passage of material through the gut. Comparing the nonruminant horse with the ruminant cow, one finds that material passes roughly twice as fast through the gut of the horse as it does through the gut of the cow (30 to 45 hours in the horse, 70 to 100 hours in the cow).

The importance of this difference emerges in a simple calculation: If the horse is two-thirds as efficient as the ruminant in extracting protein from a food but processes twice as much food in a given period of time, its rate of assimilation of protein per unit of time

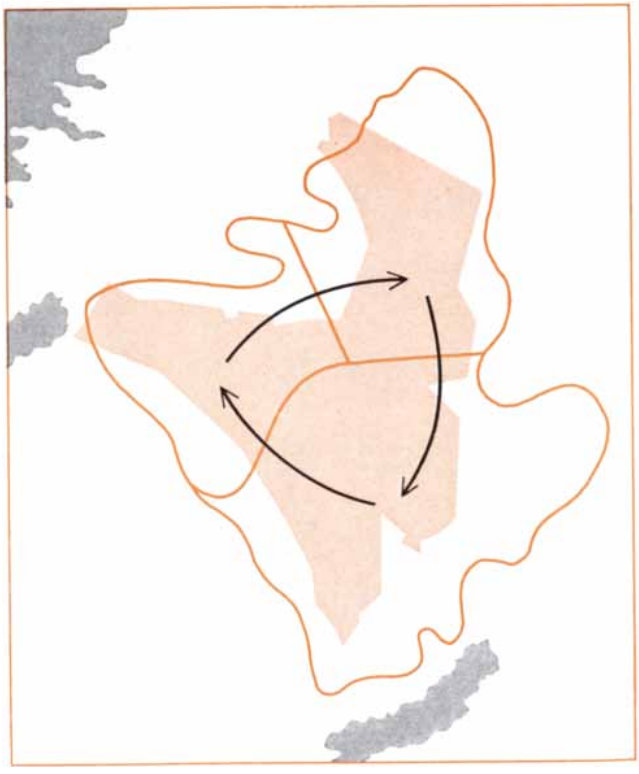
SERENGETI REGION lies along the border between Tanzania and Kenya in East Africa. It includes the Serengeti National Park, which consists of 5,600 square miles of grasslands and is within the larger Serengeti Ecological Unit, which covers 9,000 square miles.



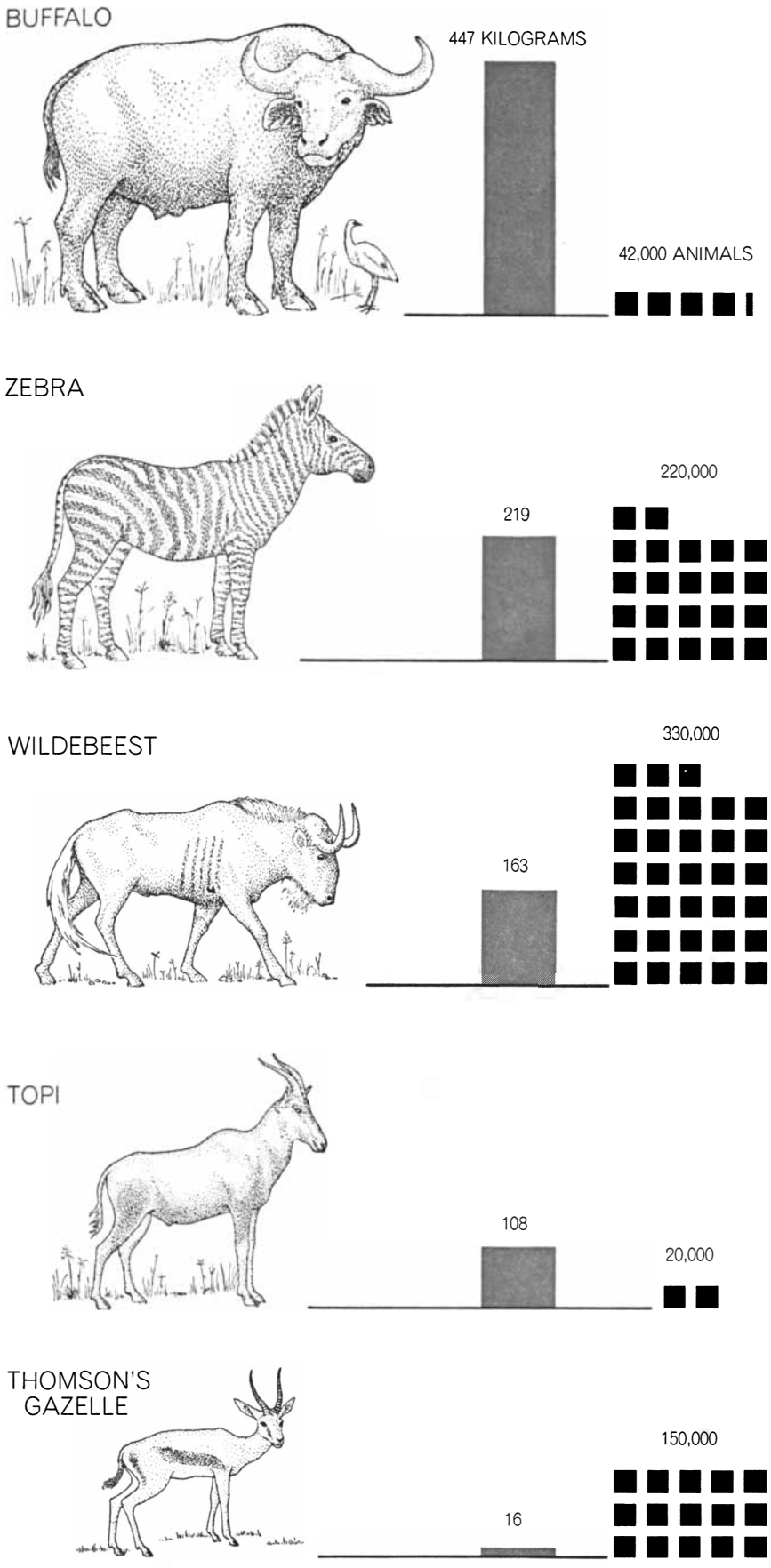
SERENGETI NATIONAL PARK (*color*) lies east of Speke Gulf of Lake Victoria. Seronera is headquarters of the park; Banagi was headquarters of the Serengeti Research Institute until 1967. The broad gray lines represent the principal roads of the region.



ECOLOGICAL DIVISIONS are characterized by high rainfall and long grass (north), low rainfall and short grass (plains region) and intermediate rainfall and intermediate grass (western Serengeti).



SEASONAL RANGES of migratory populations are outlined. The animals spend wet season (March through May) on the Serengeti Plains, early dry season in the west and late dry season in the north.



MAIN GRAZING SPECIES of the Serengeti region are depicted in order of their body weight, which is given in kilograms for an average adult female (*vertical bars*). Population of each species in Serengeti is shown at right, with each square representing 10,000 animals.

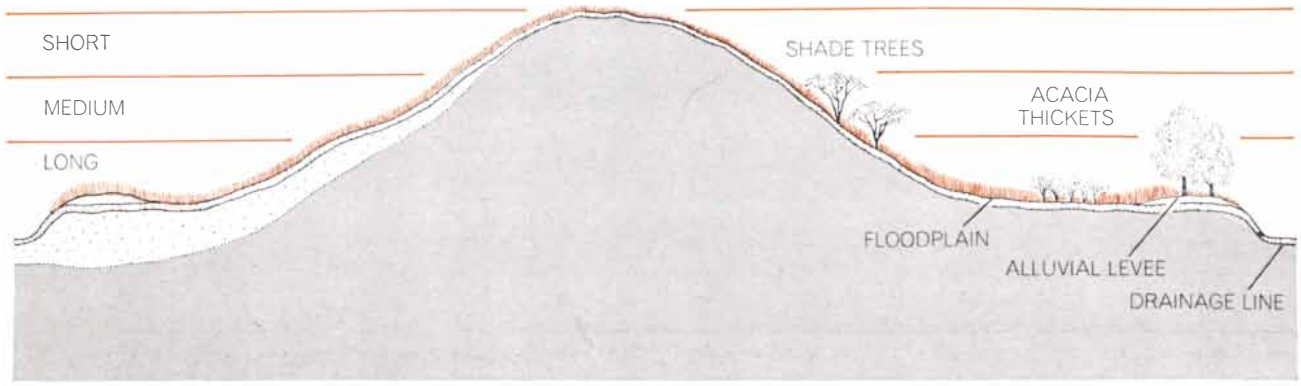
is four-thirds of the ruminant's rate. Therefore the horse (or the zebra) can support itself on a diet that is too low in protein to support a ruminant. In other words, the nonruminant is much more tolerant than the ruminant of cell wall in the diet but must maintain a high rate of intake to survive on food of this type, so that the selectivity of its feeding is much less intense than that of the ruminant.

Notwithstanding the broad division in feeding type between ruminants and some nonruminants, there is in fact a certain amount of overlap of the two groups arising from the effect of body size on the metabolic rate and therefore on the maintenance requirement for protein and energy. In general the smaller the animal, the higher the metabolic rate. In considering the matter as it affects the ecology of wild animals, one must take into account both the relative maintenance requirement (per unit weight per unit time) and the absolute maintenance requirement (per animal per unit time).

With regard to the relative maintenance requirement, the effect of the higher metabolic rates of small animals is that they need more protein and energy per unit weight per day than larger animals. For example, the daily requirement of protein of a Thomson's gazelle per unit weight is from two to four times higher than the requirement of the larger wildebeest. The result of this difference is that smaller animals lose weight more rapidly than larger animals on equal but submaintenance diets. Hence if two species of different size have the same food supply, the larger species will displace the smaller one.

Although the relative maintenance requirement is higher for a small animal than it is for a large one, the reverse situation holds for the absolute maintenance requirement. Again comparing a Thomson's gazelle and a wildebeest, the gazelle needs to consume only about 20 percent of the amount of food in a day that a wildebeest would require. (The calculation assumes equal quality of intake; the gazelle's requirement would be still lower if it were more selective than the wildebeest in feeding.) In other words, the gazelle has five times as much time as the wildebeest does to find and eat enough food to maintain itself. The smaller species can thus afford to be more selective than the larger one and can maintain itself on a food supply so sparse that the rate of intake would not satisfy the larger animal.

The survival of a small animal in place



CHARACTERISTIC TOPOGRAPHY of the Serengeti and much of the African savanna is a series of undulations. The term used to describe this topography is "catena." The distribution of long and short grass is affected by the drainage from the elevations. Water runs off the elevations and into the depressions, so that the grass in the depressions tends to be longer than the grass on the elevations. In longer grass some of the grazing animals have difficulty reaching the parts of the grass that are most nourishing for them.

of a large one requires two conditions: a high rate of assimilation of protein and energy per unit of body weight and a low absolute intake of protein and energy per animal. These conditions are supplied by components of the vegetation that have low concentrations of cell-wall constituents. Such components are often sparsely distributed and relatively inaccessible.

The larger the animal, the greater its success in conditions that depart from these, that is, the greater its tolerance for plant tissues containing higher proportions of cell wall. The more common and accessible these tissues are, the larger the animal is that can survive on them. This latter condition applies particularly to the nonruminants such as the zebra, because of their necessarily high rate of intake and their high tolerance for cell-wall material. As a result these animals have many of the characteristics of ruminants disproportionately larger than themselves.

It should be emphasized that the optimum diet for all these animals is an abundant supply of plant tissues containing a high proportion of protein and a low proportion of cell wall. The differences that I have been discussing are differences of tolerance for foods that depart from the optimum. Small animals can tolerate diets that depart from the optimum in quantity; large animals, particularly nonruminants, can tolerate departures in quality.

In order to see how the nutritional characteristics of the principal members of the Serengeti grazing fauna impose a pattern on their use of the food supply and their seasonal movements, it is necessary to examine the structure and distribution of the food supply in more detail. These considerations depend largely on the growth form of the individual grass plant. At the start of the wet season the growing shoot appears above the ground. In this early stage growth is just below the apex of the shoot, forming a series of leaves that grow up around the apex, enclosing it in their sheaths. The apex and the growing point stay near the ground. At this stage the entire plant consists of growing tissues with thin cell walls, and the leaves are easily accessible to grazers. Moreover, animals can eat the leaves without destroying the protected shoot apex and the growing point.

The next phase of growth comes later in the wet season. The stem extends below the growing point, raising the leaves away from the ground in response to the competition with other plants for light. This stage entails a great reduction in the protein content of the plant as a whole because of the structural nature of the stem.

How far the extension of stems is carried depends on two conditions. The first is the length of the growing season: taller grasses develop in wetter areas when the growing season is longer. The second is the frequency of defoliation by grazing or burning: the more the upper parts of the plant are grazed or burned, the greater the advantage is to the plant in remaining prostrate.

In the final stage the stem above the growing point extends to form the flowering culms of the grass. This part of the plant resembles the stem in containing a high proportion of cell wall to cell content. (The seeds borne by the culm are an exception, but grazing animals do not make much use of them.)

Thus at the end of the wet season the herb layer presents a series of levels of different food value and accessibility to the herbivores. The upper level consists of the low-protein culms of the grasses. Below this level are the stems and leaves of the taller grasses, and lower still are the leaves of the smaller grasses and young shoots. Also at the lowest level are the herbs proper, such as clover in a hayfield. The significance of the series of levels from a grazing point of view is that the highest concentration of protein is in the lowest level, which is relatively inaccessible to grazers because of the dense grass stems and culms above it.

Here, then, is a type of heterogeneity of exactly the kind that is important to grazing animals. It entails differences in the chemical composition and the accessibility of the food supply. For this reason we examined the stomach contents of four of the five main grazing species of the Serengeti (buffalo were not included in the study) to test the possibility that the grazers select from different levels of the herb layer.

Our procedure was to separate the stomach contents according to whether the plant was a monocotyledon (grass) or a dicotyledon (most other plants in the herb layer) and also according to parts of the plant: fruit, stem, leaf, sheath. Of the four grazing species only one, the Thomson's gazelle, contained a large amount of dicotyledon material: 40 percent, chiefly fruit. This is precisely the high-protein and high-energy package expected for that small animal. The stomachs of the three remaining species contained almost entirely grass and showed a clear separation in the proportions of leaf to stem (and culm). The wildebeest contained the most leaf, the zebra the least, and the topi was intermediate.

It therefore seems probable that these

four species take components selectively from different levels of the herb layer. Zebras take the diet highest in cell-wall constituents, from the top of the herb layer; topi and wildebeests take a higher-protein diet from the leafy levels below, and gazelles select the high-protein fruits from the ground. Except for the topi, this series corresponds with the order of decreasing body size.

These differences of selectivity, imposed by the size and physiology of the animals, have an important effect on the behavior of the ungulates in the field. The reason was shown experimentally by G. W. Arnold of the University of London in trials with sheep. The trials showed that the rate of intake by these animals is severely limited by the presence of unwanted components of the vegetation—in this case grass stem. Clearly the degree of interference by stem is greater the lower down in the herb layer the desired component is situated. Hence it is greater for Thomson's gazelles than for zebras. Similarly, for all species the interference is greatest in long-herb layers with high proportions of stem.

The differentiation of the levels in the herb layer is most marked when the layer is tall. From the point of view of the grazing animals, therefore, the most important characteristic of the herb layer is its length. Here one confronts again the implications of the plant cell: the structural constituents of the food supply act as an obstacle to the intake of the high-protein, cytoplasmic constituents of the plant.

In the Serengeti, and over much of the African savanna, one finds a pattern underlying the distribution of long and short grass. It is based on the drainage system. No area is perfectly flat; the ground always has undulations, which vary from a few inches to hundreds of feet in height and from several feet to several miles between peaks. The term used to describe this topography is "catena." Rain falling on the undulations runs off the slopes into the depressions, so that the lower areas have access to more water for longer periods of time than the raised areas and peaks. The lower areas also develop heavier soils with a substantial clay content. For this reason the grass is longer in the depressions than it is on the slopes and peaks. The series of soil and grass types on the catena is called the catenary sequence [*see illustration on preceding page*].

This pattern in the quality of the food supply has a dominant effect on the seasonal movements of the grazing animals. In the wet season, when the grasses are at their longest, all the grazers concentrate in mixed groups on the raised areas where the grass is shortest. There, by feeding and the trampling associated with it, they keep the grass in a short, growing stage of development. Under these conditions the animals have no difficulty eating grass leaves and hence obtaining sufficient protein to maintain themselves. Indeed, at this time there is usually a food surplus. It is the time when most of the calving and lactation occur. The raised parts of the catena have the additional advantage of being well drained and easier underfoot than the often waterlogged depressions.

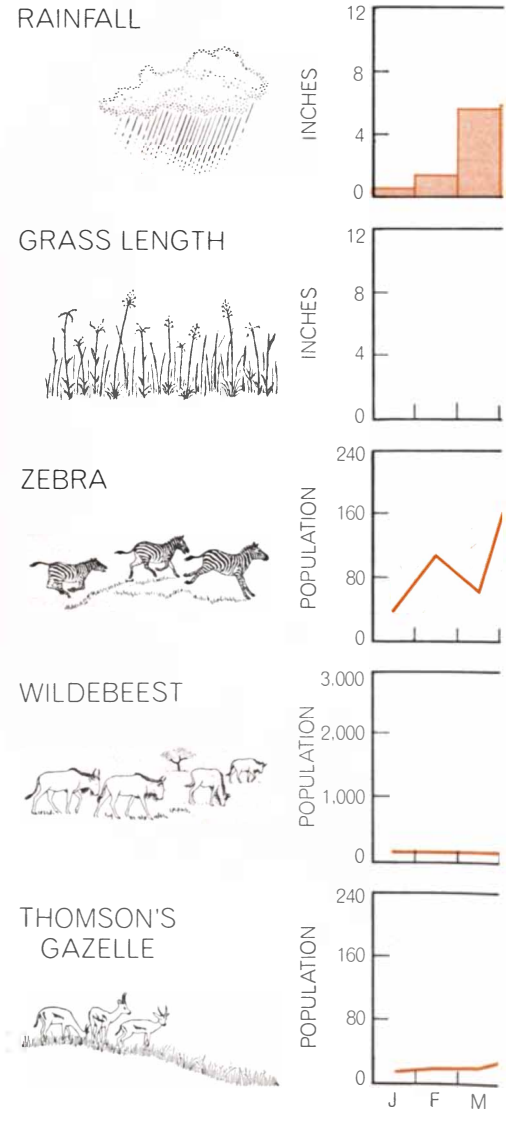
In the dry season the grass stops growing, and the short grass of the raised areas is progressively removed. As the amount of food available in those areas declines, the various species of grazers progressively become unable to maintain themselves and are forced in turn to descend into the longer grasses of the lower areas. The changes first affect the large animals (buffalo and zebra) and subsequently come to bear on the smaller animals (topi, wildebeest and Thomson's gazelle). The sequential descent of the species into the depressions of the catena is termed the grazing succession. The order of the succession corresponds to the order of selectivity in diet and, except for the topi, to the order of body weight.

In the longer grasses of the depressions it is more difficult for the grazers to select sufficiently high-protein components from the vegetation because of the preponderance of stem. This is particularly true for the smaller animals, which select from the lower levels of the herb layer. The activity of the earlier members of the succession in breaking down and opening up the dense stands of stems and culms by grazing and trampling is therefore of great assistance to the later members of the succession. The relation is expressed at this time, as in the wet season, by the association of the grazing species in mixed groups and can be documented quantitatively by a scheme of coefficients of association. The coefficients reach peaks in the wet season and the late dry season.

The conclusion is, then, that the earlier members of the succession prepare the structure of the vegetation for the following members. This carries the important implication that a reduction in the numbers of one species by disease or hunting could lead to a reduction in the numbers of another species. A decline in the population of zebras, for example, could be expected to lead to a decline in the population of the smaller wildebeests.

Up to now I have confined my discussion to the localized seasonal movements of the grazing animals in the northwestern half of the Serengeti Ecological Unit. Similar situations have been reported by workers in other regions of East Africa. I turn now to the Serengeti's most dramatic feature: the massive migration of zebras, wildebeests and Thomson's gazelles over an area about the size of Massachusetts.

These migrations are in reality exactly equivalent to the localized movements, but on a much larger scale. By a series of geological and climatic accidents the Serengeti region as a whole consists of three zones, arranged in a broad triangle, that represent on a large scale the



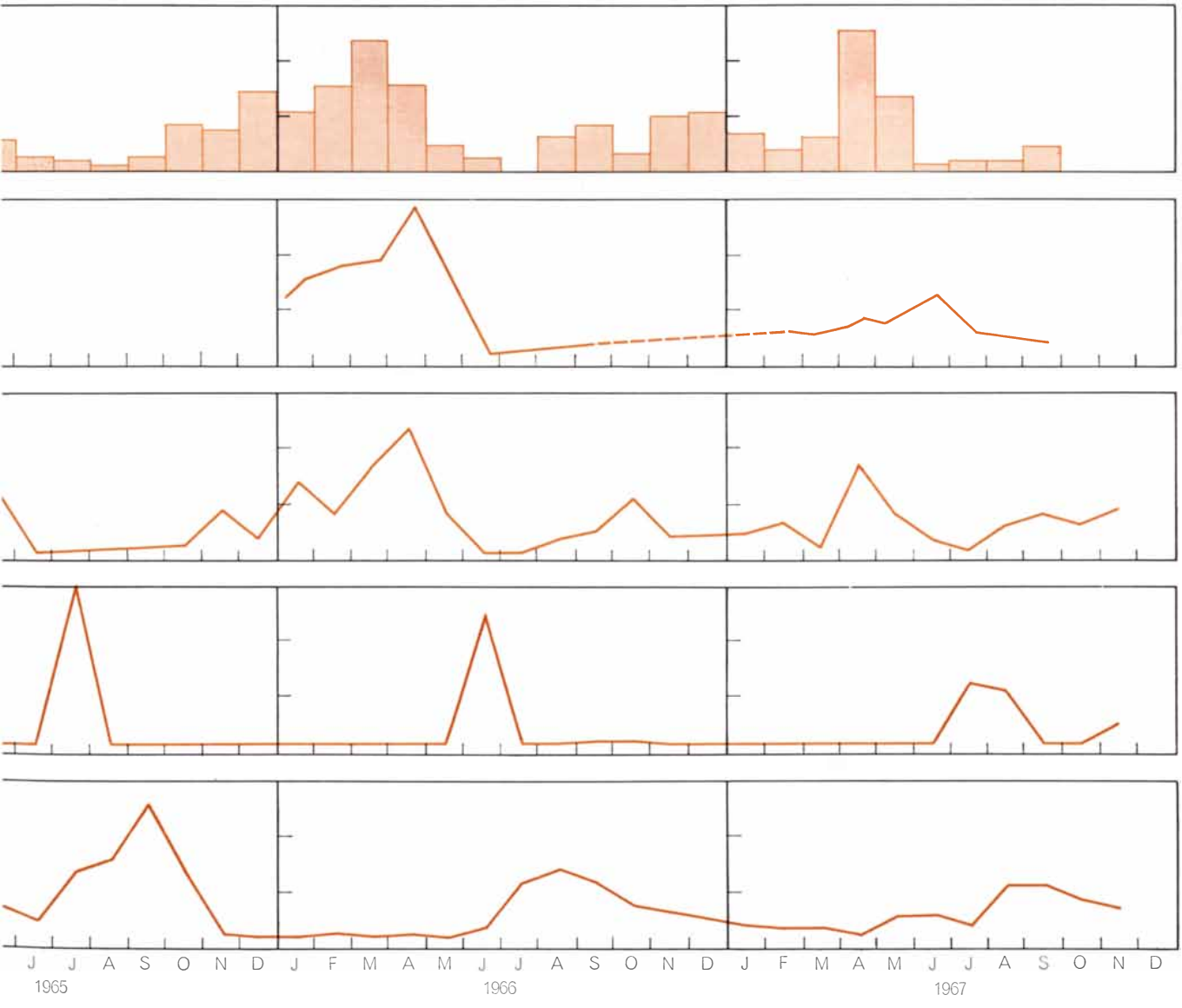
POPULATION OF MIGRATING SPECIES is shown in relation to rainfall and length

topographic undulations of the catena. In the southeast the huge open expanse of short grass of the Serengeti Plains, formed by a combination of scant rainfall and dusty volcanic soil, is equivalent to the raised areas of short grass where the animals concentrate in the wet season. Indeed, this is the site of the great concentrations of mixed species of the grazing animals in the wet season.

In the west the rainfall is higher and the grass is longer. This is where the migrators move at the beginning of the dry season. Finally, the north is the area of highest rainfall and longest grasses; into this area the migrators move at the end of the dry season. Just as in the localized seasonal movements, the migrations show a grazing succession, with the zebras leading, the wildebeests following and the Thomson's gazelles last. The pattern is repeated (with variations) all over the African savanna.

One element of the picture is lacking. Man has been part of this system from his earliest days as a predatory ground ape depending on the large game populations to feed himself in spite of his inadequate weapons; through his period as a hunter with bows and spears; through his period as a pastoralist, depending on domestic stock that shared the grasslands with the wild populations; through his period as a cultivator, at last in direct competition with the wild animals, and to his period as a tourist, escaping from his cities to revisit his original savanna.

The impact of pastoral man on the grassland, through burning and the grazing of domestic stock, is ancient. It has tended to induce and maintain areas of short grass. In other words, pastoral man has been a member of the grazing succession—not in conflict with the wild animals but in cooperation with them. Indeed, one finds evidence in the Serengeti region today that the largest concentrations of wild animals are in areas of present and past pastoral activity. In contrast, the impact of cultivation and fencing directly conflicts with the pattern of movement of the wild animals. This phase of human activity is causing a reduction in numbers and in range of the savanna fauna. It is up to man the scientist, who represents the latest human phase, to understand the working of the system as a whole and to suggest integrated patterns of land use, taking into account the interests of man the pastoralist, man the cultivator and man the tourist.



of grass for nearly three years. The figures were obtained in the western Serengeti by a series of daily transects in a strip 3,000 yards long and half a mile wide. Successive peaks during each year mark the passage of the main migratory species in the early dry season.

PHOTONS AS HADRONS

Photons with a billion times more energy than photons of visible light exhibit properties once thought to belong solely to hadrons: the class of particles that includes the proton and the neutron

by Frederick V. Murphy and David E. Yount

Unraveling the behavior and character of light has kept scholars busy and disputatious for centuries. In exhibiting properties associated with both waves and particles, light introduced investigators to the paradoxes that nature presents when it is examined at its most fundamental level—paradoxes that ultimately provoked the development of quantum mechanics. In the three decades that followed this achievement, the behavior of light quanta, or photons, became so well understood and so accurately predictable that it hardly seemed possible that further surprises lay in wait.

Surprises there have been. Over the past few years experiments carried out with high-energy accelerators have disclosed that the photon can exhibit properties once thought to be possessed exclusively by such relatively massive particles as the proton and the neutron: the heavy, durable constituents of the atomic nuclei. Such particles, which exhibit the "strong" force that binds the nucleus together, are called hadrons. It has been found that a photon with a billion times as much energy as a photon of visible light behaves as hadrons do when it is allowed to interact with hadrons. The discovery that the photon has this hadronic character at very high energies, although it was unexpected, has now been incorporated into a new group of theories that make the hadronic photon respectable.

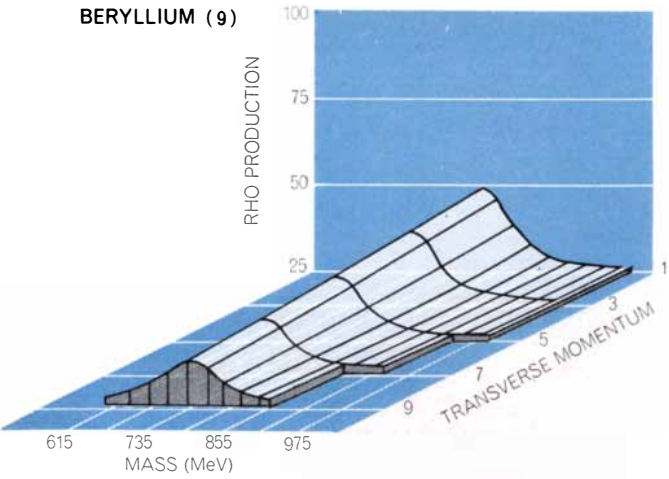
Before the 17th century it was generally believed that light consisted of a stream of particles emitted radially from light sources such as the sun and a candle. Certain materials were transparent to these particles; others were opaque or translucent and tended to reflect or absorb them in varying degrees. In 1678 Christiaan Huygens showed that if light consisted of successive waves rather than independent corpuscles, the laws of reflection and refraction could be explained in a simple way. Early in the 19th century Thomas Young and Augustin Jean Fresnel independently carried out experiments on the interference and diffraction of light that were consistent with wave motion and that even measured the wavelength of light.

In the latter half of the 19th century James Clerk Maxwell succeeded in summarizing the phenomena of electricity and magnetism with four equations that are still the fundamental operating principles of all large-scale electromagnetic devices, from electromagnets, electric motors and electronic computers to particle accelerators. Maxwell's equations predict that an oscillating electrical circuit will radiate electromagnetic waves whose velocity can be calculated from purely electrical and magnetic measurements. When the calculated velocity turned out to be precisely the velocity determined experimentally for light, physicists were convinced that light consists not only of waves but also of the particular kind of electromagnetic waves predicted by the electromagnetic equations. Thus Maxwell's equations led to the concept of a broad electromagnetic spectrum extending to each side of the visible spectrum and including the radio waves discovered by Heinrich Hertz in 1886.

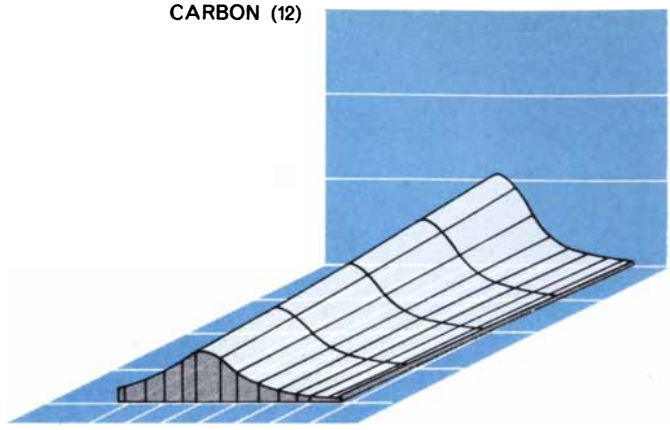
The corpuscular nature of light was reintroduced by Albert Einstein in 1905 to explain how light shining on a metal ejects electrons from the metal's surface. The light energy, instead of being distributed in an extended wave, is concentrated in tiny packets called photons. When a single photon strikes a surface electron in the metal, the photon is completely absorbed and imparts all its energy to the electron. Curiously, Einstein predicted, and Robert A. Millikan experimentally proved, that the energy of the photon, and hence of the ejected electron, depends only on the frequency of the light—as if each photon were simultaneously a wave. Another example of the corpuscular behavior of light was discovered in 1921 by Arthur H. Comp-

PRODUCTION AND SCATTERING OF RHO MESONS are studied by bombarding targets of different nuclear mass with high-energy photons. The lifetime of the rho meson is so brief (about 10^{-23} second) that it usually decays into two pions before it can reach another nucleus that could scatter it. Here the rate of rho manufacture, inferred from the decay products of rho particles, is plotted for eight different target materials, ranging from beryllium, which has only nine nucleons (protons or neutrons), to uranium, with 238. The rate of rho production per nucleon in the target is plotted on the vertical axis in arbitrary units. The rho mass appears spread out along the horizontal axis; the units are millions of electron volts (MeV). The third axis represents the square of the transverse momentum imparted to the target nucleus; the units are $(GeV/c)^2 \times 10^{-3}$. (GeV is billion electron volts; c is the speed of light.) For a given value of transverse momentum the area under the bell-shaped curve representing the rho mass distribution increases at first with the number of nucleons in the target. In the vicinity of copper (64 nucleons) the rho production per nucleon begins to level off or even decrease with increasing number of nucleons. In the heavier nuclei more and more rhos are being produced but they are also being increasingly absorbed. The data were provided by Deutsches Elektronen-Synchrotron (DESY) in Hamburg. A group from the Massachusetts Institute of Technology collaborated with DESY on the experiment.

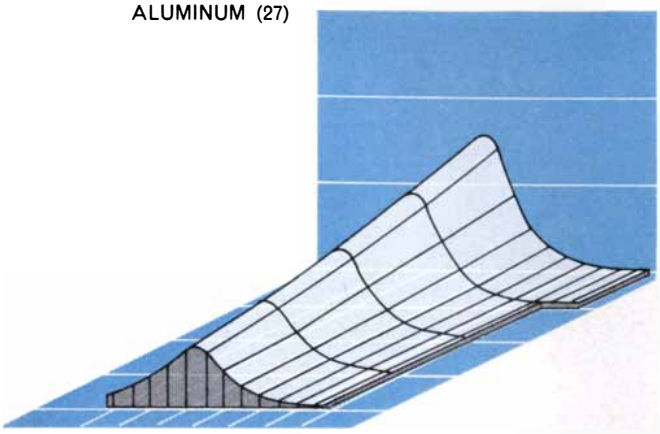
BERYLLIUM (9)



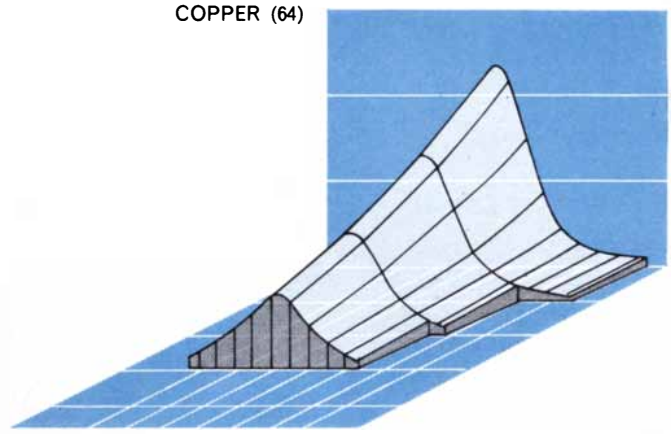
CARBON (12)



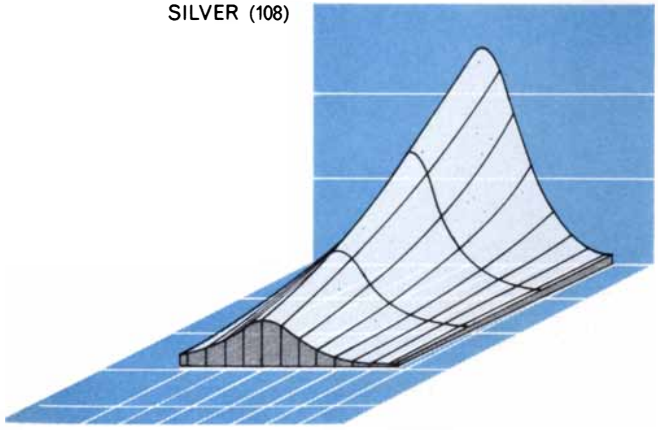
ALUMINUM (27)



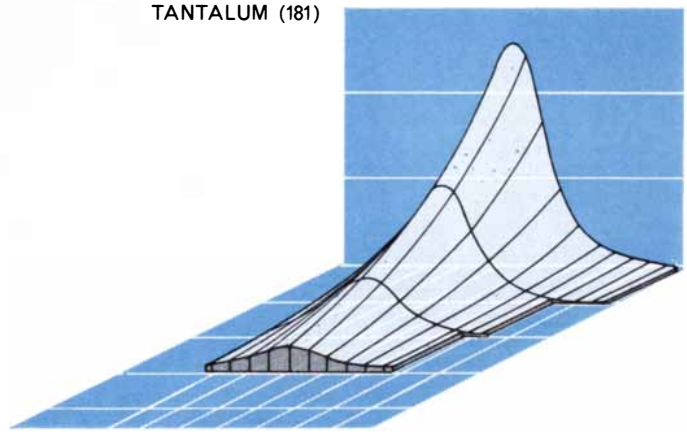
COPPER (64)



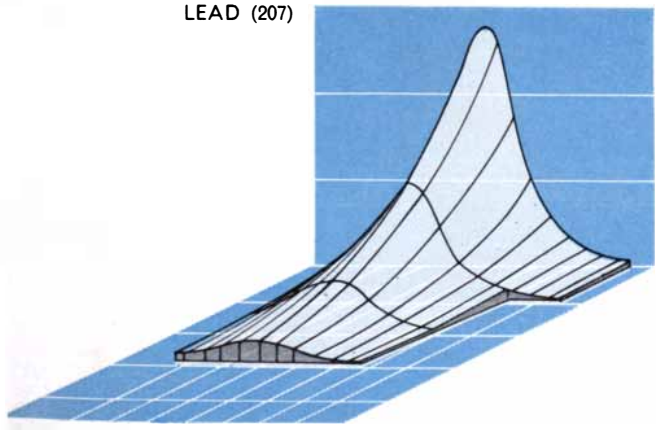
SILVER (108)



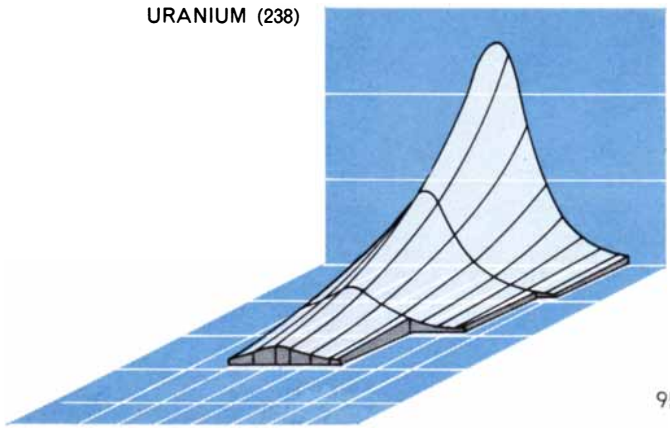
TANTALUM (181)

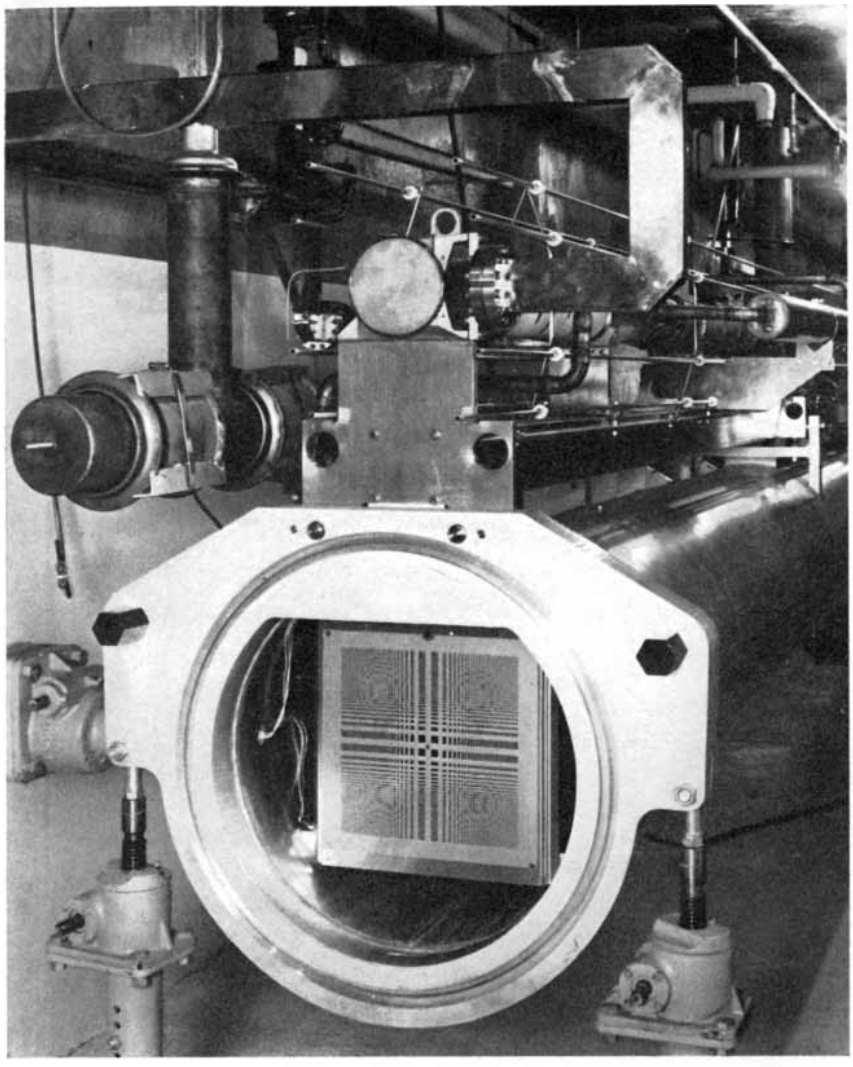


LEAD (207)



URANIUM (238)





ALIGNMENT OF TWO-MILE ACCELERATOR at the Stanford Linear Accelerator Center (SLAC) is done with the aid of a laser beam and Fresnel targets. The targets are located every 12.2 meters in a light pipe that is rigidly attached to and below the one-inch tube that carries the beam of accelerating electrons. When the targets are struck by a laser beam, they produce a characteristic diffraction pattern if the section of the tube is properly aligned. Three jacks for aligning the tube can be seen at far left, bottom left and bottom right.

ton, who observed that photons are scattered when they collide with electrons: the photons behave as material bodies having energy and momentum that are conserved in such collisions. The wave-particle duality of light is reconciled in quantum electrodynamics, perhaps the most elegant and successful theory in physics. Quantum electrodynamics predicts the electromagnetic interactions of photons with such great precision and over such a wide range of energies (more than 10^{18} to one) that many rigorous tests are possible. Since physics often advances by shattering theories, the continuing success of quantum electrodynamics has been greeted almost with dismay.

Until quite recently the interaction of photons with matter was believed to be entirely electromagnetic, that is, it appeared to involve only the electric charges and magnetic fields of the target particles and did not seem to invoke their possible hadronic, or strong, interactions. Furthermore, photons serve as the sole mediator of all other electromagnetic interactions: two electrons repel each other by firing photons back and forth. The exchanged photons exist for exceedingly short times and cannot be directly detected. For this reason they are called "virtual" [*see "a" in illustration on opposite page*].

When an electron is scattered by another electron or by any other charged particle, it is sharply accelerated and may radiate additional photons that are not reabsorbed. This radiation, called *bremsstrahlung*, is analogous to the emission of radio waves by an oscillating electrical circuit. Like radio waves, *bremsstrahlung* photons can be detected and are therefore called "real" [*see "b" in illustration on opposite page*]. Physicists apply their knowledge of electromagnetic interactions to produce large numbers of real or virtual photons of known energies under controlled conditions, that is, to produce photon beams. The availability of such controlled high-energy photon beams was the key to discovering the hadronic behavior of photons.

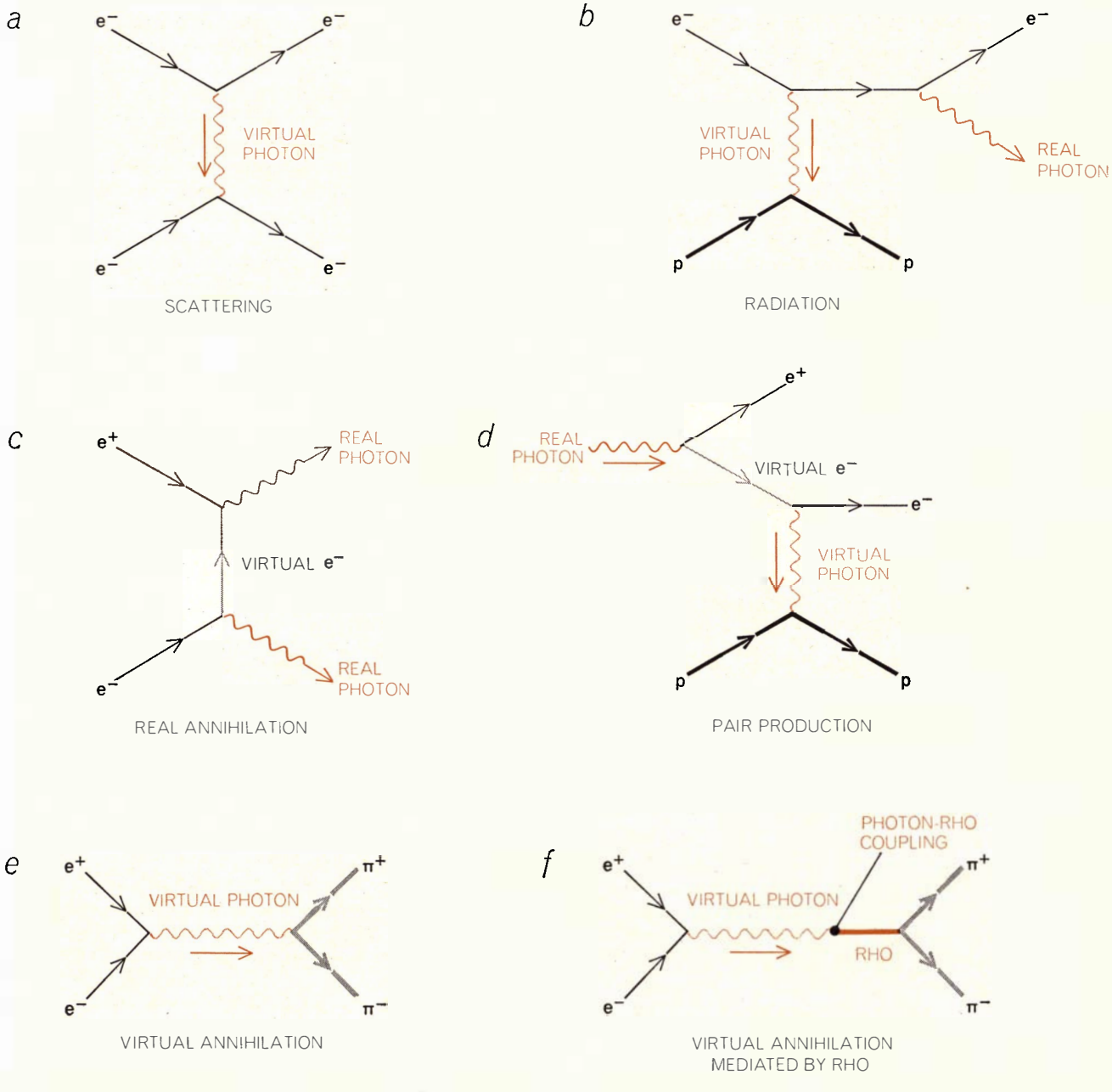
When an electron radiates a real high-energy photon during a collision with a massive atomic nucleus, the energy conveyed to virtual photons is usually small and the nucleus remains almost at rest. The energy of the real photon is then equal simply to the difference in the electron energies before and after collision. These energies can easily be measured by bending the incident and scattered electrons in strong magnetic fields. In this way a beam of "tagged" real photons is produced in which each photon in the beam has a known energy.

Physicists are also interested in processes involving virtual photons and nuclei. If no real photons are radiated during an electron-nucleus collision, the energy of the virtual photon is likewise equal to the difference in the electron energies before and after collision. When some portion of the virtual-photon energy is used to break up the nucleus or to create new particles, such as pions, the process is called inelastic. Inelastic electron scattering is thus analogous to tagging real photons and makes it possible to carry out experiments with undetected virtual photons similar to experiments conducted with real photons.

A second method for producing a beam of real high-energy photons is to scatter visible light 180 degrees with a beam of high-energy electrons, just as a thrown baseball can be "scattered" 180 degrees by a moving bat. This is Compton scattering again, except that the electrons are now moving at nearly the speed of light when the head-on collision occurs. As in Compton's original experiment, the scattered photons behave as material bodies with a definite kinetic energy and momentum that can be calculated from the experimental conditions. And whereas the baseball struck by the bat may have its energy increased tenfold, the energy of the

scattered photon may be increased by a factor of 10 billion. A beam of this type, with a high-power pulsed laser as the primary light source, demonstrates the direct correspondence between photons in widely different regions of the electromagnetic spectrum. Let there be no doubt that radio waves, visible light, X rays and gamma rays are all manifestations of the same basic electromagnetic phenomenon.

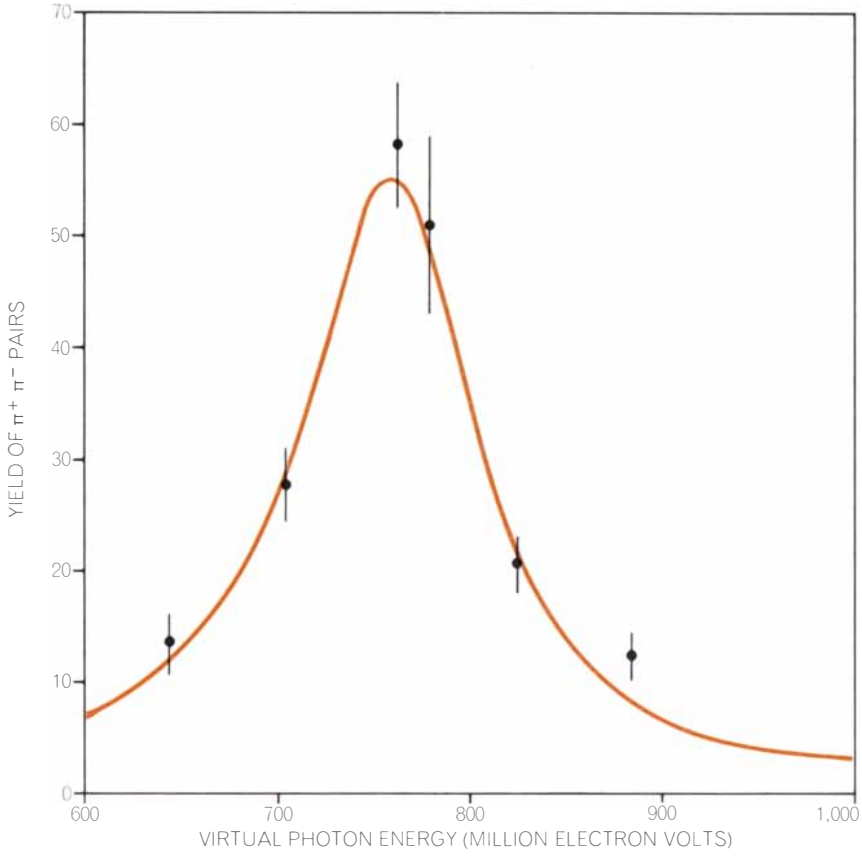
Another method for producing real or virtual photons is particle-antiparticle annihilation, such as occurs when an electron meets its antiparticle, the positron. When a beam of positrons strikes a target, a positron and an electron can interact to produce two real photons: two packets of electromagnetic energy with no mass [*see "c" in illustration below*]. Two photons, rather than just one, are required to balance energy and momentum. (This is the reverse of "pair production," in which a single real photon materializes literally by creating an electron-positron pair. A second photon is also required in pair production—a



PARTICLE INTERACTIONS are commonly represented by Feynman diagrams. Undetectable "virtual" particles are invoked to explain how energy and momentum are transferred during a reaction. According to the uncertainty principle, virtual photons can have a nonzero mass, unlike real photons, which are massless. In electron-electron scattering (*a*) a virtual photon is exchanged. When an electron is scattered by a proton (*b*), a real photon can be emitted and used in experiments requiring photon beams. Such beams can also be created from the photons produced by the annihilation (*c*) of an electron (e^-) and a positron (e^+). The inverse of annihilation (*d*) produces an electron-positron pair. Electron-positron annihilation can also yield a single virtual photon (*e*), which can then produce a pair of pions (π^+, π^-). Materialization of a virtual photon into two pions is greatly enhanced in a two-step process (*f*), in which the virtual photon first materializes into a rho vector meson. The rho decays rapidly into two pions.

virtual photon that balances energy and momentum by transferring these quantities to a nearby nucleus.) Positron-electron annihilation can also yield a single virtual photon that, if it has sufficient energy, can then materialize directly into another particle-antiparticle pair such as a proton-antiproton pair or a pair of positive and negative pions [*see "e" in illustration on preceding page*].

There is an apparent contradiction here. In electron-positron annihilation one real photon is insufficient to balance energy and momentum. How, then, can annihilation into only one virtual photon be allowed? It is possible because the virtual photon, unlike a real photon, is not a final product and cannot be detected; thus it need not be massless. Since the virtual photon exists for an exceedingly short time, there is, according to Werner Heisenberg's uncertainty principle, a correspondingly large uncertainty in its energy. Within this uncertainty, momentum and energy need not balance during the short interaction period, although they are conserved for the initial and final particles that endure long enough to be detected. Saying that momentum and energy do not balance is equivalent to saying that the virtual photon has nonzero mass: the shorter the lifetime, the wider the range of masses a virtual particle can assume. Illegally parked cars seem to obey a similar principle: the owner may risk a $2 fine for several hours but will park in a tow-away zone for only a few minutes. A real photon can also have a nonzero mass during the brief time in which it interacts, but again this momentary discrepancy cannot be directly measured. Hence real photons are always created in pairs in the annihilation of an electron and a positron because real photons can be detected. If, on the other hand, the product of annihilation is a virtual photon, one photon suffices because it becomes something else before its mass discrepancy can be observed.

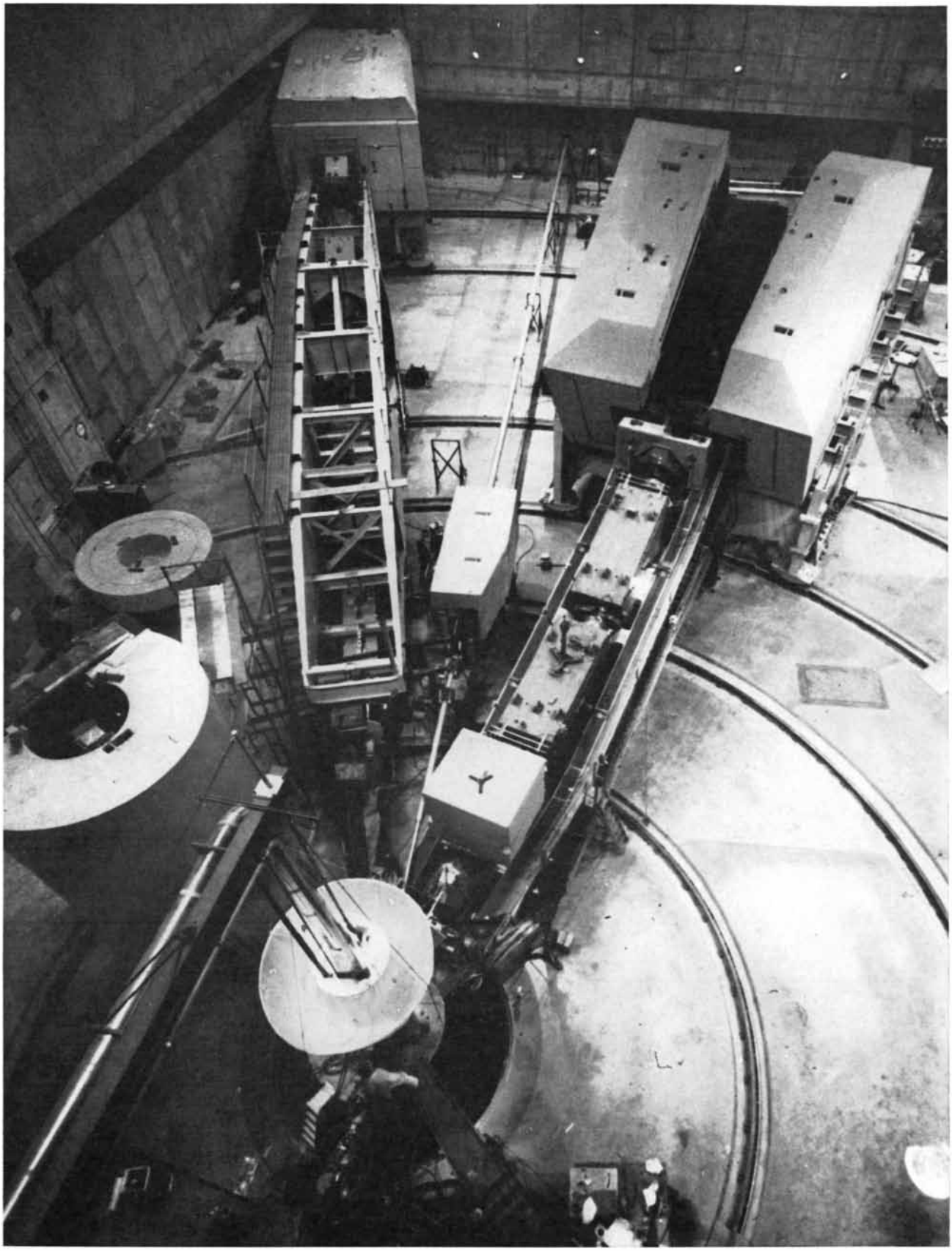


MASS OF THE RHO MESON is measured by finding the energy of the virtual photon that gives the highest yield of $\pi^+\pi^-$ pairs, the particles that are produced when the rho meson decays. The energy of the virtual photon in turn is the sum of the energies of the electron and positron that were annihilated to produce it. The data were obtained at Orsay in France by having a beam of electrons collide with a beam of positrons. The curve shows that the mass of the rho meson is 770 $\pm$ 4 MeV and that it has a "resonance" width of 111 $\pm$ 6 MeV. The greater the resonance width shown by such particles (also known as resonances), the shorter their lifetime. Most rhos decay before traveling more than a few nuclear radii. Similar but much smaller peaks are found at the masses of the omega meson and the phi meson.

The uncertainty principle plays an even more important role in the reasoning that links photons to hadrons. The ability of photons to assume a broad range of masses during high-energy interactions makes it possible for them to take on the masses and other attributes of certain hadrons called vector mesons. The elementary particles can be classified by assigning to each of them a set of labels called quantum numbers. (Examples are numbers for electric charge and spin angular momentum.) Once the quantum numbers of particles are known, the possible interactions between the particles can be predicted. This simplifies an otherwise confusing web of interactions.

The uncharged vector mesons with zero "strangeness" (a particular intrinsic property) have the same quantum numbers as the photon. Even in their free, real state, however, vector mesons are not massless. And since they interact strongly with other mesons and with nucleons, they must be classified as hadrons. Quantum mechanics teaches that if two particles differ only in mass, then whenever they have the same apparent mass, they must be capable of the same interactions. This greatly facilitates the photon-hadron interaction, which can be thought of as occurring in two steps: first the photon materializes into a vector meson in the vicinity of the interacting hadron and then the vector meson interacts strongly with that hadron [*see "f" in illustration on preceding page*]. At high energies this two-step process is more likely than a single direct interaction because the materialization of photons into vector mesons occurs more readily when the energy and the virtual-mass uncertainty are large and because the subsequent meson-hadron interaction is inherently about 200 times more likely to occur than a direct photon-hadron interaction.

A particularly useful model of photon-hadron interactions, called the vector-dominance model, was introduced by J. J. Sakurai in 1960 and was subsequently developed by a number of other workers. The model is based essentially on three relations. The first is that during a photon-hadron interaction the incident photon is equivalent to a combination of vector mesons. The second relation is that any photon surviving such an interaction is a combination of vector mesons. The third relation is a venerable principle, first discovered in ordinary optics,



END STATION "A" AT SLAC is one of several experimental areas grouped at the terminus of the two-mile accelerator. A beam of electrons, positrons or photons strikes a target located at the center of rotation of three independent magnetic spectrometers. The largest (*right*) is 170 feet long and weighs 2,000 tons. It has been used to study the absorption of virtual photons in inelastic electron-proton scattering experiments. All three spectrometers have been used to investigate the photoproduction of rho, omega and phi mesons.

that relates the absorption of light to the diffractive scattering of the same light. This relation between absorption and diffraction is known as the "optical theorem"; the adjective "eikonal" is applied to models in which the theorem is important. The term eikonal, meaning pertaining to a picture or an image, seems particularly appropriate in that the diffracted photons form an image of the target—whether it is a piece of sculpture or an atomic nucleus—and the absorbed photons account for its shadow. Essentially, then, the optical theorem connects image and shadow.

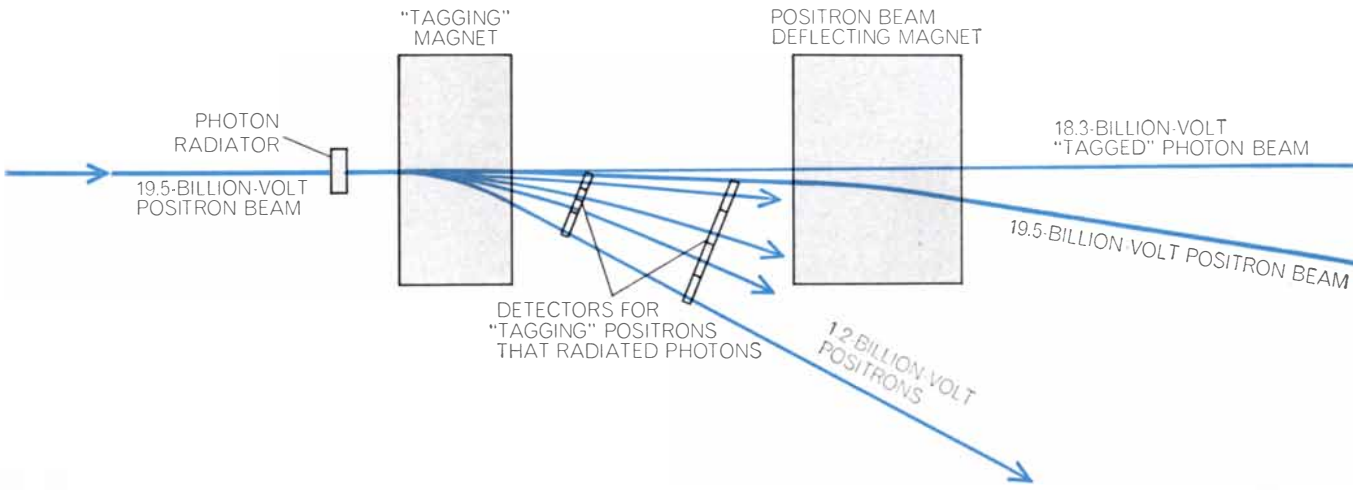
The application of the optical theorem to high-energy photons is nicely illustrated by recent experiments on the elastic scattering (diffraction) of high-energy photons by protons. Here "elastic" means that no additional particles are produced and that both the photon and the proton survive. As the angle of scattering decreases, the probability of scattering approaches the probability predicted by the optical theorem. That prediction in turn is based on the separately measured probability of absorption of photons by protons—events in which neither the photon nor the proton survives. Thus the scattering is related to the absorption; the image is related to the shadow.

Although the vector-dominance model describes a number of processes involving real or virtual photons and vector mesons, it specifically relates the following quantities: first, the constants indicating how strongly the photon is coupled to each vector meson; second, the probability that a photon striking a nucleon (either a proton or a neutron) will yield each of the vector mesons; third, the probability that a vector meson striking a nucleon will interact with it, and fourth, the total probability that a hadron will absorb a photon, in other words, the total probability that a photon striking a nucleon will interact strongly. The model is tested by measuring these probabilities and coupling constants in various experiments.

Three vector mesons with zero strangeness are currently known: the rho, the omega and the phi. The charge of the rho can be positive, negative or zero; the omega and the phi exist only with zero charge. Because photons have a zero charge they couple strongly only with the neutral vector mesons. The observed regularities and gaps in classifications of the elementary particles inspire attempts to explain and predict them. One such attempt is the quark model, devised independently by Murray Gell-Mann and George Zweig of the California Institute of Technology. They suggested that the elementary particles may be combinations of subunits, perhaps purely mathematical, called quarks. The quark model predicts that the photon should behave as if it were 75 percent rho, 8 percent omega and 17 percent phi. This really means that in a large sample of interactions the photon will behave as a rho in 75 percent of the interactions, as an omega in 8 percent and as a phi in 17 percent. Thus the rho is the most important of the vector mesons in mediating photon-hadron interactions; consequently one sometimes speaks of rho dominance instead of vector-meson dominance. Vector mesons can be readily produced by hadrons as well as by photons, so that various production mechanisms can be compared. The vector-meson lifetimes, although much longer than the lifetime of a virtual photon, are quite short. No vector meson persists long enough to form a track in a bubble chamber or a spark chamber, and each must be observed indirectly through the long-lived particles into which it decays. Typically the rho yields two pions, the omega yields three pions and the phi yields two kaons.

The mass of a particular vector meson can be calculated from the measured masses, energies and momenta of its decay products. Experimentally, even when the errors of measurement are negligible, the masses calculated from a sample of decays of the same vector meson do not coincide but instead form a "resonance" distribution, a bell-shaped curve with a certain width around some central mass value [*see illustration on page 98*]. This follows from the uncertainty principle, which requires only that the product of the average lifetime and of the width of the resonant mass distribution for a particle be constant. The broader the mass distribution, the shorter the lifetime; the narrower the distribution, the longer the lifetime.

The first predictions that vector mesons might exist were made by Yoichiro Nambu in 1957 and by William R.

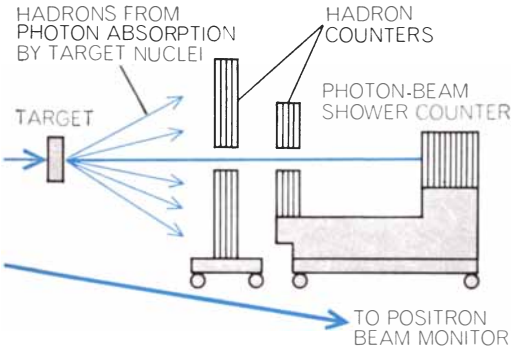


PHOTON ABSORPTION STUDIES were conducted with the experimental arrangement shown here, designed by a team from the University of California at Santa Barbara in collaboration with SLAC. Secondary positrons of 19.5 billion electron volts enter from the left and strike a tungsten target that is the photon radiator. Depending on how much energy the positrons lose in radiating the photons, they are deflected by various amounts by the "tagging" magnet. A stronger magnet, next in line, deflects any positrons that have surrendered no energy. The tagged photons (that is, photons whose energy is known from the energy lost by the positrons that produced them) travel in a straight line until they strike their intended target. If a photon is absorbed, it gives rise to pions that are recorded in the hadron counters; photons that are not absorbed enter the photon-beam shower counter. Pions and similar particles

Frazer and Jose R. Fulco two years later to explain the diffuse structure observed for protons and neutrons, particularly in the electron-nucleon scattering experiments carried out by Robert Hofstadter and his colleagues at Stanford University. The name "vector meson" is derived from the intrinsic angular momentum, or spin, of these particles, which is unity and which transforms from one spatial coordinate system to another as a vector quantity. Another quantum number, the intrinsic parity, is negative. The spin and parity, like all other internal quantum numbers, are the same for photons and vector mesons.

The cleanest method for studying the vector-meson composition of the photon is to arrange for a beam of electrons to collide with a beam of positrons, which isolates the production of vector mesons from other interactions. Such experiments have now been performed at Orsay in France and at Novosibirsk in the U.S.S.R. The annihilation of a positron and an electron frequently yields a single virtual photon, which can materialize immediately into any one of its vector-meson components. As usual, the creation of a particular vector meson is signaled by its mass, as calculated from the energies and momenta of its decay products. These decay products are in practice detected by spark chambers and counters surrounding the region of interaction.

The annihilation of a positron and an electron into a single virtual photon is purely electromagnetic and can be calculated with much confidence and precision. Furthermore, all vector mesons produced by virtual photons decay immediately, so that it is necessary only to measure what fraction of these decays are detected and to calculate the efficiency of the detection apparatus in order to determine the probability that a virtual photon will materialize. The photon composition found in colliding-beam experiments at Orsay is 75 ± 2 percent rho, 10 ± 2 percent omega and 15 ± 2 percent phi, in good agreement with the prediction of the quark model. These experiments also directly measure the central masses and the widths of the resonant mass distributions, which seem to be affected by the presence of target hadrons in other experiments. The colliding-beam experiments performed at Orsay and at Novosibirsk not only are brilliant technical achievements but also have the distinct advantage that no target hadrons are involved.



participating in strong interactions are classified as hadrons. This experiment demonstrates that photons sometimes exhibit hadronic behavior. Photons that produce positron-electron pairs or are scattered by electrons are not detected as "hadronic events"; they pass through holes in hadron counters.

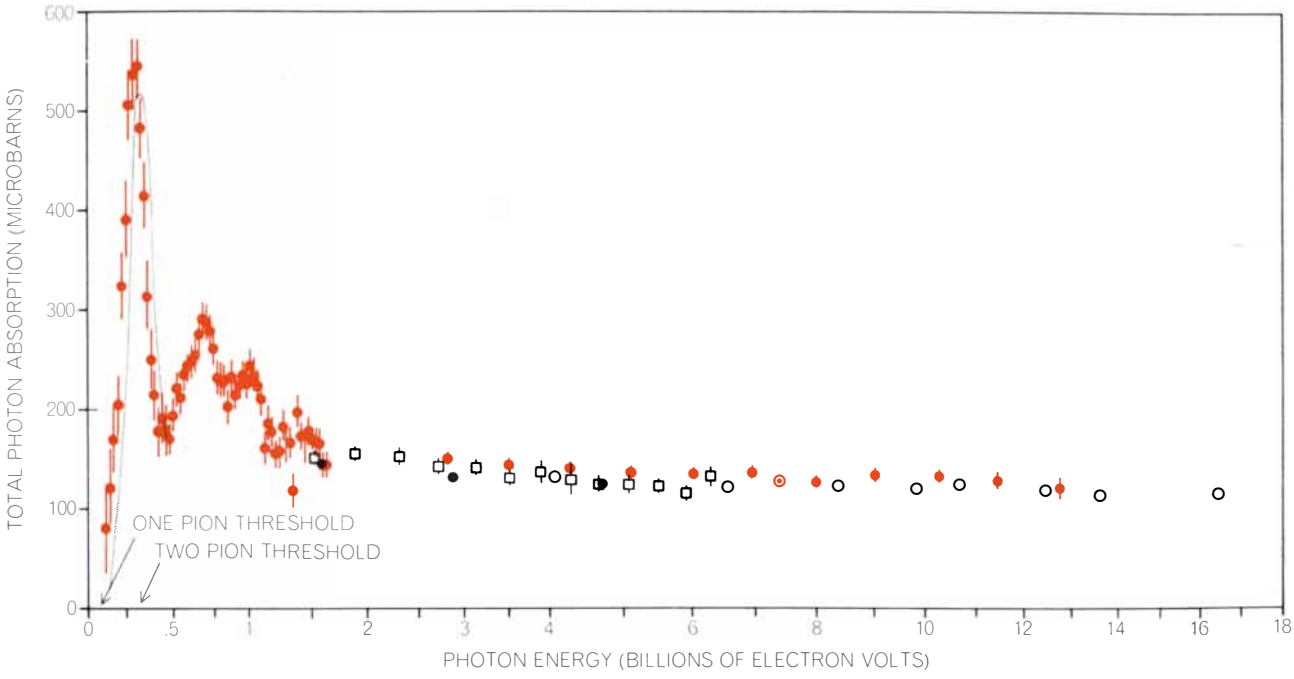
Vector-meson production by real photons has been studied in a wide variety of experiments. All the methods described above for generating beams of high-energy photons have been used successfully: *bremsstrahlung* radiation, positron-electron annihilation and backward Compton scattering of laser photons by high-energy electrons. The detectors have ranged from visual devices (bubble chambers, spark chambers and streamer chambers) to 2,000-ton magnetic spectrometers that detect single recoil particles, such as target protons, from which resonance production can be inferred.

In general the visual devices make possible a detailed analysis of both the production process and the decay process, even for events in which many final particles are involved. Electronic devices, such as magnetic spectrometers with various types of counter, yield less detailed information on particular events and cannot readily handle complex events that produce many particles. They do, however, have an important advantage in that they can be used with highly intense beams to accumulate large numbers of events or to study processes that occur only rarely after many photons have passed through the target. Besides providing data on vector-meson properties and production mechanisms, these experiments are designed to reveal new vector mesons, but none has yet been found.

It is not too difficult to determine which of the three vector mesons has been produced by analyzing their decay products. One would also like to study, however, what happens when a rho or other vector meson subsequently collides with a nucleon. But how can one hope to study rho-nucleon scattering when the path traced by a rho during its brief life is too short to be seen, so short in fact that the rho normally decays into two pions within a few nuclear radii of its point of production? The answer, suggested by Marc H. Ross and Leo Stodolsky, and independently by Sidney D. Drell and James Trefil, is that one studies rho-nucleon scattering *inside* the same nucleus in which the rho was produced. Rho production without subsequent rho-nucleon scattering can be studied with the simplest nucleus: the nucleus of ordinary hydrogen, which consists of a single nucleon. The hydrogen results can then be compared with results from nuclei of increasing diameter, such as those of beryllium, carbon, aluminum, copper, silver and lead, which increasingly show the effects of rho-nucleon scattering and absorption.

In the rho-photoproduction experiments done at Deutsches Elektronen-Synchrotron (DESY) in Hamburg, at Cornell University and at the Stanford Linear Accelerator Center (SLAC) data were taken with more than a dozen different target nuclei over a wide range of photon energies and rho-production angles. Counters and wire spark chambers, used with magnetic spectrometers, made possible high rates of data collection, and hundreds of thousands of rhos were recorded. In qualitative terms, at small angles rho photoproduction per target nucleon increases with the number of nucleons in the target nucleus up to about 64 nucleons (copper) and remains constant or even decreases thereafter. The increase per nucleon indicates that the target nucleons work together coherently to produce rhos more efficiently than the same number of independent nucleons would. This is analogous to the constructive interference observed with light from a coherent source such as a laser. In heavier nuclei more and more rhos are absorbed through rho-nucleon scattering before they escape. This absorption overcomes the advantage of coherent rho production, and thus the rate per target nucleon does not increase for elements with more than 64 nucleons [*see illustration on page 95*].

The experimental results indicate that the characteristic distance for a rho-nucleon interaction to occur in nuclear matter, called the mean free path, is about three fermis, roughly the radius of a carbon nucleus and about four times the radius of a proton. (One fermi is 10^{-13} centimeter.) When target nuclei are

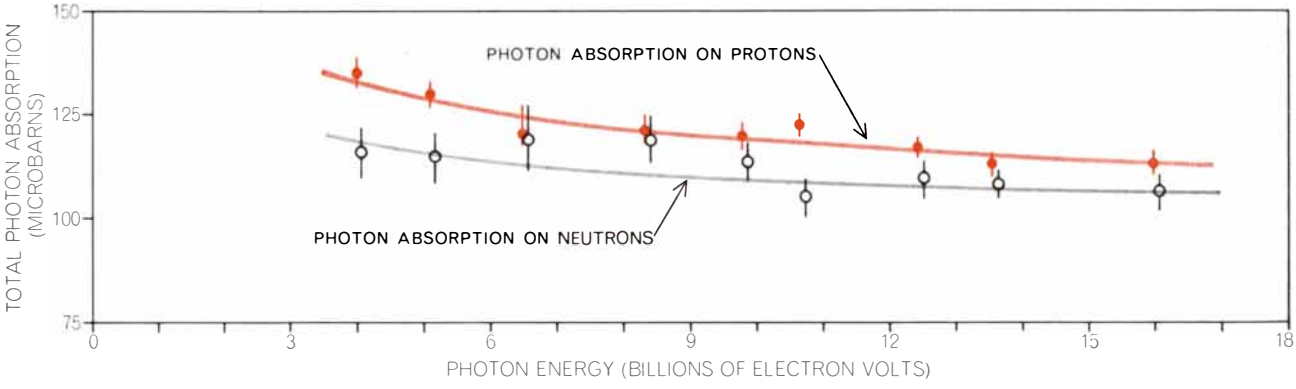


HADRONIC BEHAVIOR OF PHOTONS is depicted by this compilation of data showing the total absorption of photons by protons in hadronic events. The results for virtual photons obtained in the inelastic electron scattering experiments at SLAC (*colored dots*) are in good agreement with a variety of other experiments in which real photons were used. The sharp peaks below two billion electron volts and the smooth continuum above that value closely duplicate the curve obtained when pions (authentic hadrons) are absorbed. Thus data for both real and virtual photons show that photons can exhibit hadronic behavior.

bombarded with beams of conventional hadrons, such as protons and pions, instead of with the "rho beams" generated within individual nuclei, the results are similar. When the rho meson interacts with hadrons, therefore, it acts like a typical hadron.

A major prediction made by the vector-dominance model concerns the probability that a high-energy photon will interact strongly with a nucleon. In such a process all the photon's electromagnetic energy materializes into hadrons. Again each of the techniques we have described for producing high-energy photons of known energy has actually been used in total-absorption experiments. Specifically, photons produced by electron-positron annihilation have been used to bombard nucleons in the one-meter hydrogen bubble chamber at SLAC; similar reactions have been studied in the two-meter bubble chamber using laser photons that had been Compton-scattered by electrons. Tagged-photon beams have been used in both bubble-chamber and counter experiments at DESY, as well as in a counter experiment at SLAC. Finally, inelastic electron scattering has been studied in a spectrometer experiment at SLAC to determine the rates at which virtual photons are absorbed by protons.

These experiments, extending to an energy of 18 billion electron volts, are in

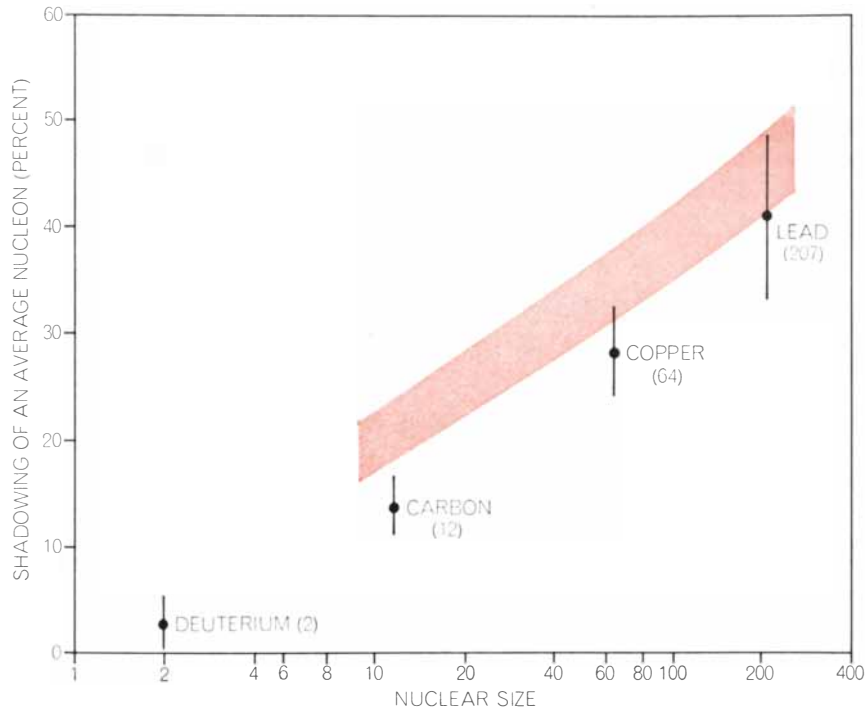


SIMILARITY OF PROTONS AND NEUTRONS is demonstrated by Santa Barbara–SLAC experiments that measured the capacity of the two nucleons to absorb photons. Theory predicts that if photons act like hadrons, the two particles should absorb very-high-energy photons at the same rate. The prediction is borne out if one extrapolates the two curves to infinity. The small divergence of the curves that appears at low photon energies may indicate that mesons other than known vector mesons contribute to the absorption.

excellent agreement [*see top illustration on opposite page*]. They lead to two general conclusions: that protons absorb real photons and virtual photons at the same rate, and that although protons absorb hadrons some 200 times more readily than they absorb photons, the variation in absorption with energy is the same in both cases. In short, whether the photons are real or virtual, their rate of absorption by protons is consistent with their having a hadronic character. This tends to confirm the importance of the two-step process: a photon materializes to a vector meson, which then interacts with a hadron.

A related prediction of the vector-dominance model is that at very high energies the absorption of photons by neutrons and by protons should be the same. This comparison has been made up to 18 billion electron volts at SLAC by members of groups from the University of California at Santa Barbara and SLAC. In this collaborative experiment photons were directed against deuterium, hydrogen and other targets. Where the nucleus of ordinary hydrogen is a single proton, the deuterium nucleus consists of one neutron and one proton. It turns out that the absorption of photons by neutrons is nearly equal to the difference between the absorption by deuterium and the absorption by hydrogen. (A correction of about 5 percent must be made to account for "shadowing" of one nucleon by the other in deuterium.) When the Santa Barbara–SLAC results are extrapolated to infinite photon energy, the absorptions by neutrons and by protons are found to be identical, as expected. At lower energies the absorption by protons is about 10 percent larger than the absorption by neutrons, suggesting that other, nonvector mesons may participate.

It is in the total absorption of high-energy photons by large nuclei, however, that the hadronic character of light has its most startling consequences. Here, more than anywhere else, the nature of light itself depends on whether or not its electromagnetic or its hadronic interaction is being observed. As we have explained, the absolute rate for the absorption of photons by protons or neutrons is roughly 200 times lower than the rates for the absorption of rhos or pions by protons. This is no surprise, since the first step—a photon materializing to a vector meson—of the two-step interaction process is rather improbable. On the other hand, the mean free path for photons, corresponding to this observed absorption rate, is about 700 fermis, more than 200 times longer than the three



SHADOWING OF NUCLEONS BY OTHER NUCLEONS occurs when atomic nuclei of various sizes are bombarded by photons. If photons behaved purely electromagnetically, the shadowing would be zero. The colored band indicates the percent of shadowing per average nucleon predicted by the "vector dominance" model of photon behavior. Santa Barbara experiments at SLAC yield somewhat lower values (*dots*) for deuterium, carbon, copper and lead. Values are an average for photons of between seven billion and 18 billion electron volts.

fermis found for rhos. It is as if 700 layers of nuclear matter were required to absorb light in its original electromagnetic form whereas only three layers of the same material sufficed to stop hadrons. Thus photons pass easily through large nuclei but hadrons barely penetrate the surface, seldom interacting with the nucleons inside.

Since photons pass undiminished through so much nuclear space, one might naïvely expect that they would illuminate all the nucleons in a particular nucleus. The absorption of photons would then be proportional to the total number of nucleons in each nucleus and hence to the volume of the nucleus. In contrast, the absorption of hadrons is known to be proportional to the number of nucleons exposed at the surface and is therefore proportional to the area of the nuclear surface, as if each nucleus were a black disk. Paradoxically the vector-dominance model and the optical theorem lead to the prediction that the ratio of nuclear absorption of photons to absorption of hadrons should depend only on the probability of a photon's materializing as a hadron and should be the same for any nucleus, with both processes varying together in proportion to the surface area and not in proportion to the nuclear volume. Thus we are asked to believe that whereas nuclei are nearly transparent to high-energy photons, surface nucleons can completely obscure interior nucleons.

This is the kind of paradox that nature is continually thrusting at those who have too much confidence in their common sense. For example, in 1819 the French mathematician Siméon Denis Poisson demolished the wave theory of light by *reductio ad absurdum*. He showed with impeccable mathematics that Fresnel's wave theory predicted that the shadow of a disk illuminated by a point source would have a bright spot at its center, which seemed clearly absurd. Alas for M. Poisson: the shadow cast by such a disk, for example by a penny illuminated with a pinhole light source, does have a bright central spot.

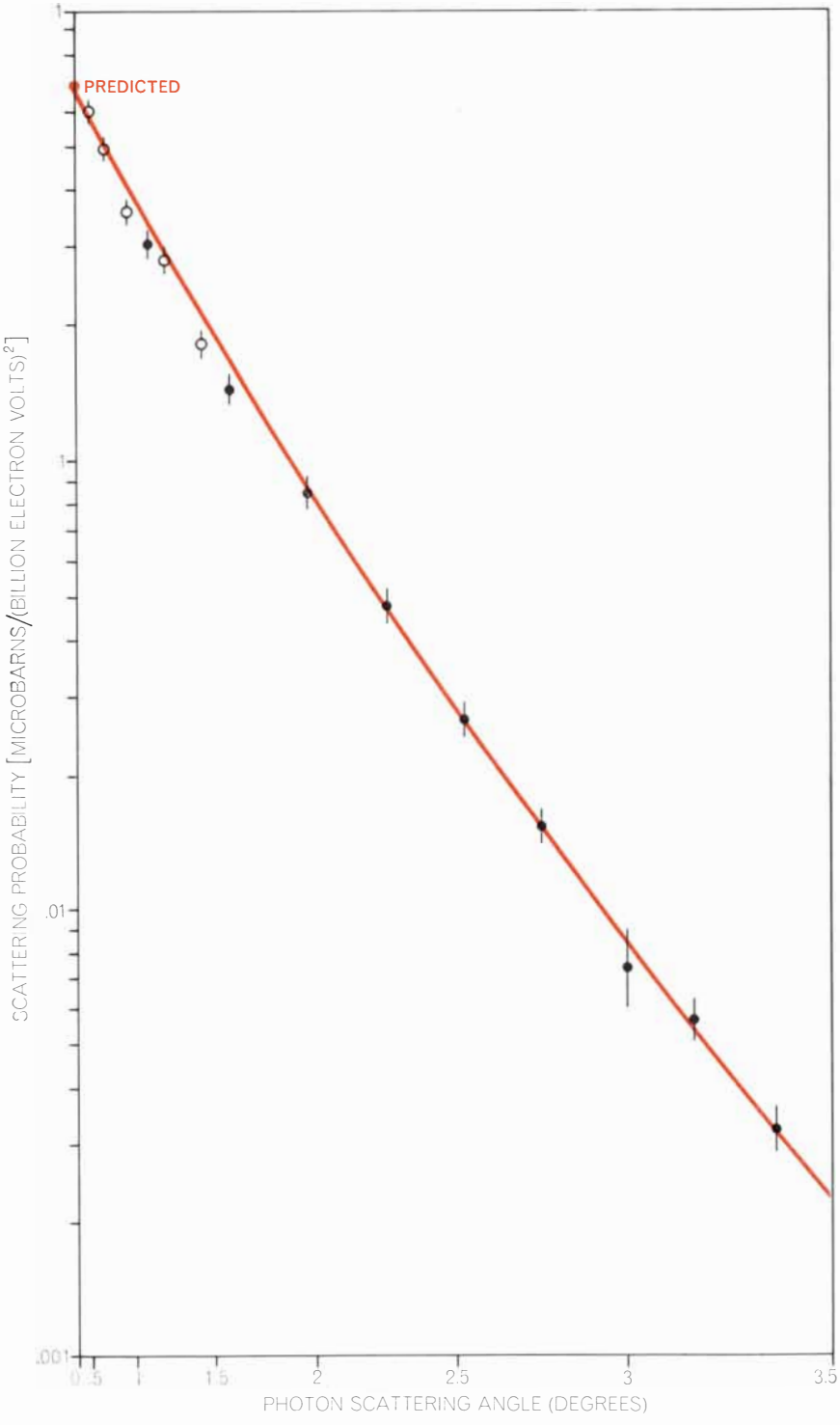
To understand the surface paradox observed with strongly interacting photons, we must once again invoke the uncertainty principle. Specifically, the position on the flight path of the photon where the photon is converted into a vector meson is uncertain by an amount proportional to the energy of the photon. When the uncertainty is large compared with the mean free path for hadrons inside nuclei (as it is for high-energy pho-

tons), then the photon must have been converted into a hadron well before reaching the nuclear surface, in which case absorption and shadowing result. The argument can be turned around. Suppose there were no shadowing. Then the photon-rho conversion must have occurred immediately in front of the absorption point. That, however, destroys the required uncertainty in the conversion point. Therefore there must be shadowing!

Does this mean that a large nucleus absorbs most of the photons striking it? Not at all; nearly all the photons pass through unaffected. The probability that a photon will show hadronic behavior and interact strongly, however, is effectively limited to the nucleons on the surface and does not extend to the nucleons inside. The photon, after penetrating the nuclear surface, is very briefly a "bare" photon stripped of its hadronic character. Suppose a photon passes through two widely separated nuclei in succession. Does the first shadow the second? No; at currently attainable energies the distance, set by the uncertainty principle, over which the photon remains bare is much shorter than the distance between nuclei. Hence shadowing of one atomic nucleus by another rarely occurs.

The prediction of photon shadowing by surface nucleons follows directly from the uncertainty-principle argument and the hadronic character of light, independently of any details of the vector-dominance model. A sufficient test for hadronic character is therefore provided by photon total-absorption studies on nuclei of increasing size. Photons can be considered purely electromagnetic only if the nucleus remains transparent, only if all the nucleons are illuminated and only if the absorption is proportional to the total number of nucleons.

The results obtained in the Santa Barbara–SLAC experiment with targets of ordinary hydrogen, deuterium, carbon, copper and lead show that photon absorption is proportional neither to the total number of nucleons nor to the number of exposed surface nucleons; the data fall somewhere in between [*see illustration on preceding page*]. Some degree of shadowing is certainly present, so that the basic hadronic character of light is confirmed. Qualitatively, therefore, the results are consistent with vector dominance. Quantitatively, however, the amount of shadowing is less than vector dominance, together with the results of other experiments, would predict. Does this mean that there are other vector mesons that couple to the photon besides the rho, the omega and the phi? Perhaps photons become vector mesons less readily than has been thought, or perhaps vector mesons are absorbed less strongly in nuclear matter than current experiments suggest. The answers to these questions, as well as the ultimate fate of the vector-dominance model, may depend on experiments already planned for the 70-billion-electron-volt accelerator now operating at Serpukhov, near Moscow, and for the 500-billion-electron-volt accelerator nearing completion at the National Accelerator Laboratory in Batavia, Ill.



ELASTIC SCATTERING OF PHOTONS ON PROTONS varies with the scattering angle. The curve connects data points obtained at SLAC for photons of 16 billion and 17 billion electron volts. As the scattering angle becomes smaller the probability of scattering approaches that predicted by the "optical theorem." As the scattering angle approaches zero more and more of the incident photons are scattered, until finally the scattering that a photon survives is related by the optical theorem to the probability of its being absorbed.

Why Johnny can't hide.

And doesn't want to anymore.

See Johnny hide. But not anymore.

Now Johnny works at 3M.

Once he thought all big organizations got that way by encouraging anonymity. Not because anyone ever told him that. But because he once had worked for one. Which will, of course, be nameless.

No one in that first organization actually told him the way to do well was to do as little as possible. Sort of stay out of the way.

But then, they never told him otherwise, either.

So he looked around. Made his own appraisal. And concluded there was a theory in operation.

To wit: Never state an opinion or take a risk and you'll never make a mistake.

Now, Johnny's no dope. He took what he thought was the hint.

He got lost in the crowd. Blended in with the background. And never, never attracted attention.

But 3M doesn't believe in crowds.

We believe in people. One at a time. And we'd like to think Johnny knew that the day he started here.

Our whole atmosphere says it. You don't stand out here unless you stand up. And speak up.

We believe in our people. Really believe. And Johnny must have seen it.

We encourage individual ideas and intelligent risk. The more 3M people innovate and accomplish, the more we reward them and the faster they advance.

We will not even hire relatives of our officers and directors. Here at 3M a better job is earned. Not awarded.

And our long and growing customer list is witness to the way 3M has always come up with products in advance of their time.

Like "Scotchgard" Fabric Protector. And "Di-Noc" Decorative Trim on the most popular station wagons. And "Scotch" Brand Hair-Set Tape.

Every new employee comes with a mind of his own. Could it be the way we treat our people has a lot to do with the way they produce ideas?

Johnny thought so. He looked around and liked what he saw and adopted a whole new approach.

Johnny came out of hiding.

And finally came into his own.

People still count here. 3M COMPANY

3 M Co., 3M Center, St. Paul, Minnesota 55101

MATHEMATICAL GAMES

Quickie problems: not hard, but look out for the curves

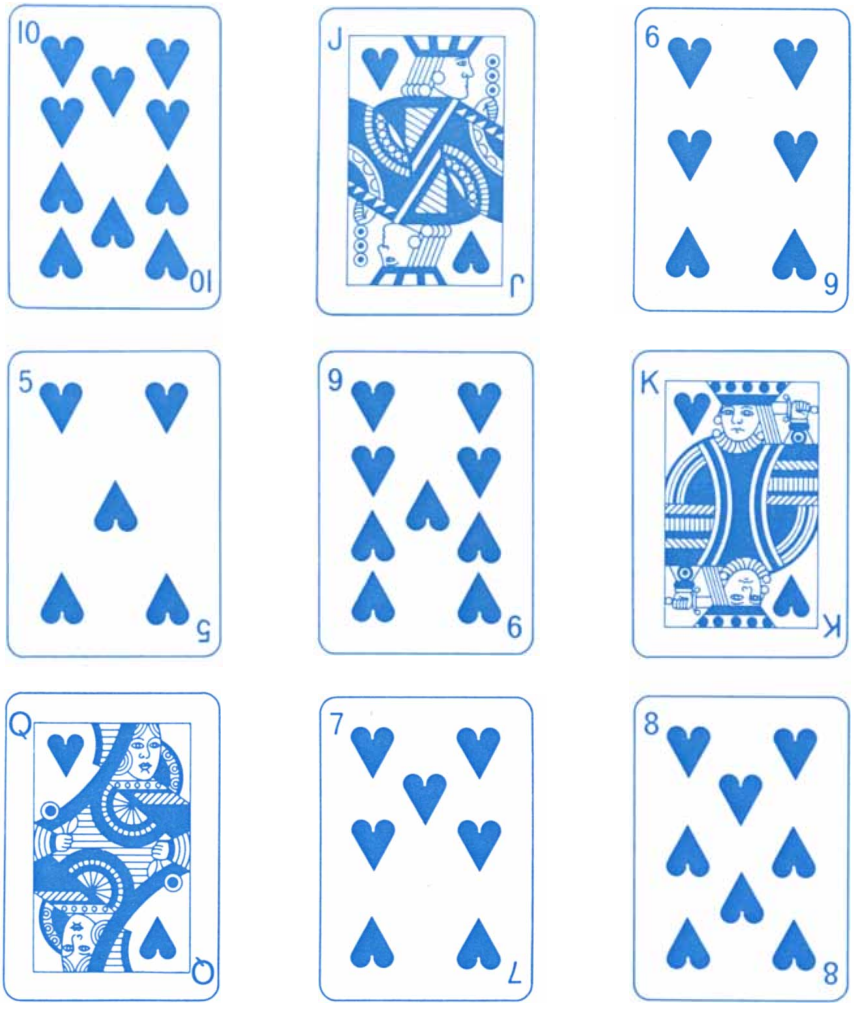
by Martin Gardner

The following problems are all of the "quickie" type. That is, they are easily stated and are not hard to crack if properly approached, but they sometimes have concealed booby traps to catch the unwary. Some of them are joke questions. The answers will be given next month.

1. You want to construct a rigid wire skeleton of a one-inch cube by using 12 one-inch wire segments for the cube's 12 edges. These you intend to solder together at the cube's eight corners.

"Why not cut down the number of soldering points," a friend suggests, "by using one or more longer wires that you can bend at sharp right angles at various corners?"

Adopting your friend's suggestion, what is the smallest number of corners where soldering will be necessary to make the cube's skeleton rigid? (Philip G. Smith, Jr.)



A magic square with nine hearts

2. An intelligent horse learns arithmetic, algebra, geometry and trigonometry but is unable to understand the Cartesian coordinates of analytic geometry. What proverb does this suggest? (Howard W. Eves.)

3. Your king is on a corner cell of a chessboard and your opponent's knight is on the corner cell diagonally opposite. No other pieces are on the board. The knight moves first. For how many moves can you avoid being checked? (From David L. Silverman's marvelous new McGraw-Hill collection of game problems, *Your Move*.)

4. Nine heart cards from an ordinary deck are arranged [*see illustration on this page*] to form a magic square so that each row, column and main diagonal has the largest possible constant sum, 27. (Jacks count 11, queens 12, kings 13.) Drop the requirement that each value must be different. Allowing duplicate values, what is the largest constant sum for an order-3 magic square that can be formed with nine cards taken from a deck?

5. Make a statement about n that is true for, and only true for, all values of n less than one million. (Leo Moser.)

6. Why would a barber in Geneva rather cut the hair of two Frenchmen than of one German?

7. With a black pencil draw a closed, self-intersecting curve of any shape you please. With a red pencil draw a second curve of the same kind on top of the first one, never passing through a previously created intersection. Circle all points where one curve crosses the other [*see top illustration on opposite page*]. Prove that the number of such points is even.

8. Place a familiar mathematical symbol between 2 and 3 to express a number greater than 2 and less than 3.

9. A six-story house (not counting the basement) has stairs of the same length from floor to floor. How many times as high is a climb from the first to the sixth floor as a climb from the first to the third floor?

10. Each of the two equal sides of an isosceles triangle is one unit long. Without using calculus, find the length of the third side that maximizes the triangle's area. (Angela Dunn.)

11. What three positive integers have a sum equal to their product?

12. A string, lying on the floor in the pattern shown in the bottom illustration on the opposite page, is too far away for you to see how it crosses itself at points A, B and C. What is the probability that the string is knotted? (L. H. Longley-Cook.)

13. If *AB*, *BC*, *CD* and *DE* are common English words, what familiar word is DCABE? (David L. Silverman.)

14. *Time*, March 7, 1938, reported that one Samuel Isaac Krieger claimed to have found a counterexample to Fermat's unproved last theorem. Krieger announced that it was $1{,}324^n + 731^n = 1{,}961^n$, where n is a positive integer greater than 2. A reporter on *The New York Times*, said *Time*, easily proved that Krieger was mistaken. How?

15. What familiar English word begins and ends with *und*? (Sid Lorraine.)

16. A man arrives at a random spot several miles from the Pentagon. He looks at the building through binoculars. What is the probability that he will see three of its sides? (F. T. Leahy, Jr., transmitted by Reed Dawson.)

17. Change 11030 to a person by adding two straight line segments.

18. A boy and a girl are sitting on the front steps of their commune.

"I'm a boy," said the one with black hair.

"I'm a girl," said the one with red hair.

If at least one of them is lying, who is which? (Adapted from a problem by Martin Hollis.)

19. A "superqueen" is a chess queen that also moves like a knight. Place four superqueens on a five-by-five board so that no piece attacks another. If you solve this, try arranging 10 superqueens on a 10-by-10 board so that no piece attacks another. Both solutions are unique if rotations and reflections are ignored. (Hilario Fernandez Long.)

20.

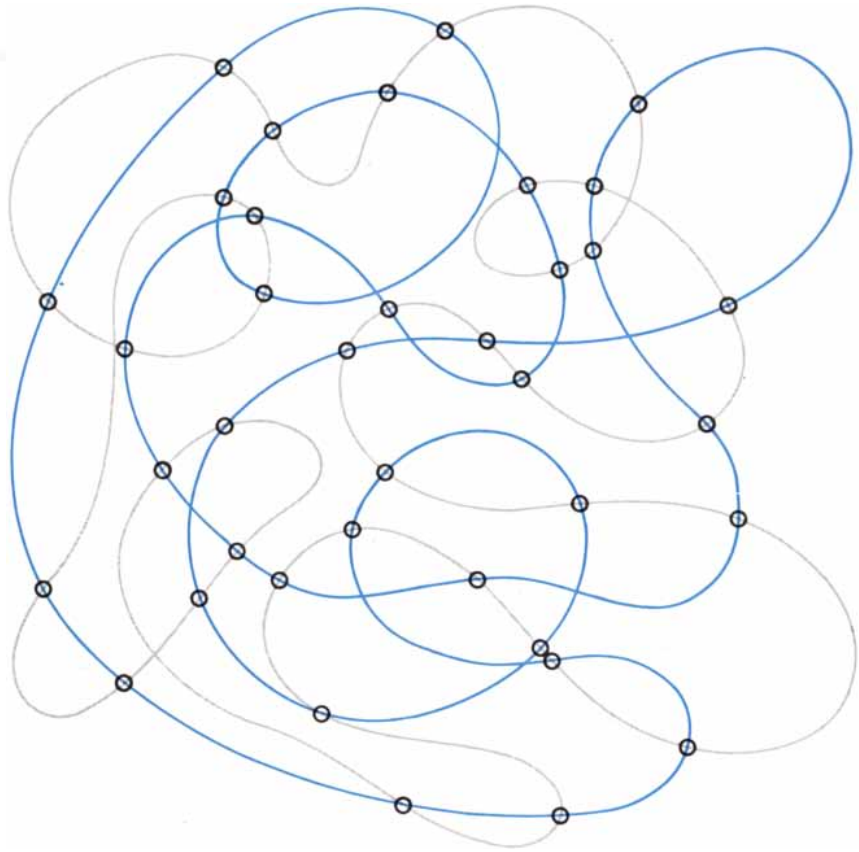
$$\begin{array}{r} ABCD \\ DCBA \\ \bullet\ \bullet\ \bullet\ \bullet \\ \hline 12300 \end{array}$$

ABCD are four consecutive digits in increasing order. *DCBA* are the same four in decreasing order. The four dots represent the same four digits in an unknown order. If the sum is 12,300, what number is represented by the four dots? (W. T. Williams and G. H. Savage.)

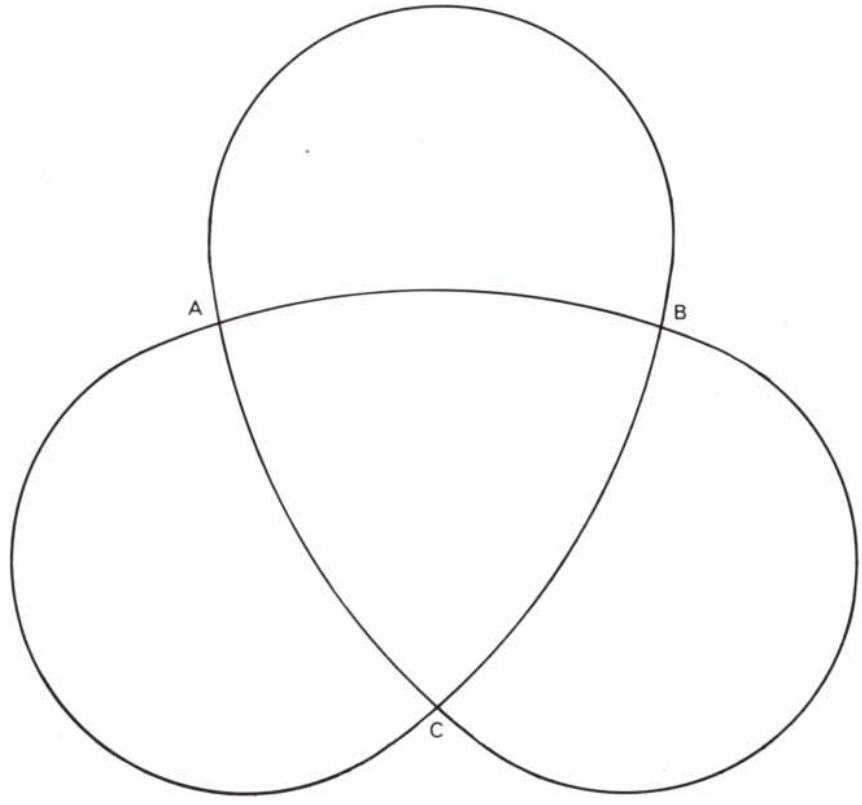
21. A "primeval snake" is formed by writing the positive integers consecutively along a snaky path [*see top illustration on next page*]. If continued upward to infinity, every prime number (color) will fall on the same diagonal line. Explain.

22. Find two positive integers, x and y, such that the product of their greatest common divisor and their lowest common multiple is xy.

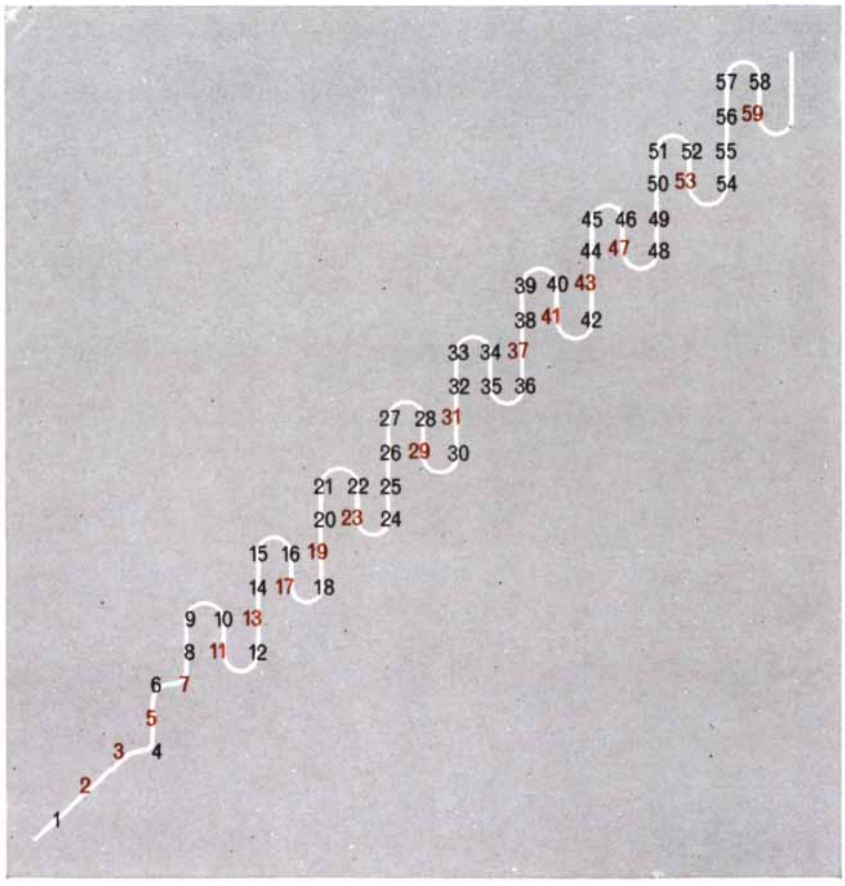
23. "Feemster owns more than a thousand books," said Albert.



A topological theorem



Is the string probably knotted?



The primeval snake

"He does not," said George. "He owns fewer than that."

"Surely he owns at least one book," said Henrietta.

If only one statement is true, how many books does Feemster own?

24. In this country a date such as July 4, 1971, is often written 7/4/71, but in other countries the month is given second and the same date is written 4/7/71. If you do not know which system is being used, how many dates in a year are ambiguous in this two-slash notation? (David L. Silverman.)

25. Why are manhole covers circular instead of square?

26. How many different 10-digit numbers, such as 7,829,034,651, can be written by using all 10 digits?

27. Many years ago, on a sultry July night in Omaha, it was raining heavily at midnight. Is it possible that 72 hours later the weather in Omaha was sunny?

28. What well-known quotation is expressed by this statement in symbolic logic?

$$2B \vee \sim 2B = ?$$

29. Regular hexagons are inscribed in and circumscribed outside a circle [*see illustration at left at bottom of this page*]. If the smaller hexagon has an area of three square units, what is the area of the larger hexagon? (Charles W. Trigg.)

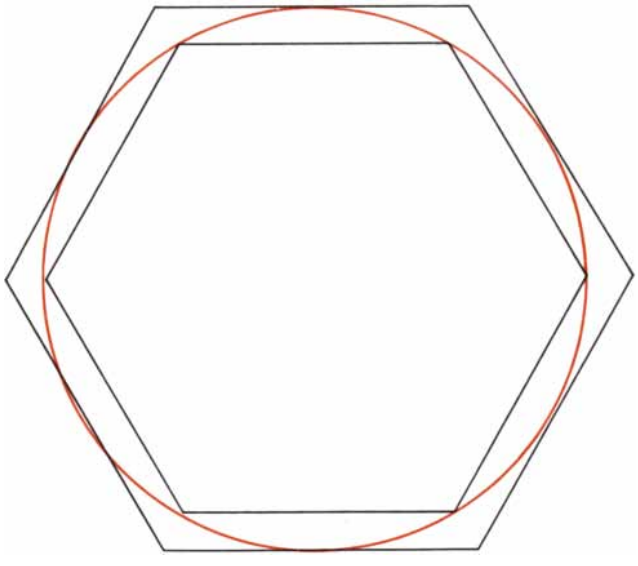
30. "I was n years old in the year n^2," said Smith in 1971. When was he born? (S. J. Farlon.)

31. If you think of any base for a number system, I can immediately write down the base without asking you a question. How can I do this?

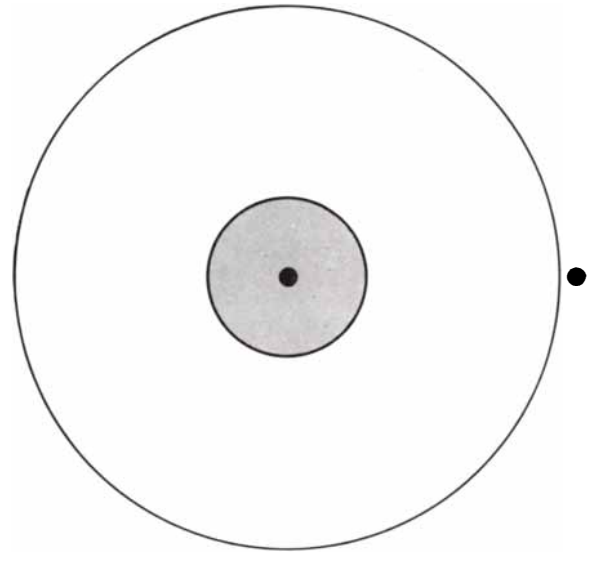
32. What was the name of the Secretary General of the United Nations 35 years ago?

33. You have one red cube and a supply of white cubes all the same size as the red one. What is the largest number of white cubes that can be placed so that they all abut the red cube, that is, a positive-area portion of a face of each white cube is pressed flat against a positive-area portion of a face of the red cube. Touching at corner points or along edges does not count.

34. What four consecutive letters of the alphabet can be arranged to spell a familiar four-letter word? (Murray R. Pearce.)



Hexagon problem



Lake, island and trees

35. The illustration at right at the bottom of the opposite page is a diagram of a deep circular lake, 300 yards in diameter, with a small island at the center. The two black spots are trees. A man who cannot swim has a rope a few yards longer than 300 yards. How does he use it as a means of getting to the island?

36. A boy, a girl and a dog are at the same spot on a straight road. The boy and the girl walk forward—the boy at four miles per hour, the girl at three miles per hour. As they proceed the dog trots back and forth between them at 10 miles per hour. Assume that each reversal of its direction is instantaneous. An hour later, where is the dog and which way is it facing? (A. K. Austin.)

The only problem last month was to guess why Arthur C. Clarke chose HAL as the name of his talking computer in *2001: A Space Odyssey*. If each letter in HAL is shifted forward one letter in the alphabet, the result is IBM. This was called to my attention by Robert T. Wainright. I had assumed that it was intentional but Clarke has since told me it was quite coincidental; indeed, he was surprised when it was pointed out!

Several readers have called my attention to a recent article, "Multi-dimensional Map-folding," by W. F. Lunnon, in *The Computer Journal* (published by the British Computer Society in London), Vol. 14, No. 1, February, 1971, pages 75–80. Using an ingenious diagram based on two perpendicular slices through the center of the final packet, Lunnon was able to write a simple backtrack program for x-by-y maps, extend the problem to higher dimensions and discover several remarkable theorems. For example, the edges of one cross section always diagram x linear maps of y cells each, and the edges of the other cross section diagram y linear maps of x cells each. His program found 1,368 folds for the three-by-three map (152 folds for each cell on top).

The two-by-three, two-by-four, two-by-five, two-by-six and three-by-four maps have respectively 60, 320, 1,980, 10,512 and 15,552 folds. The order-4 square has 300,608, the order-5 has 186,086,600. In all cases the number of folds is the same for each cell on top and the cyclic law mentioned last month necessarily holds for all rectangular maps. The order-2 cube, folded through the fourth dimension, has 96 folds; the order-3 cube has 85,109,616. Many other results are tabulated by Lunnon, but a nonrecursive formula for even planar maps remains elusive.

THE AMATEUR SCIENTIST

How to ensure a good hologram and how to build an unusual kind of barometer

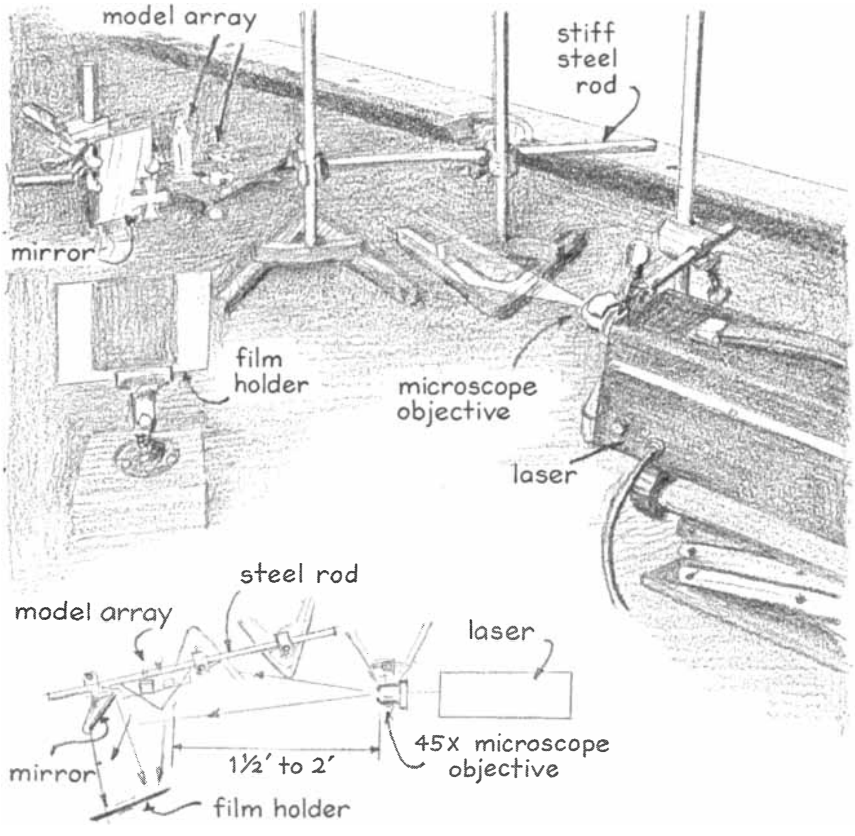
Conducted by C. L. Stong

Many amateurs have built a helium-neon laser. Relatively few have succeeded in using the laser to make a good hologram: the type of photograph in which the coherent light of the laser is used to create an interference pattern that, when it is illuminated with coherent light, yields a three-dimensional image of an object. Most of the failures can be traced to vibrations that produce minute changes in the distances between the various parts of the holographic optical system, particularly the distance between the object to be photographed and the film holder.

In the making of a hologram the coherent light of a laser is projected along two paths to the sheet of film. The rays traveling along one path constitute the reference beam; they fall on a flat mirror, from which they are reflected to the film. The rays traveling along the second path constitute the object beam; they fall on the object to be photographed, from which they too are reflected to the film. The rays of the reference beam are virtually equal in length. The rays of the object beam are reflected by the various surfaces of the object and are therefore of different lengths. As a result the light waves in the two beams are out of step and interfere with one another when they combine at the surface of the film. At some points on the film the crests of waves from the two beams coincide. The light is brightest at these points, and the film gets maximum exposure. At other points the crest of a wave in one beam coincides with the trough of a wave in the other beam. The light is dimmest at these points and the film gets minimum exposure. The resulting photographic pattern is essentially a record of the local differences in the length of the rays of the reference beam and the object beam. If the differences in path length so recorded arise solely from the shape of the object, the hologram yields a remarkably realistic image of the object. In practice, however, some portion of the difference invariably arises from the vibration of the apparatus. The quality of the hologram is degraded in proportion to the intensity of the vibration.

Professional workers who make holograms resort to elaborate schemes for isolating their apparatus from vibrations, such as mounting the parts on a multiton slab of granite, on a base that floats on air or on a platform equipped with a servo system that cancels the vibrations by means of feedback. People who do not have access to such apparatus can still make reasonably good holograms by means of a simple technique recently developed by LeRoy D. Dickson of the IBM General Systems Division in Rochester, Minn.

"In 1967," writes Dickson, "I found that good three-dimensional holograms could be made by minimizing vibrations that alter the distance between the object and the mirror. Vibrations between the film and the laser, and between the laser and the object, have substantially less effect on the quality of the hologram, although they should be minimized. Essentially what I did was tie the two reflecting elements of the system together with a steel rod so that they moved as a unit. I attached the mirror to one end of the rod by an apparatus clamp and cemented the object to be photographed to a base of sheet metal that was in turn cemented to the rod a few inches beyond



LeRoy D. Dickson's setup for making a hologram

the mirror. The remaining portion of the rod was clamped to a pair of apparatus stands for support [*see illustration on opposite page*]. This improvised optical bench was placed on a solid base that also supported the laser and film holder as separate units. A 45-power microscope objective lens was inserted in the beam of the one-milliwatt helium-neon laser to spread the narrow beam of the laser into diverging rays. The rays were directed toward the optical bench to simultaneously illuminate both the mirror and the object. The resulting holograms were sharp, bright and displayed good parallax.

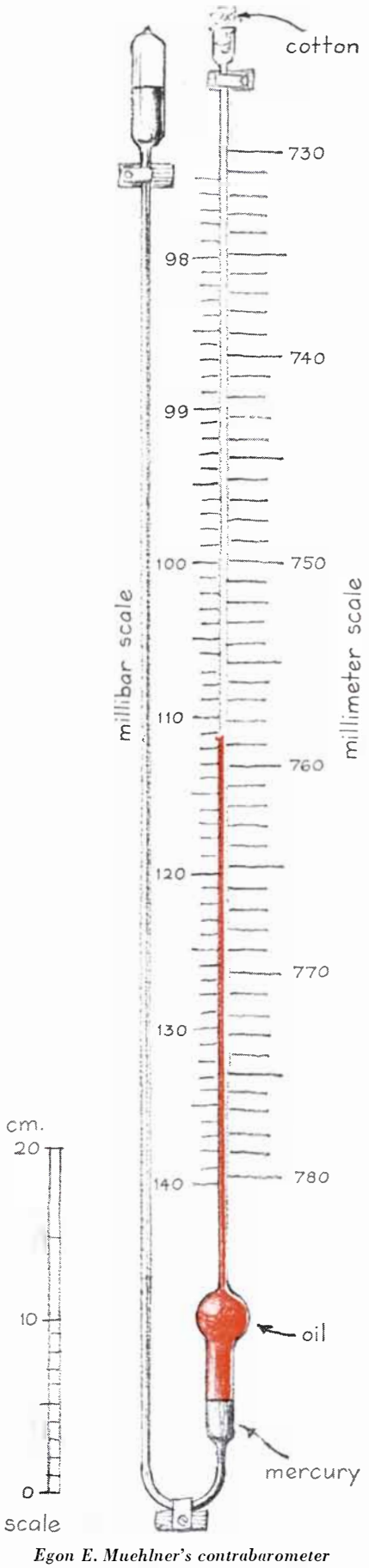
"Two students at the Mayo High School in Rochester, Brandon Dallman and Tom Smyrk, experimented with the system and achieved excellent results on their first try. For objects they used miniature models of buildings and automobiles and mounted them as close to the mirror as possible. Indeed, one object, a Maltese cross, was mounted directly on the mirror. The one-milliwatt helium-neon laser, the microscope lens, the film holder and the rod assembly were placed on a pool table that rested on the concrete floor of a basement. I recommended the use of four-by-five-inch Kodak spectroscopic plates (Type 649-F).

"Light reflected to the film by the mirror should be approximately four times brighter than the light reflected to the film by the object. To achieve this ratio I use rough-surfaced objects of comparatively high reflectivity, such as unpainted models. The relative brightness of light that reaches the film can be judged by inserting a piece of white cardboard in the film holder and, with the room lights off, alternately blocking the rays from the mirror and from the object with a piece of black cardboard. Light reflected by the mirror can be reduced by moving the mirror either toward the edge of the beam or farther away from the laser. The best illumination ratio is 4:1, but anything from 3:1 to 10:1 will work. The angular position of the mirror should be adjusted to illuminate the film uniformly.

"To make an exposure, turn on the laser and let it warm up for 15 minutes. Turn off the room lights, block the laser beam with a piece of black cardboard, open the film holder and after a few seconds unblock the laser beam. The laser must of course be enclosed by a housing that blocks all light except the beam. Do not touch the apparatus with anything that might cause the parts to vibrate, including the cardboard used for blocking the beam. Try several exposures of 30 seconds, one minute and two minutes. Develop the films in Kodak D-19 developer for six minutes at 68 degrees Fahrenheit. Rinse, fix and dry according to Kodak instructions. Good holograms appear as transparencies that range from light gray to somewhat darker gray when they are examined in white light. The image can be reconstructed by looking through the hologram toward diverging rays from the laser. (It would be dangerous to look down the beam as it comes from the laser, but the diverging lens greatly reduces the intensity of the light.) Rotate the film around its vertical axis until the brightest image appears. The image can also be seen by looking toward a point source of ordinary light, but the quality will be degraded somewhat. A satisfactory source can be improvised with a 35-millimeter slide projector. Insert a piece of red gelatin in the space normally occupied by the slide and insert a pinhole mask over the front of the projection lens. The mask can be a piece of cardboard perforated by a hole about three millimeters in diameter. The perforation should be located at the center of the projection lens."

The aneroid barometer, which displays changes in atmospheric pressure with a rotating hand actuated by a flexible closed chamber, can be found in millions of U.S. homes. Among household scientific instruments it is outnumbered only by the clock and the thermometer. Egon E. Muehlner, an engineer of Santa Ana, Calif., observes that barometers of this type are neither as reliable, as accurate nor as handsome as the contrabarometer, in which the changes in pressure are displayed by the level of fluid in a glass tube. Muehlner recently built a contrabarometer, one of the few working models in this hemisphere. He discusses its virtues and explains how to make one:

"The contrabarometer," writes Muehlner, "is basically a *U*-shaped glass tube that contains two fluids. One arm of the *U* is closed at the top and holds a column of mercury that extends upward a short distance into the other arm, which is open to the atmosphere. The open arm, which is surmounted by a small bulb and a capillary tube, holds a column of brightly colored oil that rests on the mercury. Movement of the mercury in response to variations of barometric pressure causes a relatively large excursion of oil in the capillary [*see illustration at right*]. When the pressure rises, the meniscus of the oil falls, and vice versa. This action accounts for the name of the instrument. Contrabarometers adorned

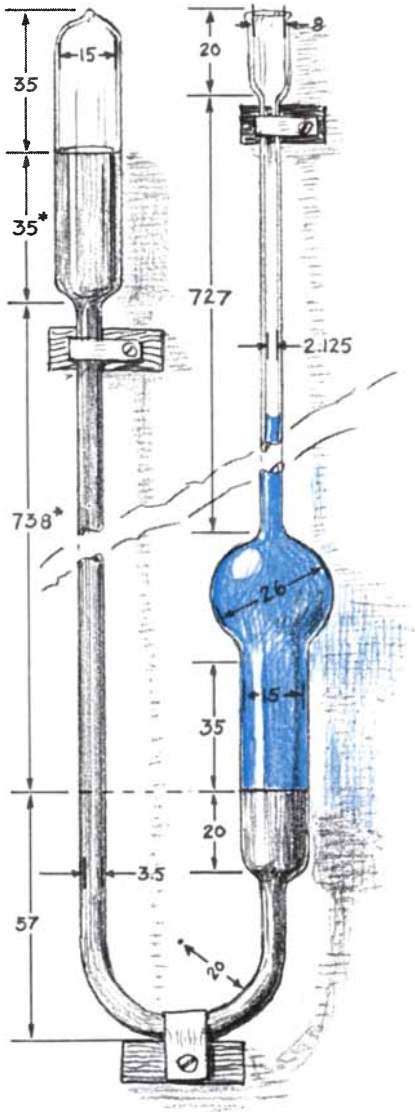


Egon E. Muehlner's contrabarometer

many living rooms here and abroad during the 19th century, providing the head of the house with a scientific foundation for his statement: 'It's going to rain.'

"The contrabarometer is significantly more reliable and accurate than the aneroid type, which tends to stick and to suffer from all the other ailments that beset mechanical contrivances. Moreover, when the geometry of the contrabarometer is properly proportioned, the instrument requires no temperature correction, as the conventional mercury barometer does.

"The construction calls for some glass-blowing that, with a little practice, can be achieved by most beginners. Shops that specialize in the repair of neon signs also accept work of this kind. The construction begins with the acquisition of the glass tubing specified in the accompanying drawing [*bottom left*]. The sizes are available from most distributors of scientific supplies. The substitution of other sizes will alter the scale factor of the instrument and the relative volumes of fluid required for temperature compensation. Appropriate dimensions that differ from those I specify can be calculated by means of the accompanying formulas [*top of opposite page*]. For the indicating fluid I use Meriam Red Oil D-2673, which has a specific gravity of .827 at four degrees Celsius and a vapor pressure of one torr at 25 degrees C. The oil is brilliant red. It can be obtained from the Charles Meriam Company, Inc., 5017 Telegraph Road, Los Angeles, Calif. 90022. Other oils can be substituted by taking their density and thermal expansion into account in making the calculations. Changes in the total volume of fluid should be accommodated by altering the size of the spherical bulb, which in my design has an inner diameter of 26 millimeters.



Dimensions of the contrabarometer

"My instrument is designed to be used at sea level. It has a scale factor of 10: the oil rises or falls one centimeter when barometric pressure changes one millimeter. The reference scale is graduated from 775 to 730 millimeters of mercury, a pressure range of 45 millimeters. At locations of higher elevation the instrument can be modified by reducing the height of the mercury column as measured between the upper and the lower meniscus of the metal.

"Before the instrument is filled the glassware should be firmly mounted on an attractive base. Much of the satisfaction of having a good barometer comes from its beauty. My base was made of 3/4-inch walnut. I gave it four coats of marine varnish and hand-rubbed it first with pumice and water and then with rottenstone and linseed oil. A friend who made a contrabarometer carved his base by hand from a block of oak. The carving features embellishments and ornaments in the style of the 1880's.

"The glassware is mounted on the base with three clamps, two at the top and one at the bend of the *U*. The clamps are preferably made of polished brass. A short length of slit rubber tubing is placed around the glass at the clamping points. The glass must not be clamped too tightly to the base because it might break if the wood warped. The oil capillary should clear the wood by at least 1.5 millimeters to leave space for inserting the scale.

"The barometer can be filled by either of two procedures. The easier one calls for a vacuum pump. If this technique is employed, the bulb at the top of the mercury column must initially include a short tube about seven millimeters in diameter. This tube is softened in a flame and constricted in a narrow zone to a diameter of about four millimeters just above the point where it will be sealed into the bulb. Air is pumped from the bulb through the tube, creating a vacuum that draws mercury into the system. The instrument requires approximately 150 grams of triple-distilled uncontaminated mercury. Impure metal will stick to the glass. Mercury of the required quality can be obtained from a distributor of chemicals or perhaps a dentist. With the instrument in the upright position connect the vacuum pump to the exhaust tube and start the pump. Place a dishpan of water under the barometer to catch metal that may be spilled accidentally. Pour mercury into the open arm of the *U* with a small funnel. Continue adding mercury carefully until the meniscus rises to a height of 29 millimeters in the bulb at the top of the mercury column. The metal will then stand at a height of 26 millimeters in the 15-millimeter tubing of the other column. The vacuum pump should be capable of reducing the pressure to less than .05 torr. Let the pump operate for 15 minutes after the barometer has been filled. Then seal off the tube through which the air was exhausted by heating the constriction to the temperature where the glass softens.

"The barometer can also be filled without a vacuum pump. Procure from a distributor of scientific supplies a 10-foot length of size 'A' Intramedic polyethylene tubing, a product of the Clay-Adams Company. This material is manufactured for surgical procedures. It has an outside diameter of .038 inch and a bore of .023 inch, and it fits a No. 23 gauge hypodermic needle. Lubricate the exterior surface of the plastic tubing with the Teflon suspension Slip Spray, a product of the Du Pont Company. Thread the tubing through the barometer until the end of the plastic capillary rests inside the bulb at the top of the mercury column. The bulb should be sealed. If difficulty is encountered in inserting the plastic capillary, use a leader of fine music wire before sealing the bulb. Rest the instrument on its edge with the sealed bulb lowermost. Elevate the opposite end about 30 centimeters. With a small improvised funnel pour mercury through the plastic capillary and into the bulb. Avoid trapping an air bubble in the bulb at the point where it joins the glass tubing by elevating the opposite end still

more as necessary. Lower the elevated end slowly as the tubing fills. Add mercury until the metal half-fills the *U* bend. Remove the plastic capillary. Rotate the instrument to the upright position. Mercury will rise in the open tube, creating a vacuum in the bulb at the closed end. With an eyedropper add mercury to the specified level. Avoid spilling mercury. The fumes can be toxic.

"Oil is added on top of the mercury. Through the glass capillary insert an appropriate length of plastic capillary tubing and push the end into the glass sphere. Oil can be siphoned into the system, but it flows slowly. The job can be speeded up by forcing oil through the plastic capillary with a hypodermic syringe. Attempts to pour oil directly into the glass capillary fail. Bubbles of air become trapped in the narrow bore. Fill the capillary to the point where the indication of a specific barometric pressure is desired. If this is to be accomplished, the current barometric pressure must be known. I learned the current pressure by telephoning the local airport. Airports conventionally report barometric pressure as the 'altitude setting' in inches of mercury. I placed a bit of adhesive tape on the glass capillary at the point where the upper edge of the tape coincided with the meniscus of the oil. The tape thus marked the current barometric pressure. Inevitably a few air bubbles will become trapped in the oil capillary. To get them out rotate the barometer slowly in the counterclockwise direction. This motion gradually lowers the bulb at the closed end of the instrument. Metal drains into the bulb and simultaneously lowers the level of oil in the capillary. Eventually all the oil will drain into the sphere, where the bubbles will break. When they have broken, restore the instrument to the upright position.

"One must now determine the scale factor of the instrument: the distance in millimeters that the oil meniscus is displaced by a change in barometric pressure of one millimeter. By the application of the formula I designed my instrument for a scale factor of 10. Glass tubing, however, is not manufactured to close tolerances. My instrument turned out to have a scale factor of 10.7, as measured by a water manometer. To calibrate the instrument I connected one end of a *U*-shaped glass tube partly filled with water to the open end of the barometer. The other end of the water manometer was fitted with a short length of flexible hose and a pinch clamp. I blew into this hose (or sucked it) and clamped it. I then measured the resulting displacement of the water. The pressure thus applied to the barometer, measured in millimeters of mercury, is equal to the displacement of the water, also measured in millimeters, divided by 13.55. A series of calibration points was established and marked on the glass capillary with bits of adhesive tape by the similar application of other pressures to the water manometer.

$$\text{Scale factor} = \frac{1}{\left(\frac{d}{D}\right)^2\left(2 - \frac{S_o}{S_m}\right) + \frac{S_o}{S_m}}$$

in which:

D = diameter of mercury meniscus
d = diameter of oil meniscus
S_o = specific gravity of the oil
S_m = specific gravity of the mercury

$$V_o = \frac{0.785\ HD^2 - V_m}{\frac{E_o}{E_m}\left(2 - \frac{S_o}{S_m}\right)}$$

in which:

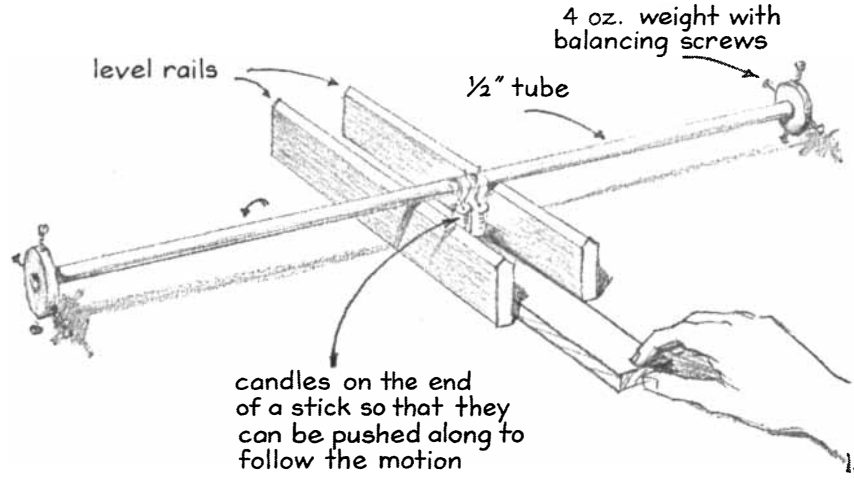
V_o = total volume of oil, cm.^{3}
V_m = total volume of mercury, cm.^{3}
E_o = expansion coefficient of the oil *
E_m = expansion coefficient of mercury**

* = 67.5×10^{-5}/degrees Celsius for D-2673 oil
** = 18.1×10^{-5}/degrees Celsius for mercury

Design equations for the contrabarometer

"I made the permanent scale of my instrument from a strip of white Formica, the plastic sheeting used for covering kitchen counters. The principal graduations were applied with black adhesive tape 1/16 inch wide and the subdivisions with tape 1/32 inch wide. Numerals were applied with transfer sheets. These materials are available from dealers in art and drafting supplies. The scale was calibrated both in millimeters of mercury and in millibars, after which it was positioned according to the tape tabs and fastened to the base with brads. One inch is equal to 25.4 millimeters of mercury, a millimeter of mercury (one torr) is equal to 1.333 millibars and a millibar is equal to a pressure of 100 newtons per square meter. I equipped the instrument with a sliding fiducial point for observing relative changes in pressure with respect to time and also mounted a thermometer in the center of the base.

"One word of caution: If you make



J. M. Clarke's ultrasimple heat engine

the contrabarometer, be careful when you move it from place to place. If it is not held upright, fluid can easily spill out of it and even spoil the vacuum at the top of the mercury column. If anyone has any question about the making of the instrument, my address is 13882 Dall Lane, Santa Ana, Calif. 92705."

Thermal engines of unusual design have been described in these columns from time to time. J. M. Clarke, an engineer of Basingstoke, England, now challenges anyone to devise a simpler heat engine than one he recently built. Clarke's engine has only one moving part: a metal tube with weights at the ends that can be adjusted radially for balancing the mass with respect to its long axis. The tube rolls on a pair of level rails and is powered by a pair of candle flames [*see bottom illustration on preceding page*].

"The weights," writes Clarke, "exert a bending moment at the center of the tube. Four nuts, spaced at equal radial distances around the weights, can be adjusted to compensate for slightly bent or out-of-round tubes. The tube is preferably of thin-wall aluminum at least half an inch in diameter and 24 inches long. Heat causes the tube to expand in the center and bow toward the opposite side. Gravity causes the bowed tube to roll away from the source of heat. The motion can be maintained by pushing the candles forward as the tube advances. Each element of the tube near the candles is subjected to an alternating stress modified by the thermal expansion. The tube absorbs heat on the side that expands and loses it on the opposite side. Clearly the device converts some portion of the heat into mechanical work that accelerates the rod. In practice I have found that the performance of the engine is limited by its poor capacity for losing heat. I have succeeded in measuring the power output by attaching one end of a length of twine to the tube and a small weight to the other end. The twine coils around the rotating tube and lifts the weight."

Roger Hayward, who illustrates this department, suggests a way of making Clarke's engine fully automatic. He would place a small tank of water between the rails, perhaps a tank made of aluminum foil. The candles would ride across the water in an aluminum boat. The boat could be propelled by a short mast at the bow that engaged the advancing edge of the tube. Those who accept Clarke's challenge can reach him at Forge Cottage, Dogmersfield, Basingstoke, Hampshire, England.

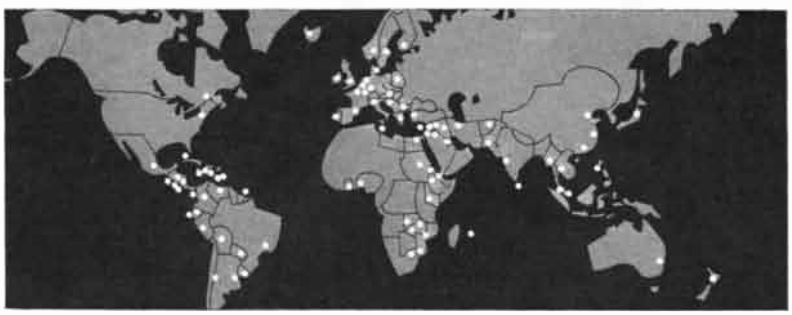
Now, our whole world is precision optical instruments for engineering, geodesy, photogrammetry and microscopy.

From 200 square feet in 1921 to the largest optical instrument complex in Switzerland today. And a sales and service organization upon which the sun never sets.

Size isn't everything. How we got that way is. We're people who believe in, and practice, the most meticulous craftsmanship. We insist that our instruments perform even better than we say they will. We make a fetish of improving our products as fast as technology permits. We're responsive to the needs of our instrument users, and to the environments where they use them.

We're four-thousand people in Switzerland and throughout the world who know that what our customers think of our instruments, they think of us. And we're very sensitive.

WILD HEERBRUGG LTD. CH-9435, HEERBRUGG, SWITZERLAND



IT'S 50 YEARS LATER

50 JAHRE ANS YEARS AÑOS 1921 1971 WILD HEERBRUGG

Take any 2 books for only $2.95...

(values to $34.00)

(retail prices shown)

57900. LASERS. Second Edition. *Bela A. Lengyel.* Re-written, updated version of the author's 1962 trailblazing volume. **$14.95**

52710. THE HIDDEN SEA. *Photographs and Notes by Douglas Faulkner. Text by C. Lavett Smith.* The shimmering world beneath the ocean's surface. **$14.95**

44350. ELEMENTS OF CARTOGRAPHY. Third Edition. *Arthur Robinson and Randall Sale.* **$10.95**

84130. THERMAL PHYSICS. *Charles Kittel.* **$10.95**

63340. MUSIC BY COMPUTERS. *Edited by Heinz von Foerster and James W. Beauchamp.* Includes four 33⅓ rpm records. **$14.95**

42070. THE DEVELOPMENT OF MODERN CHEMISTRY. *Aaron J. Ihde.* **$14.95**

82690. TELESCOPE HANDBOOK AND STAR ATLAS. *Neale E. Howard.* **$12.00**

50350. GEOLOGY OF THE MOON. *Thomas A. Mutch. (Counts as 2 choices.)* **$17.50**

44190. ELEMENTARY PARTICLES AND THEIR CURRENTS. *Jeremy Bernstein.* **$12.00**

69930. POPULATIONS, SPECIES AND EVOLUTION. *Ernst Mayr.* A sweeping exposition of the theory of evolution—complete with latest refinements. **$10.00**

67650. THE PENTAGON OF POWER. *Lewis Mumford.* Examining the fundamental difference between the good life and the "goods life" of our modern culture. **$12.95**

82880. TEXTBOOK OF POLYMER SCIENCE. Second Edition. *Fred W. Billmeyer.* The physics, chemistry and technology of polymers, the most important synthetic materials ever developed by industry. **$15.95**

34390. ASIA, EAST BY SOUTH. Second Edition. *Joseph Earle Spencer and William L. Thomas.* An unusual geography that immediately clarifies much of today's front-line news. **$16.50**

53400. HISTORY OF ROCKETRY AND SPACE TRAVEL. *Wernher von Braun and Frederick I. Ordway III.* **$17.50**

34110. APL PROGRAMMING AND COMPUTER TECHNIQUES. *Harry Katzan, Jr.* The first introductory work on APL—a new, amazingly flexible programming language. **$12.00**

38310. CLASSICAL ELECTRODYNAMICS. *John David Jackson.* **$13.95**

69270. PICTORIAL HISTORY OF AMERICAN MINING. *Harold N. and Lucille L. Sloane.* **$12.50**

79530. A SOURCE BOOK IN PHYSICS. *Edited by William Francis Magie.* **$12.50**

46580. THE EVOLUTION OF MAN AND SOCIETY *C. D. Darlington.* **$12.95**

36310. BIOLOGY AND THE FUTURE OF MAN. *Edited by Philip Handler.* Survey of the life sciences from the molecular level to the ecological. **$12.50**

65210. OCEANOGRAPHY: An Introduction to the Marine Environment. *Peter K. Weyl.* **$12.50**

34050. ANTHROPOLOGICAL ESSAYS. *Oscar Lewis.* **$12.95**

74180. THE RISE OF ANTHROPOLOGICAL THEORY. *Marvin Harris.* **$16.50**

85230. THE INVISIBLE PYRAMID /THE UNEXPECTED UNIVERSE. *Loren Eiseley.* The renowned naturalist eloquently explores man's place in nature. *(Both books count as one.)* **$12.70**

79510. A SOURCE BOOK IN CHEMISTRY, 1400-1900. *Edited by Henry M. Leicester and Herbert S. Klickstein.* **$11.95**

45510. ENVIRONMENT, POWER AND SOCIETY. *Howard T. Odum.* Applies systems analysis to today's challenging ecological problems. **$9.95**

47380. EXPLORING THE OCEAN WORLD. *Edited by C. P. Idyll.* **$14.95**

59840. MAN-MACHINE COMMUNICATION. *Charles T. Meadow.* Ways and means whereby man and computer can pool their talents. **$13.95**

70350. AN INTRODUCTION TO PROBABILITY THEORY AND ITS APPLICATIONS, Vol. I. Third Edition. *William Feller.* **$12.95**

70360. AN INTRODUCTION TO PROBABILITY THEORY AND ITS APPLICATIONS, Vol. II. Second Edition. *William Feller.* **$15.95**

87210. WILDLIFE IN DANGER. *J. Fisher, N. Simon, J. Vincent. Preface by Joseph Wood Krutch.* **$12.95**

36290. BIOPHILOSOPHY. *Bernhard Rensch.* An unprecedented natural philosophy of life based on recent discoveries in biology. **$12.50**

56540. INTRODUCTION TO MATRIX ANALYSIS. Second Edition. *Richard Bellman.* **$14.75**

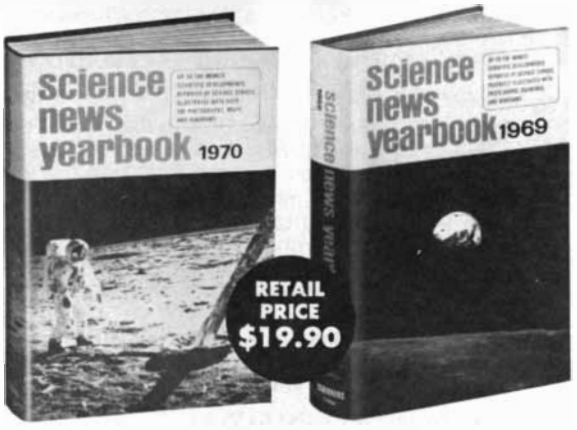
63910. NATURE AND FUNCTION OF SCIENTIFIC THEORIES. *Edited by Robert G. Colodny.* **$12.95**

50110. GENETICS OF THE EVOLUTIONARY PROCESS. *Theodosius Dobzhansky.* A completely rewritten 4th version of the classic *Genetics and the Origin of the Species.* **$10.95**

and get these 2 FREE:

Packed full of facts, photos, maps and diagrams, these are volumes you'll refer to again and again. It's all here: the tremendous advances in space research, crucial experiments in particle physics, new elements discovered in chemistry—and an especially illuminating section on the way the world *views* the uses of science. Just choose two other books you want for only $2.95—and you'll get these two free with a brief trial membership in

The Library of Science



TRIAL MEMBERSHIP APPLICATION

The Library of Science 2-312
Riverside, N.J. 08075

Upon acceptance of this order, please enroll me as a trial member and send the two books whose numbers I have indicated—plus my two FREE Yearbooks. Bill me just $2.95 plus shipping and handling. If not delighted, I will return all books within 10 days and my membership will be cancelled. As a trial member, I need accept as few as 3 more selections during the next 12 months, always at reduced member's prices plus shipping and handling and I may cancel membership any time thereafter. I understand that savings range up to 30%, occasionally more. Each month I will receive advance reviews describing the forthcoming selection, along with a convenient form for requesting alternate selections or no book at all. I understand I may choose a bonus book for every 4 selections purchased. (Introductory offer counts as my first selection toward a bonus book.)

My 2 books for only $2.95 are (write in numbers)

_____ _____

Name_____

Address_____

City_____ State_____ Zip_____

Offer good in Continental U.S. and Canada only.

BOOKS

Archaeology as the meeting ground of the two cultures

by Philip Morrison

THE IMPACT OF THE NATURAL SCIENCES ON ARCHAEOLOGY: A JOINT SYMPOSIUM OF THE ROYAL SOCIETY AND THE BRITISH ACADEMY, organized by T. E. Allibone, F.R.S., and others. Oxford University Press ($24). NOBEL SYMPOSIUM 12: RADIOCARBON VARIATIONS AND ABSOLUTE CHRONOLOGY, edited by Ingrid U. Olsson. John Wiley & Sons, Inc. ($45). SCIENTIFIC METHODS IN MEDIEVAL ARCHAEOLOGY, edited by Rainer Berger. University of California Press ($20). Three symposia are recorded here: one in London at the end of 1969, one in Uppsala a few months before that and one in Los Angeles in 1967. All unite topics from the second law of thermodynamics to the king lists of antiquity with an intimacy to delight the heart of anyone who mourns the division of culture. A couple of authors made it to all three of the meetings.

The physicist and the chemist can read the subtle microinscriptions with which nature dates many objects. The famed radioactive isotope carbon 14 has a mean lifetime of about 8,000 years; it is built into every organism during life and begins its decay at death. It has a heavy impact on the study of the human past, and it is the main thread running through all three volumes. Other threads interweave; the London volume follows two at some length. The first is prospecting for finds below the sea and under the soil, pursued with remarkable virtuosity by geophysical methods, mainly magnetic; the second is microanalysis for what can be learned from the material itself, with methods that do not mar the curator's treasures. (An ancient debaser of the good silver coinage of honest King Yezdegerd I of Persia seems to have been convicted 1,500 years after the fact.) There are promising new forms of internal clock, such as the light that is released on heating ceramic grains; the heating "anneals" electrons that have been excited by radioactivity over a long span of time and then have remained in metastable traps within the crystal lattices of this insulating material. There is an account of one case of the powerful mathematizing trend: the serial assignment of tombs in order of time by advanced statistical treatment of similarity in many dimensions (used to arrange the diverse types of artifacts buried within).

Counting annual layers and reading written documents remain the most reliable means of dating the past. They too are neither simple nor infallible. We moderns date the dynasties of Egypt most securely by our astronomy: we know when Sirius rose, and we are lucky enough to have read some Egyptian documents that tell what day it was by their civil calendar. The first date for which we know this fact is 1872 B.C. Dating the 15 centuries of pharaohs before that still depends mainly on royal annals, particularly on the king lists of the Turin Royal Canon, a fragmentary papyrus that records a total of 955 years "from the beginning of the First Dynasty to the end of the Eighth." There are various cross-checks from monuments, eclipses and the like. History, gaps and all, can be in error overall by nearly a century all the way back to 3120 B.C. For Babylon we date by evidence given us by Ptolemy without error back to 750 B.C. (Alas, his Handy Tables, more practical than his Almagest, are lost; they had a fine king list.) Before that Assyrian king lists work; they are tied to the Egyptian sequence by diplomatic correspondence among ancient rulers. We can almost precisely date the still earlier First Dynasty of Babylon, (there remains an ambiguity of 56 years) because we have their lists of observations of the visibility of Venus as a morning and an evening star. One way or another there is no uncertainty greater than a few decades left in Mesopotamian dates back to about 2100 B.C.

The bristlecone pine forests of the dry White Mountains of the Sierra Nevada provide a marvelous run of tree rings with little error, linking back not through single trees (although one such pine is known that lived for 4,900 years) but through an ingenious and painstakingly checked multiple fit. The Uppsala volume gives not a king list but a ring list, dating all the way back to 5142 B.C.! The oaks of the Main and the Rhine valleys, trees shorter-lived than the bristlecone pine, supply ring dates for their region back to A.D. 800. Carbon-14-dated relics from Egypt agree beautifully with the carbon activity found in old bristlecone wood. The absolute carbon date, however, wanders slowly out of phase; it fits well only back to Roman days. By the time of the First Dynasty of Egypt the carbon date is too young by a millennium.

The carbon of the past was more radioactive than that of the present. The explanation is not certain, but an admirable fit has been proposed by V. Bucha of Prague that ascribes the entire effect to the known slow decline in the earth's magnetic field, the reduced field allowing lower-energy cosmic ray protons to enter the earth's atmosphere. This effect increased the radioactivity of carbon nearly 50 percent at the maximum of 10,000 years ago—about up to the present level of the post-thermonuclear-test burden. The sunspot cycle shows up pretty plainly during the past few hundred years, but it has a much smaller effect in modulating the flux of cosmic rays. Some other wiggles show up in the calibration curves.

The Uppsala conference goes somewhat beyond the dating of man's works to include an account of the oxygen-isotope analysis of the water in an ice core drilled a mile deep right through the Greenland ice cap. The core pieces are kept in a deep freeze by the Army Cold Region Research Engineering Laboratory in Hanover, N.H., and a Danish group analyzed the oxygen isotopes. The isotope ratio gives the temperature of the ancient clouds. Unfortunately the isotopes showed that the thin seasonal layering washes out by molecular diffusion in a rather short time, even in the solid ice. No radioactive method will date the ice without large samples or taking mea-

surements in the field; W. Dansgaard and his partners apply a theory of glacial ice-flow rate to establish their time axis. The scheme checks very well with what is known of the more recent glacial changes. The Danish workers' temperature curve shows that we have been living in a rather uniform postglacial period since most of the ice retreated 12,000 years ago. Our present long warm spell has not been matched for 80,000 years.

The third symposium has a limited theme, one that demands precision from the carbon-14 workers since the time span of interest is an order of magnitude less than the half-life of the isotope. The main problem of precision is often the sample itself: a great roof timber may have died on being felled, but then it may have been reused centuries later, when its outer, last-grown sections had long since been lost. Masonry styles and the design of carpentry joints as date criteria have not yet been studied closely enough, the experts concede. A lintel in an important style-dated barn was dated by carbon 14; it turned out to have been a repair made two centuries after the building went up!

The most heroic proposal is for a study of the silver production of imperial Rome by measuring traces of lead in the Greenland ice, since the smelting of silver releases lead vapor. In antiquity 400 tons of lead had to be smelted to win a ton of silver. Crude estimates from slag heaps suggest that total Roman production of silver—an economic mainstay—reached rates not matched worldwide until about 1750. This might be checked by steam-drilling a wide 1,000-foot vertical shaft into the ice cap, mining from its walls a few hundred big blocks of ice under conditions of extreme lead-cleanliness, melting the blocks into bottles and flying home. In a smaller operation that reached back to 1750 "the investigators, at the time of collecting the blocks, were encased in plastic suits and gloves and worked with Teflon-handled stainless steel tools that had been cleaned in nitric acid." Ten men would live on the crest of the ice cap from summer through summer again, digging the shaft one season and mining the samples the next. One hopes that economic theory will be able to make use of these genuinely hard data.

THE GLOBAL CIRCULATION OF THE ATMOSPHERE, edited by G. A. Corby. Royal Meteorological Society, London ($15). "The first global experiment of the Global Atmospheric Research Programme [will] start in the period 1974–1975." They will have set the date by July, 1971. Here is the background to GARP. It presents 10 review papers at the specialist's level, delivered in the summer of 1969 at a conference in London organized jointly by the American and British meteorological societies (with Canada).

The program is described in some detail by its Swedish chairman, and the actual performance of weather satellites and free balloons is analyzed and described in happy preparation. One can see the first recorded vertical temperature sounding of the atmosphere from orbit, all the way from sea level to an altitude of 50 kilometers, in agreement with the balloon results to better than one degree Celsius. (The balloon of course stops at about 30 kilometers.) The feat can be managed by high-resolution infrared spectroscopy, using the widths of various spectral lines to fix the temperature curve. Wind speeds would come from watching the clouds drift; how high the clouds are would be gained from more infrared studies. Surface pressure might just be found by careful measurement of scattered sunlight, but the new radio altimeters on the horizontally floating balloons look easier: they weigh less than half a pound complete with antenna, and they manage on an average power of a hundredth of a watt. They can fix the true altitude (of a given pressure surface) to within a meter or two—better than the radar used to check them.

The instrumental, systems-design and computing virtuosity displayed in some of these papers is not enough. We need a theory of the general circulation of the atmosphere, a theory heroic in its realism and yet secure in its physical foundations. What it doesn't know can hurt it: water vapor, mountains, ozone 40 kilometers up, sea ice. It needs a net fine enough to catch any lively little embryo of a tropical storm but not so fine as to overload the memory of the big digital looms that weave the net. GARP is only beginning; the long-range global weather forecast is not at hand.

It is nonetheless a hopeful beginning. The free thermal convection of water in a six-inch ring, rotating at two revolutions per minute and heated at the walls, shows model flow patterns remarkably like the global patterns, and a strategy of parameters is taking form. The physical quantities are here, related by equations of approximate validity, but the averages and the maps and tables that hold the input data are intricate beyond the analyst, and so far remain beyond even his nimblest computers. The comparison of smoothed and totaled results with observations can in the end give us the clues we need, but "one of the main problems, as with all energy interactions in the atmosphere, is to separate cause and effect."

The first and most general of the reviews, by E. N. Lorenz of the Massachusetts Institute of Technology, is on the nature of the global circulation. We still do not understand quite what happened to the wonderful idea of G. Hadley, who first argued (in 1735) that the trade winds were the necessary consequence of an obviously elegant circulation: the cold air sinks at the poles and the hot air bubbles up at the Equator, so that aloft the air moves poleward and down below the cool air seeks the Equator. The Equator-seeking air conserves the rotational momentum and gives rise to the westward trades, plain to all who sailed the full-rigged ships. The Hadley cells are in all the geography books but not in the atmosphere. We do not know why. For a generation we have realized that the great eddies we call cyclones replace Hadley's simple loop. We have weather, not mere climate. One meridian is not totally like another. Is it because the Hadley ideal flow is unstable? Or does no such steady and symmetrical flow pattern exist, even under ideal circumstances?

In 1956 a computer model could support Hadley flow; when the flow was perturbed, "cyclones were soon evident" in the printout. Still, we lack a clear answer. According to some views, the weather system moves in distinct wavelike eddies. Others hold that a frankly statistical approach is the best. Yet the patterns are not as random as turbulence or as smooth as wave oscillations. They still elude our theory to show up in this or that numerical experiment. Pressure was the chosen variable of the past generation; men always spoke of highs and lows. Nowadays motion seems to be the key: the flow of energy, angular momentum, mass. The next generation may emphasize water, in the droplets of clouds. It is a wistfully ingenuous theory we now follow; it assumes that water immediately falls out as rain whenever it condenses into droplets of any size. Clouds, however, are real; they sheathe the blue ground of this spinning planet in white, as the space photographs show us. In them a subtle microphysics is at play: dust and ions, nucleation, updraft, lightning.

The equations in the book are typo-

graphically unattractive and even a little confusing. The plates are full of interest. This is a useful, thorough work, aimed at the professional, with an ample research bibliography.

ASYMMETRIC ORGANIC REACTIONS, by James D. Morrison and Harry S. Mosher. Prentice-Hall, Inc. ($24.95). It is close to a round 100 years since chemists confidently began to think of their invisible molecules as real structures in space, generalizing from the snake-ring dream of Kekulé and the patient sorting of tiny crystals by Pasteur. Although physicists before Planck puzzled over the failures of equipartition and built philosophies on the abstractness of energy, contemporary chemists, particularly the great organic synthesizers such as Emil Fischer, became the adepts of atoms in space. Their remarkable insights continue, based on the intricate logic of the nature and yield of subtly chosen large sets of reactions. They draw only here and there on such modern physical results as X-ray diffraction or the magnetic-resonance sampling of the electronic environment of various atomic nuclei.

This up-to-date technical review concerns one of the deepest and most intriguing of stereochemical problems: the occasional unequal behavior of left and right. It is not at all a book for the general reader. Its first chapter is a compact theoretical and historical introduction, and its last sections present some quite general material. The bulk of the volume, however, consists of critical surveys of the work of some 1,200 chemists, all those who have used the notions of left and right for the study of reaction mechanisms and of molecular configurations, or for the preparation of optically active compounds, between 1933, when the last major review appeared, and 1969. The reader is assumed to know the fundamentals of stereochemistry. The dotted-line-and-wedge formulas, with the many terms of the art that freely appear, are generally not even defined.

Our interest lies, however, in the overall state of the science. Enantiomers—molecules related like two gloves—are fully equivalent (at least barring the negligible weak-interaction processes). Nonetheless, the variable flexibility of chemical bonds to rotation, the existence of molecules combining unlike enantiomeric portions that can separately take part in reactions, and the helical packing of symmetrical units (say of urea or of silicon dioxide) into asymmetrical crys-

tals that accordingly take on handedness —all give the subject a puzzling richness suited to the method of the organic chemist. This book has almost no mathematics beyond simple tables of yields, yet it establishes the links of long reaction chains and makes judgments on the approaches and rotations of molecular components, far beyond the present powers of the quantum theorist and his computer.

It is only 20 years since the subject was finally "liberated from the vitalistic heritage of optically active substances." The senior author of the book was one of the pioneers. It is true that handed molecules—the Greek gives us for them the term chiral—can be produced from achiral substances in such a way that unequal amounts of enantiomers can form. Such results can be produced by reagents, by catalysts, even by solvents. What is producing them is now clear. It is no mysterious asymmetrical force exuding from a nearby center but an entirely expected effect of the atomic neighborhood, biasing, say, the ease of approach of the various reacting molecules to one of the possible sites of action.

There are a few results worth citing for their general interest, far removed from the specialist's domain of organic reaction mechanisms. Urea crystallizes into a structure with a helical pattern that has handed interstices on the molecular scale; such crystals can pick up handed molecules from a mixture of left-handed and right-handed ones. Crystals seed their own kind like the crude organisms they are. At the turn of the century C. H. Soret crystallized sodium chlorate many times in sealed glass ampoules. "The molecules of sodium chlorate are achiral but the crystal is chiral. Dextrorotatory [right-handed] crystals formed 433 times; levorotatory [left-handed] crystals, 411 times; and a mixture of both, 94 times." In 1969 bromine was shown to react with the chiral crystals of a molecularly achiral but complicated compound to yield an excess of 6 percent of a molecularly chiral product. The classic physical production of excess enantiomer was carried out in 1930, when Werner Kuhn, using circularly polarized light of a wavelength chosen for its disrupting effect, preferentially broke up one of the members of a 50-50 molecular mixture of a certain ester. Still unconfirmed is the recent report that irradiation with the chiral electrons from the beta decay of strontium 90 (parity nonconservation tends to align their spins) destroys the amino acid D-tyrosine at a smaller dose than it does the mirror molecule L-tyrosine. But the D-L mixture was not made measurably more left-handed during a few months of exposure.

THE ORIGINS OF FEEDBACK CONTROL, by Otto Mayr. The M.I.T. Press ($7.95). James Watt was both a highly practical man and a reflective one. He wrote in 1783: "There is no end of millwrights once you give them leave to set about what they call machinery:—they have multiplied wheels upon wheels until it has now almost as many as an orrery." One of those pieces of wheelwork was the new centrifugal pendulum Watt's partner Matthew Boulton saw in a mill in 1788, which regulated the distance of the upper and the lower millstone as the speed changed. By the end of that year the drafting room at Boulton and Watt had sent to the shop the working drawings for the most famous of the early feedback control devices, the flyball governor. Watt knew he had adapted a millwright's device; he never tried to take out a patent. On the evidence of hearsay governors had already been used to regulate millstones for a generation. At first they were not true feedback devices; only in 1787 did Thomas Mead describe a system in which the flyballs worked to control the sail area of a windmill.

Otto Mayr is a mechanical engineer, now a curator at the Smithsonian, who translated this compact and knowing book from his own German original. He follows feedback from the earliest times, applying the block-diagram techniques of control theory to distinguish true closed-loop feedback from those systems, such as friction-loaded prime movers, that have no physically distinct comparators or sensors and are only formally closed-loop systems. Nor can we accept devices in which man closed the loop; human feedback has roots deep in evolutionary history.

A regulated water supply for the water clock was the first feedback device. There was a float whose position controlled the valve orifice, thus maintaining a constant water level. It is described in Alexandrian times; the tradition remains strong throughout the Arab world until the 13th century. Then "between 1206 (al-Jazarī) and the mid-18th century no references to float valves are known." Even the 1575 edition of Hero of Alexandria, who describes several of them, had no influence; the baroque authors simply ignored the float regulator, although Hero himself became celebrated. The first new feedback device since ancient times was a thermostatic damper control for incubators for chick eggs; it was invented by the ingenious Netherlander Cornelis Drebbel, if we can accept an impressive but not contemporaneous set of records. Robert Boyle knew that Drebble ("that great, singular, learned mechanick") had found out how to stabilize temperature a generation earlier, but he did not know how it had been done.

All three lines of development flowered into practical use during the last half of the 18th century. Steam boilers routinely had float-valve water-level controls (so did water closets); bimetallic thermostats spread after Bonnemain's patent of 1783 (even his first name is unknown); steam-engine governors became ubiquitous. The author suggests that this reversal of order has a philosophical origin: men of the baroque period believed in automata, in programmed control, in mercantilism and in fixed law. With the epoch of Adam Smith, of liberalism, *laissez faire* and the self-regulating market, the time of the feedback loop had arrived. Both the climate of economic thought and the inventors' concerns reflected "the intellectual currents of the era."

The general theory of feedback had to wait even beyond the development of specific formal theories of automatic control. Not until the 1930's did electrical and electronic work offer scope for novel and generalizing nonlinear theories. Out of that beginning control engineering grew to maturity during the forced draft of World War II; on both sides there arose a proved technology, a general and useful theory and a large number of trained men.

The first truly automatic servo (a servo is a closed loop that can follow a *changing* command signal) was the fantail, a small windwheel mounted at right angles to the main sails and geared to turn the entire top of the windmill. Once the wind blows parallel to the plane of the tail wheel the action stops. The device was patented first by the Lancaster blacksmith Edmund Lee about 1745 and spread to general use in Britain and Germany, but never to the rest of Europe.

LETTER AND IMAGE, by Massin. Translated by Caroline Hillier and Vivienne Menkes. Van Nostrand Reinhold Company ($20). The theory of information rests on one massive insight: even symbols must in some sense be materially realized. DNA itself is not merely a

magical charm more powerful than the incantations of Faust; it is a coiled polymer of certain nucleotides, and its cross-links, its breaks, its elasticity and its reactions transcend chemistry to become biological actors in the evolutionary drama.

Men's symbols too must have form. Designers have known it for millenniums. "A leading French graphic designer" has put together more than 1,000 diverse and fascinating illustrations around the theme of the symbol turned real, and has given us as well an informed and particularly well-documented text. There are two large divisions: the letters of the alphabet drawn to have meaning beyond the symbolic, and "calligrams" (as Apollinaire named his "lyrical ideograms," the written word given visual form).

We have here a cornucopia in one small corner of graphic history. Letters abound, whether they are in illuminated manuscript, in the cut-steel matrixes of the Renaissance, in engravings, in lithographs or in an animated cartoon (*Rythmetic,* produced by Norman MacLaren, in which you can see "the number 5 scratch its back with its transverse bar"). There are *A*'s from all the periods of European books: *A* as a pair of dividers, as a bird with a fish (in illumination), as a bishop handing a nobleman a holy book, as a couple of flying buttresses, as a cast-iron triangle, as a man hoeing, as an easel, as a pair of nudes (in several versions, chastely acrobatic or erotic), as clowns, as grandees, as Mephistopheles in silhouette leading the little black dog, as a sculptor at work, as birds, as flowers, as the floor plan of an imaginary palace, as a railroad station (a huge *A*-frame). The cordwood of the old *Nature* masthead is omitted, but little else. And the list goes on to Z! There are letters formed by the pieces in chess problems, music manuscript with half notes made of turbaned Turks, the shorter notes the heads of their marching black spearmen. Figured poems and prose span time from a version of Aratus' astronomical poem in Latin done on parchment (10th century, with the text shaped into the zodiac figures) up to the wonderful poster by Ryuichi Yamashiro done for a reforestation campaign a couple of years ago (with the Japanese character for *tree* brushed in all sizes and groups to make a true forest on the page). Computer graphics, those portraits done in letters and symbols, are nothing new; here is Abraham Lincoln's face constructed long ago in a copperplate version of the Proclamation of Emancipation.

INDEX OF ADVERTISERS

JULY 1971

BIBLIOGRAPHY

Readers interested in further reading on the subjects covered by articles in this issue may find the lists below helpful.

THE CENSUS OF 1970

THE CHANGING POPULATION OF THE UNITED STATES. Conrad Taeuber and Irene B. Taeuber. John Wiley & Sons, Inc., 1958.

NEGROES IN CITIES: RESIDENTIAL SEGREGATION AND NEIGHBORHOOD CHANGE. Karl E. Taeuber and Alma F. Taeuber. Aldine Publishing Company, 1965.

CENSUS, U.S.A.: FACT FINDING FOR THE AMERICAN PEOPLE, 1790–1970. Ann Herbert Scott. The Seabury Press, 1968.

THE CHALLENGE OF AMERICA'S METROPOLITAN POPULATION. Patricia Leavey Hodge and Philip M. Hauser. Praeger Publishers, 1970.

POPULATION GROWTH AND AMERICA'S FUTURE: AN INTERIM REPORT TO THE PRESIDENT AND CONGRESS. Commission on Population Growth and the American Future. U.S. Government Printing Office, 1971.

THE INDUCTION OF INTERFERON

VIROLOGY MONOGRAPHS, 6: INTERFERON. J. Vilček. Springer-Verlag New York Inc., 1969.

INTERFERON: PROCEEDINGS OF A SYMPOSIUM SPONSORED BY THE NEW YORK HEART ASSOCIATION. Little, Brown and Company, 1970.

DOUBLE-STRANDED RNAS (POLY I:C) IN THE PREVENTION OF VIRAL INFECTIONS. Maurice R. Hilleman in *Archives of Internal Medicine,* Vol. 126, No. 1, pages 109–124; July, 1970.

A NEW CLASS OF DIODE LASERS

GALLIUM ARSENIDE LASERS. Edited by C. H. Gooch. John Wiley & Sons, Inc., 1969.

GaAs–$Ga_xAl_{1-x}As$ HETEROSTRUCTURE INJECTION LASERS WHICH EXHIBIT LOW THRESHOLDS AT ROOM TEMPERATURE. I. Hayashi and M. B. Panish in *Journal of Applied Physics,* Vol. 41, No. 1, pages 150–163; January, 1970.

OPTICAL COMMUNICATIONS RESEARCH PROGRESS. S. E. Miller in *Science,* Vol. 170, No. 3959, pages 685–695; November 13, 1970.

PREPARATION OF MULTILAYER LPE HETEROSTRUCTURES WITH CRYSTALLINE SOLID SOLUTIONS OF $Al_xGa_{1-x}As$: HETEROSTRUCTURE LASERS. M. B. Panish, S. Sumski and I. Hayashi in *Metallurgical Transactions,* Vol. 2, No. 3, pages 795–801; March, 1971.

GaAs–$Al_xGa_{1-x}As$ DOUBLE HETEROSTRUCTURE INJECTION LASERS. I. Hayashi, M. B. Panish and F. K. Reinhart in *Journal of Applied Physics,* in press.

PATHWAYS IN THE BRAIN

SYNAPTIC DISTRIBUTION OF CENTRIPETAL AND CENTRIFUGAL NERVE FIBRES IN THE OLFACTORY SYSTEMS OF THE RAT: AN EXPERIMENTAL ANATOMICAL STUDY. Lennart Heimer in *Journal of Anatomy,* Vol. 103, Part 3, pages 413–432; November, 1968.

CONTEMPORARY RESEARCH METHODS IN NEUROANATOMY: PROCEEDINGS OF AN INTERNATIONAL CONFERENCE HELD AT THE INSTITUTE OF PERINATAL PHYSIOLOGY. Edited by W. J. H. Nauta and O. E. Ebbesson. Springer-Verlag New York Inc., 1970.

THE SACRED SOURCE OF THE SEINE

RAPPORT SUR LES DÉCOUVERTES ARCHÉOLOGIQUES FAITES AUX SOURCES DE LA SEINE. Henri Baudot in *Mémoires de la Commission des Antiquités du Département de la Côte-d'Or,* Vol. 2, pages 95–144; 1842–1846.

LES BRONZES D'ART DES SOURCES DE LA SEINE. Henry Corot in *Mémoires de la Commission des Antiquités du Département de la Côte-d'Or,* Vol. 20, No. 1, pages 107–120; 1933.

SCULPTURES EN BOIS DÉCOUVERTES AUX SOURCES DE LA SEINE. R. Martin in *Revue Archéologique de l'Est et du Centre-Est,* Vol. 14, No. 1–3, pages 7–35; January–September, 1963.

SUPERNOVA REMNANTS

THE STORY OF THE CRAB NEBULA. N. U. Mayall in *Science,* Vol. 137, No. 3524, pages 91–102; July 13, 1962.

SUPERNOVAE AND SUPERNOVA REMNANTS. A. Poveda and L. Woltjer in *The Astronomical Journal,* Vol. 73, No. 2, Part 1, pages 65–74; March, 1968.

SUPERNOVAE. I. S. Shklovsky. Wiley–Interscience, 1969.

ROTATING NEUTRON STARS AND THE NATURE OF PULSARS. Thomas Gold in *Nature,* Vol. 221, No. 5175, pages 25–27; January 4, 1969.

MAGNETIC DIPOLE RADIATION FROM PULSARS. James E. Gunn and Jeremiah P. Ostriker in *Nature,* Vol. 221, No. 5179, pages 454–456; February 1, 1969.

A GRAZING ECOSYSTEM IN THE SERENGETI

REPORT ON AN ECOLOGICAL SURVEY OF THE SERENGETI NATIONAL PARK, TANGANYIKA. W. H. Pearsall in *Oryx,* Vol. 4, No. 1, pages 71–136; April, 1957.

A STUDY OF THE GAME OF THE SERENGETI PLAINS. M. Grzimek and B. Grzimek in *Zeitschrift für Säugetierkunde,* Vol. 21, pages 1–61; 1960.

SELECTION OF VEGETATION COMPONENTS BY GRAZING UNGULATES IN THE SERENGETI NATIONAL PARK. M. D. Gwynne and R. H. V. Bell in *Nature,* Vol. 220, No. 5165, pages 390–393; October 26, 1968.

PHOTONS AS HADRONS

TOTAL PHOTOABSORPTION CROSS SECTIONS UP TO 18 GeV AND THE NATURE OF PHOTON INTERACTIONS. D. O. Caldwell, V. B. Elings, W. P. Hesse, G. E. Jahn, R. J. Morrison, F. V. Murphy and D. E. Yount in *Physical Review Letters,* Vol. 23, No. 21, pages 1256–1260; November 24, 1969.

PHOTOPRODUCTION OF NEUTRAL RHO MESONS. H. Alvensleben, U. Becker, William K. Bertram, M. Chen, K. J. Cohen, T. M. Knasel, R. Marshall, D. J. Quinn, M. Rohde, G. H. Sanders, H. Schubel and Samuel C. C. Ting in *Nuclear Physics B,* Vol. B 18, No. 2, pages 333–365; May 1, 1970.

HIGH-ENERGY PHOTONUCLEAR REACTIONS AND VECTOR-MESON DOMINANCE. E. M. Henley in *Comments on Nuclear and Particle Physics,* Vol. 4, No. 3, pages 107–115; May–June, 1970.

TOTAL HADRONIC (γ,p) AND (γ,d) CROSS SECTIONS FROM 4 TO 18 GeV. D. O. Caldwell, V. B. Elings, W. P. Hesse, R. J. Morrison, F. V. Murphy, B. W. Worster and D. E. Yount in *Physical Review Letters,* Vol. 25, No. 9, pages 609–612; August 31, 1970.

THE AMATEUR SCIENTIST

COFFEE-TABLE HOLOGRAPHY. John Landry in *Journal of the Optical Society of America,* Vol. 56, No. 8, page 1133; August, 1966.

HOLOGRAPHY FOR THE SOPHOMORE LABORATORY. Robert H. Webb in *American Journal of Physics,* Vol. 36, No. 1, pages 62–63; January, 1968.

INTRODUCING THE VEGA GT. AND THE VEGA GT.

When we designed the Vega, we knew it would ride and handle better than most little cars. Because we built it low. And wide. And we built it with precise steering, and coil springs at each wheel. On paper, Vega was not only a great little economy car, it was a great little sports car.

Much to the credit of our engineers, Vega is even better on the road than it was on paper. It darts around town like a little champ. And it cruises freeways like a much bigger car.

So we decided to build a GT. For handling, we put in front and rear stabilizers, and our 110-hp (93 SAE net) aluminum overhead-cam engine. Plus special 13 x 6 wheels and A70 x 13 bias belted ply white lettered tires. Then we added instrumentation: tach, clock, ammeter, temperature gauge. And a sporty soft vinyl-covered four-spoke steering wheel.

And we spruced it all up, too. With a black grille. Special GT nameplates. Woodgrain accents.

And more.

Then, we were so happy with the way our GT came out, we decided to come out with two.

The Vega GT Hatchback Coupe. And the Vega GT Kammback Wagon. Whee.

GM
MARK OF EXCELLENCE

Buckle your seat and shoulder belts.
It's an idea you can live with.

VEGA
CHEVROLET

Next we'll deliver the people to live in it.

What you see here is housing by air. With a lift from United Aircraft. This idea of using our Skycrane® helicopter to deliver factory-built houses has a lot going for it. Low cost. Speed. Hope for easing the acute housing shortage. And another of our big copters with military experience could easily meet another urgent need: fast, efficient movement of people between cities. New York to Boston in 1 hour 11 minutes, for example. Downtown to downtown, with no long airport rides. Carrying people and things isn't all. What we make at United Aircraft—from jet engines to microcircuits—also carries answers to some of the nation's most critical needs.

United Aircraft®

PRATT & WHITNEY AIRCRAFT • HAMILTON STANDARD • SIKORSKY AIRCRAFT • NORDEN • UNITED TECHNOLOGY CENTER • UNITED AIRCRAFT RESEARCH LABORATORIES • UNISEM • UNITED AIRCRAFT OF CANADA LIMITED • TURBO POWER & MARINE SYSTEMS •