SCIENTIFIC AMERICAN

MOTION PERCEPTION

$1.25

June 1975

Truck drivers have always known what motorists need to know about aluminum.

That aluminum can take it. That it can be engineered to perform as well as the material it replaces. Yet, because it's a weight-saver, aluminum gives you a lighter vehicle that saves gas, brake wear, tire wear, even registration costs in some states.

Aluminum has performed so well in trucks, which are built to last hundreds of thousands of miles, that automotive engineers are now evaluating additional aluminum applications for the automobile—specifically, hoods, wheels, trunk decks, bumpers and bumper reinforcements . . . even, in some cases, aluminum doors. They understand better than most of us that American cars must become less expensive to operate, without sacrificing strength. So, they're specifying strong, lightweight aluminum alloys specially designed for passenger cars.

Sturdy aluminum components that save weight in trucks.

The men who drive truck rigs equipped with aluminum cabs, frame rails, radiator parts, bumpers, fuel tanks, suspension parts and forged disc wheels know how well aluminum stands up over the long haul. Under rough treatment. In all kinds of weather. They also know that lightweight aluminum components pay off financially. Because a lighter rig can carry more payload, yet requires less fuel when rolling. Either loaded or empty.

If you'd like a more in-depth look at how aluminum in automobiles helps in the conservation of energy, send for our brochure, *Energy, Aluminum and the Automobile.* Write to Aluminum Company of America, 986-F Alcoa Building, Pittsburgh, PA 15219.

The reasons for using aluminum are found in aluminum itself.

ALCOA

Established 1845 **SCIENTIFIC AMERICAN** June 1975 Volume 232 Number 6

ARTICLES

DEPARTMENTS

PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. COPYRIGHT © 1975 BY SCIENTIFIC AMERICAN, INC. ALL RIGHTS RESERVED. PRINTED IN THE U.S.A. NO PART OF THIS ISSUE MAY BE REPRODUCED BY ANY MECHANICAL, PHOTOGRAPHIC OR ELECTRONIC PROCESS, OR IN THE FORM OF A PHONOGRAPHIC RECORDING, NOR MAY IT BE STORED IN A RETRIEVAL SYSTEM, TRANSMITTED OR OTHERWISE COPIED FOR PUBLIC OR PRIVATE USE WITHOUT WRITTEN PERMISSION OF THE PUBLISHER. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $15 PER YEAR, U.S., POSSESSIONS AND CANADA; $18 PER YEAR, ALL OTHER COUNTRIES. EUROPEAN SUBSCRIPTION COPIES MAILED C/O PUBLICATIONS DISTRIBUTION SERVICE, ST. KATELIJNEVEST 14, B2000 ANTWERP, BELGIË.

THE COVER

The photograph on the cover shows what the world might look like to a stroller if the eye were literally like a camera. Since the human visual system has no shutter, the images formed by the lens of the eye would flow across the retina in a blur. The fact is, of course, that the stroller does not perceive his surroundings as a blur; they appear sharp and fixed across his entire field of view. This demonstrates that, although the eye itself may act as a camera, the visual system as a whole acts more like a computer operating according to specific mathematical rules (see "Visual Motion Perception," by Gunnar Johansson, page 76). The scene in the photograph is a street in Manhattan, looking south toward the new World Trade Center towers. The photographer, Fritz Goro, fitted his camera with a zoom lens (originally developed for motion-picture photography), the focal length of which can be continuously changed. When the lens was zoomed during a brief time exposure, the effect was as if the camera had been moved.

THE ILLUSTRATIONS

Cover photograph by Fritz Goro

Page	Source
14	Sterling Wortman, Rockefeller Foundation
16–21	Ilil Arbel
23	Edward S. Reynolds, Harvard Medical School
24–27	Tom Prentiss
28–29	Ilil Arbel
30–31	Tom Prentiss
32–35	Sarah Landry
36	Sarah Landry (*top*); Edward O. Wilson, Harvard University (*bottom*)
51	Stanford Linear Accelerator Center
52–62	George V. Kelvin
66–75	Dan Todd
77–86	Alan D. Iselin
87	Gunnar Johansson, University of Uppsala (*left*); Alan D. Iselin (*right*)
91	James Tormey
92–97	Allen Beechel
99	Ben Rose
100	National Central Library, Florence (*top*); Ilil Arbel (*bottom*)
101	Ilil Arbel
102	Stillman Drake, University of Toronto
103	Ilil Arbel
106–110	Ilil Arbel
113	John Sanford
114–120	Jerome Kuhl

LETTERS

Sirs:

In "Visual Pigments and Color Blindness" [SCIENTIFIC AMERICAN, March] W. A. H. Rushton conjectures that the English physician Thomas Young wrote little about his trichromatic theory of color vision because he did not want to disquiet his patients by advertising "his addiction to science." I suggest that the story is a bit more complicated.

In 1802, in the first of a famous series of lectures that laid the foundations of physical optics, Young wrote: "As it is almost impossible to conceive each sensitive point of the retina to contain an infinite number of particles, each capable of vibrating in perfect unison with every possible undulation, it becomes necessary to suppose the number limited, for instance to the three principal colors, red, yellow and blue, and that each of the particles is capable of being put in motion more or less forcibly by undulations differing less or more from perfect unison. Each sensitive filament of the nerve may consist of three portions, one for each principal color."

Within the year Young was moved to make an important correction: "It becomes necessary to modify the supposition I advanced... respecting the fibers of the retina, substituting red, green and violet for red, yellow and blue."

That is all he ever wrote about the "particles" and "fibers" of the retina in relation to color vision. Since he lectured and published freely in many fields of science, including vision and the eye, I submit that he was not reluctant to be known as a scientist.

The editor of the second (posthumous) edition of Young's lectures omitted his suggestion of three kinds of retinal receptors. James Clerk Maxwell, in 1872, credited Hermann von Helmholtz with supplying this element in the theory. On Maxwell's authority it is known today as the Young-Helmholtz theory. Yet Helmholtz himself, in 1860, having read the original lecture as published in *Philosophical Transactions of the Royal Society,* had credited Young with originating the notion of the three retinal receptors. Thus Helmholtz does not deserve (any more than he needs) the attachment of his name to this theory.

If any second name should be attached to Young's, it is that of the English chemist George Palmer. In 1777, 25 years before Young, he wrote: "Each ray of light is composed of three rays only: one appears yellow, one red, and the other blue.... The surface of the retina is composed of particles of three different kinds, corresponding to the three rays of light...."

Note that Young also used the term "particles." It is not known, however, whether he ever read Palmer.

DAVID L. MACADAM

Eastman Kodak Company
Rochester, N.Y.

Sirs:

Philip Morrison's review of *A Plague of Corn,* by Daphne A. Roe [SCIENTIFIC AMERICAN, March], mentions that the hereditary theory of pellagra was still cited in an American medical journal of 1916. The relation between pellagra and heredity had consequences that continued long after that time.

According to a forthcoming book by Allan Chase, *Killers of Hope,* a team of eugenicists led by Charles B. Davenport was carrying out around 1915 genetic studies on human groups with high pellagra incidence. At the same time Joseph Goldberger and his team were doing nutritional studies on pellagra. Three years after Goldberger, in a major scientific triumph for the U.S. Public Health Service, had established that pellagra was a nutritional deficiency (later found to be due to the lack of niacin) Davenport and his associates published their 500-page report including pedigrees that purported to prove that pellagra was a genetic disease, present in the inferior, chronically pauper stock of the poor people in the South of the U.S. Goldberger's work was cited in a footnote.

According to Chase, the work of these eugenic "scholars" was mainly responsible for the fact that the U.S. Government failed to act on Goldberger's discovery. Only the Great Depression, forcing the installation of relief programs, brought about the eradication of pellagra. By 1945, after white bread was required by law to be enriched with vitamins, the disease had practically disappeared.

It is interesting to compare the claims of the early eugenicists concerning pellagra with those of present-day I.Q. hereditarians. Both sets of claims are presumably the products of "research" biased by social prejudice and likely to find sympathetic echoes in biased audiences.

S. E. LURIA

Center for Cancer Research
Massachusetts Institute of Technology
Cambridge

Sirs:

While reading a collection of old letters recently, I was gratified to discover that *Scientific American* was encouraging women in the sciences as long ago as 1879. A young lady in an astronomy class at Vassar College wrote that year:

"We have to look every night for the comet. After chapel we make a grand rush for the telescopes.... We thought we had found the comet one night, but it turned out to be a star as seen with a poor focus. We get ten dollars a month for calculating the rising and setting of the planets for the *Scientific American,* that is just a dollar apiece. Query:—how many are there in the class? Prof. Mitchell says, however, that we shall not be paid in the world for poor work, and she thinks we might as well learn the lesson now, so if on looking over our work she finds a mistake, she deducts fifty cents and gives it toward the scholarship. She gave it about ten dollars that way last year."

ELIZABETH N. SHOR

La Jolla, Calif.

Scientific American, June, 1975; Vol. 232, No. 6. Published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Gerard Piel, president; Dennis Flanagan, vice-president; Donald H. Miller, Jr., vice-president and secretary; George S. Conn, treasurer; Arlene Wright, assistant treasurer.

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

Advertising correspondence should be addressed to Harry T. Morris, Advertising Director, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017.

Offprint correspondence and orders should be addressed to W. H. Freeman and Company, 660 Market Street, San Francisco, Calif. 94104. For each offprint ordered please enclose 30 cents.

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

NAME

NEW ADDRESS

OLD ADDRESS

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

JUNE, 1925: "Light may be spread out into its spectrum by means of the diffraction grating, which consists of a polished surface having about 20,000 parallel lines ruled across it for each lineal inch. Knowing this fact, how would one go about the problem of similarly spreading out the X rays, which are about 10,000 times shorter? A sufficiently fine grating could not be ruled, but in 1912 M. von Laue, a German scientist, discovered another method of doing the same thing. It occurred to him that if the atoms of matter were spaced as closely together as theoretical considerations indicated they should be, then they would themselves make an excellent diffraction grating for X rays. Von Laue at once put his theory to the test and was rewarded by a clear demonstration that it was valid. The atoms of crystals constituted a grating of just the right spacing, and were neatly and geometrically arranged in rows and tiers. The method has been perfected by Sir William Bragg, and it enables scientists to find out exactly how the atoms of various substances are grouped."

"The chimpanzee is familiar to most of us because he has become a motion-picture performer. Few of the performances of Hollywood apes are, however, true ape acts but are only human actions taught to apes. Wolfgang Köhler, professor of philosophy at the University of Berlin, has spent a number of years studying how untrained chimpanzees act and think. He has found that, given a stick of sufficient length, a chimpanzee will use it as an extension of his arm to draw toward him a piece of fruit he cannot otherwise reach. Would he, however, be able to form the necessary mental picture in advance to enable him to make use in a similar manner of a short stick to draw toward him a long stick that would in turn enable him to reach the prize? The chimpanzee Sultan solved this puzzle."

"Professor H. S. Jennings of Johns Hopkins University has made a contribution to the controversy of heredity v. environment in which he criticizes adversely most current interpretations of the Mendelian law. He finds much more of a place for the effects of environment than is usually granted. Jennings points out that men and other organisms do not inherit 'characteristics' at all; what their parents leave them are certain packets of chemicals. Under one set of conditions these chemicals may produce one set of characters and under other conditions other characters."

"The legislature of Tennessee has forbidden the teaching of the theory of evolution in state-supported schools. Three other Southern states have taken similar action."

SCIENTIFIC AMERICAN

JUNE, 1875: "It has long been supposed that no direct mechanical effects could be produced when luminous rays were allowed to fall upon one end of a most delicately balanced lever arm suspended in vacuum. Mr. William Crookes has now proved conclusively, by experiment, that such mechanical effects can be achieved. To do so he has employed a new instrument, called by him a radiometer. The radiometer consists of four small pith disks, fixed at the extremities of two crossed arms of straw, balanced upon a pivot at the point where the straws cross each other, so that they can spin around on the pivot. The pith disks are white on one side and blackened with lampblack on the other. The entire arrangement is enclosed in a glass bulb from which the air is removed. The disks and crossed arms spin around rapidly when submitted to the action of light. Strange indeed are the thoughts started by this revelation that light, pouring upon bodies freed from atmospheric friction, is in itself an active force."

"A popular theological dogma declares that life is the grand object of creation, that the composition as well as the contour of the earth's surface has special reference to its habitability, and that all things show a ruling design to fit the world to be the home of sentient creatures, more especially of man. Strictly speaking, Science has nothing to do with such dogmas. It has no means of discovering the ultimate purposes of things, and no time to waste on their discussion. Nevertheless, it is difficult sometimes not to take an indirect interest in the claims of those who presume to decide such questions, at least so far as to notice how aptly the facts of Nature contradict their assertions. Thus in the present case it would be much easier to sustain the contrary thesis, namely that so far from having been made what it is that it might be inhabited, the earth became what it is through being inhabited. In short, life has been the means to, not the end of, the earth's development. In the light of recent discoveries, Byron's poetic extravagance, 'The dust we tread on was alive!' becomes a simple statement of observed fact."

"Professor Cornu, of the École Polytechnique in Paris, has put into use a new instrument for measuring the velocity of light between two stations. An electrical registering apparatus is used, giving, it is believed, more accurate measurements than the well-known toothed-wheel arrangement of Fizeau. Foucault fixed the velocity of light, by his instrument, at 185,157 miles per second. Professor Cornu, by his new instrument, fixes the velocity of light at 186,660 miles per second."

"England and other nations are now adopting the torpedo system of attacking an enemy's vessels. The boats specially constructed for this submarine warfare carry their deadly bombs on the ends of spars extending from the bows of the boat. The torpedoes are cylindrical in form, with hemispherical heads bearing percussion fuses or some other means of detonating the explosive. The torpedo launch rams the enemy boat with sufficient force that the torpedo, striking below the water line, is exploded."

"M. Pasteur, the distinguished French chemist, has recently published an exhaustive treatise on the diseases of the silkworm, resulting from his investigations in the French silk-manufacturing district. The enormous mortality that, during certain years, has happened among the silkworms M. Pasteur ascribes to two diseases. The first is characterized by the presence, in all the organs of the worm or butterfly, of small ovoid corpuscles, invisible except when magnified 400 or 500 times. The second is recognized by the presence of a particular ferment in the digestive tube or stomach."

"A gas-burning cooking stove has been invented by B. Giles, of Blackheath, England. He claims to have succeeded in cooking the most delicate dishes without their imbibing the slightest flavor from the products of combustion."

PLANNED UN-OBSOLESCENCE.

The Bell System's research and manufacturing people design and build phones to last. In fact, our phones must pass "torture tests" that simulate 20 years of average use. Not that you'll keep your phone that long. But isn't it good to know that if you finally decide to get a new phone it's because you want one instead of need one? The Bell Companies know you want a dependable, reliable phone.

WE HEAR YOU.

THE AUTHORS

STERLING WORTMAN ("Agriculture in China") is an agricultural researcher, educator and administrator who is currently a vice-president of the Rockefeller Foundation. Born in Oklahoma, he attended Oklahoma State University, receiving a B.S. in 1943. He did his graduate work at the University of Minnesota, earning his Ph.D. in plant genetics there in 1950. His association with the Rockefeller Foundation began soon afterward, when he joined the staff as a geneticist in charge of corn breeding. His tenure at the foundation was interrupted by two extended stints at the Pineapple Research Institute in Hawaii, first as head of the department of plant breeding from 1955 to 1959 and later as director of that institute from 1964 to 1965. In between he worked as an assistant director of the Rockefeller Foundation's International Rice Research Institute in the Philippines. He returned to the New York headquarters of the foundation in 1966 to become director of the agricultural sciences division and was appointed a vice-president in 1970.

ATTALLAH KAPPAS and ALVITO P. ALVARES ("How the Liver Metabolizes Foreign Substances") are professors at Rockefeller University. Kappas holds an A.B. degree from Columbia University and an M.D. from the University of Chicago. After medical school he spent seven years devoted to clinical training in internal medicine and scientific training in steroid biochemistry at medical institutions in Boston and New York. He then returned to Chicago and taught medicine there for 10 years before joining the Rockefeller faculty in 1967. He was appointed physician-in-chief of the university's research hospital in 1974. Alvares did his undergraduate work at the University of Bombay in his native India. He came to the U.S. in 1958 and went on to obtain his M.S. in biochemistry from the University of Detroit in 1961 and his Ph.D. in pharmacology from the University of Chicago in 1966. After two years of postdoctoral training at the University of Minnesota he joined the Wellcome Research Laboratories in 1967 as a senior research biochemist. He has been at Rockefeller University since 1970.

EDWARD O. WILSON ("Slavery in Ants") is professor of zoology at Harvard University and curator in entomology of that university's Museum of Comparative Zoology. A graduate of the University of Alabama, Wilson acquired his Ph.D. in biology from Harvard in 1955; he has been a member of the Harvard faculty ever since. His research interests focus on the biology of the social insects, the classification of ants, sociobiology and biogeography, and he has published a number of books on these subjects, including *The Insect Societies* (1971). He comments: "Having just finished a treatise on all aspects of social life in animals and early men (*Sociobiology: The New Synthesis,* published by the Harvard University Press), I am now devoting full time to my laboratory and field studies on caste and slavery in ants."

SIDNEY D. DRELL ("Electron-Positron Annihilation and the New Particles") is professor of theoretical physics at Stanford University, where he also serves as deputy director of the Stanford Linear Accelerator Center (SLAC). He did his undergraduate work at Princeton University and his graduate work at the University of Illinois, receiving his Ph.D. in 1949. In addition to his research, which won him the Atomic Energy Commission's Ernest Orlando Lawrence Memorial Award in 1972, Drell has been active in the area of science and public affairs. He has served as a consultant to various Government agencies and laboratories, and he was a member of the President's Science Advisory Committee from 1966 to 1970. At present he is chairman of the advisory panel on high-energy physics for the U.S. Energy Research and Development Administration. For relaxation he enjoys playing chamber music on the violin.

JOHN R. PERCY ("Pulsating Stars") is associate professor of astronomy at the University of Toronto. A native of England, he was raised and educated in Toronto, obtaining all his degrees (a B.Sc. in mathematics and physics in 1962, an M.A. in astronomy in 1963 and a Ph.D. in astronomy in 1968) from the University of Toronto. He has been on the faculty of that university since 1968, except for the year 1972–1973, when he was a visiting fellow at the University of Cambridge. "In the past few years," he writes, "I have derived great pleasure and satisfaction as an active but junior collaborator in molecular biology with my wife, who is an immunochemist."

GUNNAR JOHANSSON ("Visual Motion Perception") is professor of psychology at the University of Uppsala in Sweden. His doctorate is from the University of Stockholm, where he taught for a number of years before joining the Uppsala faculty in 1957 as head of the psychology department. Since then he has spent a couple of years in the U.S. as a fellow of the Center for Advanced Study in the Behavioral Sciences at Stanford University. Johansson reports that in addition to his basic research on the visual perception of motion he is engaged in "applied research on man-machine systems, particularly in the field of traffic safety."

JAMES N. BUTLER ("Pelagic Tar") is Gordon McKay Professor of applied chemistry at Harvard University, where he is also a member of the faculty of geological sciences and the committee on oceanography. After receiving his Ph.D. from Harvard in 1959 for research on photochemistry Butler taught for four years at the University of British Columbia. He then spent several years directing research projects in electrochemistry and environmental science at the Tyco Laboratories in Waltham, Mass., before joining the Harvard faculty in 1971. In the past three years his principal interest has turned toward understanding the pollution of the oceans by petroleum residues. Together with several students he is currently conducting a research project on pelagic tar and its effect on the *Sargassum* community under the sponsorship of the National Science Foundation's Office of the International Decade of Ocean Exploration.

STILLMAN DRAKE ("The Role of Music in Galileo's Experiments") is professor of history at the Institute for the History and Philosophy of Science and Technology at the University of Toronto. After graduating from the University of California at Berkeley in 1932, Drake went to work in the investment field, a career he left to join the Toronto faculty in 1967. This is the third article he has either authored or coauthored for SCIENTIFIC AMERICAN; the first two were "Galileo's Discovery of the Law of Free Fall" in the May 1973 issue and "Galileo's Discovery of the Parabolic Trajectory" (with James MacLachlan) in the March 1975 issue. Concerning the origin of the present article he writes: "I worked this out a couple of years ago, but found it so hard to believe at first that I held it back. Now I have studied all the other notes long enough to be pretty sure that this one was done before the law of free fall was known to Galileo, and that it explains a step in my first article that I then charged off to good luck."

The way we make chlorine saves enough electricity to light the homes of Philadelphia for a year.

Chlorine is chlorine, and for the most part, one supplier's chlorine is just as good as the next guy's.

So since everybody's chlorine is basically created equal, Diamond Shamrock found a better way to make it.

Chlorine is manufactured by supplying salt brine and large amounts of electricity to electrolytic cells. At Diamond Shamrock, we design cells which utilize a Dimensionally Stable Anode (DSA*) made of exotic metals rather than a graphite anode. This Diamond Shamrock innovation allows the manufacturer of chlorine a 15% to 20% reduction in power consumption.

In fact, its use by the North American chlorine industry saves 2½ billion kilowatt-hours per year—more than enough electricity to light the homes of Philadelphia for a year.

At present, only 50% of the chlorine manufacturers use the diaphragm cell with an energy-saving Dimensionally Stable Anode. But if everyone did, think of the electricity you would save. With the energy situation being what it is, the electricity saved will have an impact on our entire national economy.

The way we make chlorine saves more than electricity.

By using the Diamond Shamrock diaphragm cell to make chlorine, not only is electricity saved, but the environment is helped.

The need to open up a Dimensionally Stable Anode Cell for cleaning is less frequent because it operates cleaner than other cells. Therefore, less contaminants are emitted to the atmosphere. Also, chlorinated organic waste is reduced 90%. Other disposable materials used up in the cells are less than half of what was expected in the past.

Since we make chlorine the way we do, it costs less.

Because many U.S. industries depend on Diamond Shamrock for their supply of chlorine and its co-product, caustic soda, Diamond Shamrock recently started up a 1200 ton per day chlorine/caustic soda plant near Houston, Texas—increasing our overall capacity approximately 50%.

This means availability of more chlorine for the manufacture of plastics, synthetic detergents, drugs, dyes, refrigerants, insecticides, weed killers, solvents, as a bleaching agent for paper and textiles, and for killing bacteria in water purification and sewage treatment.

Philadelphia

We can keep it lit for a year.

The way we make chlorine saves 2½ billion kilowatt-hours of electricity.

The additional caustic soda will be used for the manufacture of drugs and dyes, food processing chemicals, petroleum refining, soaps and detergents, and the manufacture of aluminum.

Diamond Shamrock produces both liquid and dry caustic, in various forms and grades.

For information about how we make chlorine and caustic soda, or about Diamond Shamrock, the international corporation that's turning industries around with chemistry and energy, write: Diamond Shamrock, 1100 Superior Avenue, Cleveland, Ohio 44114.

After all, it doesn't take much energy to write.

The way we make chlorine also helps the environment.

*DSA is a registered trademark of Electronor Corporation.

Diamond Shamrock

Established 1845 SCIENTIFIC AMERICAN June 1975 Volume 232 Number 6

Agriculture in China

The most populous nation appears to have achieved the objective of producing enough food for all its people. It has done so largely by the adoption of improved strains of rice and wheat

by Sterling Wortman

Visitors to China in recent years consistently report that the population appears to be healthy and adequately nourished. In the light of China's reputation as a country where hunger has been no stranger for centuries and catastrophic regional famines used to be almost annual events, many Western agriculturalists have reacted to these reports with skepticism. Last year I served as chairman of a group of 12 visitors from the U.S., including some of our foremost agricultural scientists, who were able to judge for themselves how matters stood. We traveled in China for 28 days in August and September.

One major purpose of the visit was to arrange an exchange of germ plasm of Western and Chinese strains of plants. At the same time the group looked forward to learning what it could about contemporary Chinese agricultural practices and to testing by observation the official government position that agricultural production in China is now adequate to the task of feeding the nation's growing population.

In the course of its travels the group met with a number of Chinese agricultural scientists and technicians representing various government agricultural organizations. We visited agricultural communes in the provinces of Kwangtung, Shensi and Kirin and also on the outskirts of two of the three principal cities in China: Peking and Shanghai. We first traveled to Kirin in China's vast wheat-growing northeast (where sorghum, soybeans, millet and maize are also important crops). We flew from there to Sian, the capital of Shensi province in China's western wheat belt, and then traveled by train to Nanking in Kiangsu province, passing through some 600 miles of agricultural land devoted to wheat and rice farming. We ended our visit touring the vicinity of Canton, in the heart of China's multiple-cropping rice country. Wherever we went we found that, compared with farming in other developing nations, Chinese farming generally was going well.

An appreciation of agricultural practices in China is enhanced by the comparison of some equivalent Chinese and U.S. statistical data. The total land area of China (some 973 million hectares, or roughly 2.3 billion acres) is larger than that of the 48 states of the continental U.S. Much of China, however, is too mountainous or too dry for agriculture. Arable land represents only a little more than 15 percent of China's territory, whereas the arable land in the 48 states makes up more than 20 percent of the U.S. total.

When the comparison between the two nations is put in terms of arable land actually under cultivation, it is quite another matter. For example, the estimate of arable hectares actually planted in the U.S. in 1964 is 116 million out of a total of 156 million. That same year China, with a total of only 107 million hectares of arable land, raised crops on 150 million hectares. This seemingly impossible statistic reflects the traditional Chinese practice of multiple cropping: raising two or more crops per year on the same land.

Another significant comparison is the proportion of the total population of China and of the U.S. that engages in agriculture. The population of China is currently estimated at between 850 and 900 million. The number of Chinese engaged in one or another kind of agricultural activity is generally agreed to be from 80 to 85 percent of the total population. Thus estimates of the agricultural labor force in China range from a minimum of 680 million to a maximum of 765 million. (The difference between the high estimate and the low one is greater than the populations of France, Holland and Belgium combined.) In comparison the U.S. agricultural labor force is quite small. It consists of only 2 percent of the total U.S. population of 210 million, or 4.2 million workers.

China's agricultural activities are not distributed evenly across the nation. A substantial part of all farming is done in the plains and deltas of three great river systems: the Yellow River in the north and northwest, the Yangtze in eastern and central China and the Pearl River in the south. The Manchurian plain in the northeast, which includes the three provinces of Kirin, Liaoning and Heilungkiang, is also a major production area. Indeed, most food and fiber production would fall within the region

INTENSIVE LAND USE is a tradition in Chinese agriculture. Here, in the loess belt of central Shensi, crops are grown both in the valley, where water for irrigation is available, and on the terraces cut into the hillsides, which are watered only by rainfall. Wheat is the traditional crop in this area near Sian, but sorghum, maize and fast-maturing varieties of soybeans are also grown.

IRRIGATED RICE FIELDS in the Yangtze River valley below Nanking displayed an exceptionally good crop when the author and his colleagues visited the area in 1974. By introducing high-yield strains China has become the world's leading rice producer.

bounded by 21 degrees and 48 degrees north latitude and 100 degrees east longitude, and the greater part of the farming is confined to the eastern two-thirds of this same north-south band [*see illustration on next two pages*]. In terms of diversity of climatic zones and of crops suited to local conditions this geographical range resembles that found in the U.S. between northern Minnesota and Florida.

A worldwide agricultural practice that is applied in China to an extent unequaled elsewhere is irrigation. About a third of all cultivated cropland in China, or some 33.5 million hectares, is irrigated. By way of comparison fewer than 16 million hectares of the arable land in the U.S., or little more than 10 percent, are irrigated.

China's arable land is a small percentage of the whole: 15.3 percent. Land that is too arid, too high or too poor in soil to be suitable for either grazing or forestry accounts for 18 percent more. Grasslands occupy another 28 percent and forests 8 percent. The missing 30 percent or so in this accounting is land that was once cultivated but has since been abandoned. Originally forest or hilly rangeland, it was farmed piecemeal at various times in the past and then left to erode. Whether many of these lost 300 million hectares can be reclaimed for agriculture or even for high-quality forestry is quite doubtful. Reforestation would, however, at least halt further erosion, runoff and silting, thereby helping to stabilize stream and river flow. Such a reforestation program is reportedly in progress but our group gathered no data on it.

China's principal food grain, in terms not only of area sown but also of tons harvested, is rice. In terms of area sown wheat comes next. Because rice is often double-cropped and almost no wheat land yields two wheat crops a year, the harvest tonnages of the two grains are quite different. Benedict Stavis of Cornell University estimates that in 1971 the Chinese planted 34.2 million hectares to rice and 27.5 million hectares to wheat. According to his estimate, the 1971 paddy-rice harvest (that is, the tonnage before milling removes the husk) was 111 million metric tons and the wheat harvest was only 42 million metric tons.

Other food grains are also important in Chinese agriculture. Estimates of areas sown and tonnages harvested, however, are hard to obtain. Principal among the lesser grains are sorghum (the Chinese kaoliang), millet and maize; barley, oats, rye and buckwheat are also grown. Finally, the Chinese raise such tubers as the white potato, the sweet potato and the yam. Harvest data for the tubers are often lumped with the statistics on food-grain production by reckoning a fourth of the tubers' fresh weight as being equal to the same weight in grain. For example, estimates for 1971 indicate that 51 to 59 million hectares were sown to grains other than rice and wheat and that another 10 to 15 million hectares were planted to tubers. Between 70 and 75 million metric tons of miscellaneous grains and from 24 to 36 million tons of tubers were harvested.

The 1971 estimates give observers of Chinese agriculture a kind of baseline for later statistics. It was after tabulation of the 1971 harvest that the Chinese government announced attainment of agricultural self-sufficiency. The government reported a total 1971 yield of 250 million metric tons of food grains. Estimates by Stavis and by the Food and Agriculture Organization of the United Nations arrive at much the same figure: 247 and 249 million metric tons respectively.

It is common to find conflicting statistics regarding Chinese agricultural production. It is not surprising; most of the available statistics are frankly presented as estimates. Nevertheless, Chinese agricultural activity has unquestionably advanced since the People's Republic came to power in 1949. Much of the increase seems to have been achieved between 1957 and 1971, a period that witnessed a rise of some 65 million metric tons in the production of food grains.

The use of modern high-yield agricultural methods, although applied to only some 20 percent of China's cultivated land, seems to have been responsible for more than two-fifths of the 1957–1971 gain. How did this increase in productivity come about? Extrapolating backward from what we observed in 1974, we find it possible to suggest some (although not all) of the answers to the question.

The 12 members of our group jointly drafted a report of our visit, and this article is largely based on that report. Individual members also observed and reported on their particular fields of interest, and so primary credit for specific comments should be apportioned as follows: rice and the problems of organizing research, Nyle C. Brady, director of the International Rice Research Institute; wheat, forestry and the current status of agriculture, Norman E. Borlaug, director of the International Wheat Improvement Program; soybeans, Richard L. Bernard of the Agricultural Research Service in the U.S. Department of Agriculture; maize and sorghum, George F. Sprague and Jack R. Harlan of the University of Illinois; millet and forage grasses, Glenn W. Burton of the Agricultural Research Service; plant pathology, Arthur Kelman of the University of Wisconsin; vegetable crops, Henry M. Munger of Cornell University, and exchange of genetic resources, Harlan, John L. Creech, director of the U.S. National Arboretum, and the author. During our travels Philip A. Kuhn of the University of Chicago, a historian whose interest is in China from the 17th century to the present, provided us all with valuable insights, as did Alexander P. DeAngelis, a China specialist with the organization that sponsored our visit, the National Academy of Sciences of the United States.

A good starting place for our review is the Chinese program to improve rice yields. Internationally the key "green revolution" advances have been the development of high-yield dwarf strains of the two principal types of rice: "japonica" for temperate climates and "indica" for tropical ones. For example, the indica dwarf known as IR 8 was developed at the International Rice Research Institute in the Philippines. We learned that similar dwarf indica varieties had been developed in China by means of local breeding programs that were initiated in 1956. These Chinese indica dwarfs were put into commercial production in southern China early in the 1960's. Chinese japonica dwarfs had been planted widely in northern China some years earlier. The Chinese use of independently developed varieties of high-yield rice does not mean that China's agricultural scientists are ignoring Western advances in the field. For example, IR 8 was first released by the International Rice Research Institute in 1966. IR 8 seed was brought to China the next year and has now been widely tested (as have many others, including IR 26, the dwarf variety most recently released by the institute).

We were told that by 1965 high-yield indica dwarfs were grown on 3.3 million hectares of southern rice land. By 1973 the area devoted to the high-yield dwarfs had increased to 6.7 million hectares. That is roughly a fifth of all the land sown to rice in China. Moreover, the figure does not include the area sown to high-yield japonica dwarfs in northern China. Wherever we traveled we saw only dwarf varieties of rice, except for some nondwarf plants growing in plots at experiment stations.

Multiple cropping of rice land, a common practice in the south of China, has benefited from these improved varieties. The high-yield dwarfs are quick to mature. Heavier applications of fertilizer have also increased productivity. We found, however, that where triple cropping was practiced it was not customary to plant rice all three times. Instead, particularly in areas where only two crops a year had been raised in the past, the pattern was rice-rice-wheat or rice-rice-barley. Raising of the alternative crop was confined to the months least suited to rice culture.

On balance, we credit the combination of extended irrigation, increased application of fertilizer and the introduction of high-yield varieties of rice as being instrumental in making China by far the largest producer of rice in the world today.

Wheat is China's second most important food grain. New high-yielding varieties of winter wheat have been developed locally, beginning in the early 1960's. The first of the new wheats to be generally sown was a tall-strawed type; by 1965 it was growing on as many as 2.5 million hectares. A variety of dwarf winter wheat with about twice the yield potential of the tall-strawed wheat was first planted in 1972; now two or three other locally bred dwarf varieties are also available. As with rice, the Chinese have not ignored the progress that has been made by Western agricultural scientists. Quantities of seed from several varieties of dwarf spring wheat, developed by Mexican scientists and by the International Wheat Improvement Program staff in Mexico, have been purchased by the Chinese. In 1973 5,000 tons of Mexican seed were imported. In 1974 the total was 15,000 tons.

Late August and early September is not the time to survey wheat crops; we saw no wheat (nor any barley, oats or rye) growing during our visit. The 1974 winter and spring wheat had already been harvested, and the planting of winter wheat does not begin until the end of September and early October.

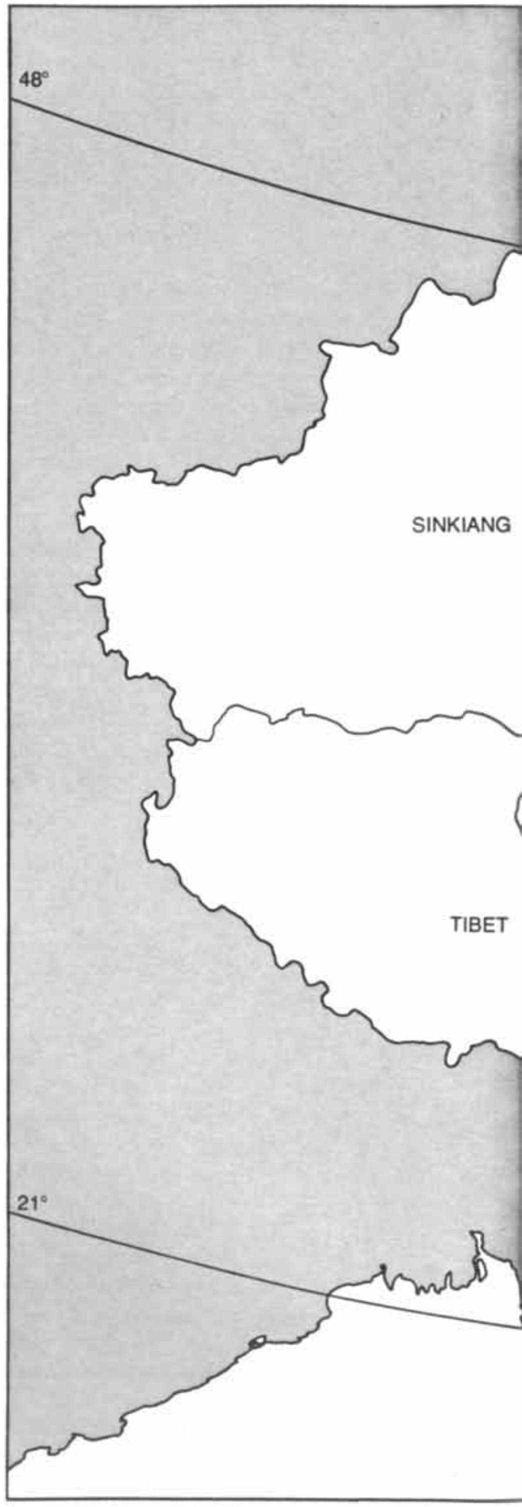
To increase wheat yields the Chinese not only are introducing new varieties but also are extending southward the zones where both winter wheat and spring wheat are raised. In some of the new areas winter wheat is grown in the interval following harvest of the late rice crop and preceding the planting of the early rice crop. The success of the program in these areas will depend on the use of quick-maturing varieties of rice and winter wheat and the application of increased amounts of fertilizer. In pushing spring-wheat culture southward the Chinese are sowing spring wheats in the fall in areas where little or no wheat was formerly grown. The use of early-maturing, high-yielding varieties of spring wheat from Mexico has made possible this expansion.

In the northeastern provinces the spring-wheat harvest comes too late to allow any but very modest efforts to follow it with a second crop, such as vegetables or buckwheat. Most of the wheat land lies fallow until the next spring. During the past two years, however, promising experimental results have come from planting rows of maize between rows of spring wheat about a month after the wheat has been planted. Mechanical wheat harvesting is not possible when that is done, but for the Chinese, with their labor-intensive approach to agriculture, it is not a serious obstacle.

Soybeans are grown throughout China but are considered a major commercial crop only in the three provinces of the northeast. There they are planted in late April or May and harvested in September and October. Because of the shortness of the northeastern growing season, soybeans are not double-cropped. In the fields we observed in Kirin, however, the plantings consisted of alternate strips of soybeans, six to eight rows wide, and much wider strips of maize. Irrigation was rarely in evidence, harvesting was done with hand sickles and machines served only for threshing. Most of the harvest is sold to the state and is used to make the traditional soy products—bean curd and soy sauce—that represent an important high-protein part of the Chinese diet.

Elsewhere in China soybeans serve a fill-in function, being seeded on the banks of ditches and streams, roadsides, railroad rights of way, field margins and any odd piece of ground too small or too steep for a commercial crop. They are also intercropped in orchards and are the predominant species in many farmers' individual gardens. Some soybeans go to market as green pods and some serve as livestock feed, but most of the fill-in crop is consumed by the growers themselves.

We observed soybean-breeding programs being undertaken both at the Kirin Academy of Agricultural Sciences and at the Northwest College of Agriculture near Sian. Some 80 percent of the soybeans grown in Kirin are improved varieties that were developed by the academy. At the Northwest College of Agriculture a five-year breeding program had culminated in the development of a new soybean variety in 1970. The improved strain matured in 105 to 110 days and thus could be successfully grown as a second crop following the winter-wheat harvest in mid-June. This has now become a general practice in central Shensi.



AGRICULTURE in China is almost entirely confined to the rectangle (*color*) bounded

The use of hybrid varieties of maize and sorghum has contributed significantly to the increasing yields of these crops during recent years. For example, we were told that in Kirin between 60 and 70 percent of the area used to grow maize was sown to hybrid varieties. Maize is by no means a newcomer to China, having first been introduced some centuries ago. Hybrid maize now predominates, particularly in northern and central China; some of the hybrids are derived from older U.S. inbred lines.

Maize planting, cultivation and harvesting is done by hand. The stalks are cut at ground level; the ears are husked and hauled to a drying floor, and the stalks are stored for winter feed. After partial drying the ears are shelled and the grain is further dried until its moisture content drops to a level suitable for

on the west by 100 degrees east longitude. Most of the country's food is produced in the eastern two-thirds of the rectangle. From north to south the climatic zones in China are roughly equivalent to the zonation from northern Minnesota to Florida in the U.S.

storage. Maize is not a particularly popular foodstuff among the Chinese; in the northeast at least, where the winters are long and severe, some of the crop is used as pig feed.

Sorghum, a grass that was first domesticated in Africa, entered China long enough ago to evolve several subraces that are not found elsewhere except in Japan and Korea. Today these original sorghums have been almost completely replaced by hybrids. The hybrids under cultivation in Kirin were different from those in Shensi and had been developed by local agricultural experimenters. In Kirin sorghum is eaten more or less like rice. The seed is "pearled" to remove its brown coat, the bran is fed to pigs and the pearled grain is boiled until it is tender. In Shensi, where the cereal tradition is based on wheat, the whole sorghum seed is ground into flour and baked. Although the northeast is the traditional sorghum belt, both there and in Shensi the crop is steadily being replaced by maize; in the northeast today sorghum and maize occupy about equal amounts of arable land.

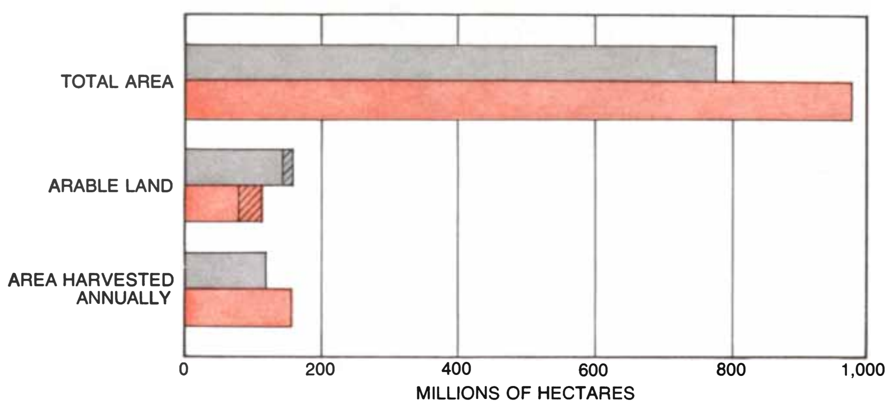
Millet was the first food grain to be grown in China, beginning more than 6,000 years ago. Its original wild progenitor can still be found in Shensi, and the species of millet that is grown widely in China today, *Setaria italica*, may well have evolved in Shensi. A grain high in protein (12 percent of dry weight), it makes an excellent golden porridge when it is boiled; it is easy to understand why millet is the preferred cereal in Kirin even though rice, wheat, sorghum and maize are available.

The principal advantage of millet as a crop is that it can be raised successfully in semiarid regions where occasional rainfall is the only source of water and other grains will not thrive. Moreover, its straw is rated ahead of other cereal straws as a feed for livestock. As a result millet is often sown broadcast and is cut before it reaches maturity to produce a high-quality fodder. In semiarid parts of northern and northeastern China, where irrigation is not feasible at least for the present, and where some cattle, sheep and goats are raised, the future of millet as a food and forage grass seems assured.

With respect to domestic animals, the distribution of livestock other than pigs and poultry in China is significantly regional. For example, almost all sheep and goats are found only in the mountainous areas of China's western provinces and in the extensive grasslands of Inner Mongolia to the north and Sinkiang to the northwest. Except for a few dairy herds in the suburbs of major urban areas, cattle are similarly confined to the west and the frontier. Horses, mules, donkeys and water buffaloes, all maintained almost exclusively as draft animals, are virtually the only large mammals found in agricultural China [*see illustration on page 21*].

A domesticated animal that is of paramount importance throughout China, however, is the pig. In 1972, according to a conservative U.S. Department of Agriculture estimate, there were nearly 260 million pigs in China. This is approximately four times the number in the U.S., a nation that prides itself on its pork production. A comparison of the two countries' production methods is illuminating. In the U.S. the objective is to produce a 100-kilogram pig in the shortest possible time, usually less than six months. The young pig is fed a balanced ration of maize and soybeans, supplemented with minerals, vitamins and antibiotics. In China pigs are fed mainly the by-products of agriculture: the weeds and water plants collected by farmers, the pods and leaves of soybeans, peanuts and rape, cornhusks, corncobs, cabbage and sweet-potato leaves and the bran from rice, wheat, sorghum and millet. Since many of these plant products are fibrous and difficult to digest they are usually chopped up, soaked and fermented before being used as feed. Needless to say, pigs gain weight more slowly on this kind of diet. They usually come to market at eight to 12 months of age, weighing between 70 and 100 kilograms; the state will not buy pigs that weigh less than 50 kilograms.

Swine husbandry is mainly an individual enterprise. The breeding stock is maintained by the commune, and the piglets are provided to individual households, generally two to a family. Hand-raising a pig or two not only guarantees the farmer some pork in his diet but also adds to the family income. Perhaps above all, the pigs are valued for the manure they produce; it is one of the key ingredients in the organic compost that still makes up a large part of China's fertilizer.



LAND AREAS available for agriculture in China (*color*) and the U.S. (*gray*) are roughly comparable. The 48 contiguous states of the U.S. are smaller in area than China by some 200 million hectares (*top bars*), but the U.S. area includes nearly 50 million more hectares of arable land (*middle bars*). Not all U.S. arable land is cropped each year, however, whereas some Chinese arable land yields two or three crops per year. As a result nearly 35 million more hectares are harvested in China each year than in the U.S. (*bottom bars*). Shading on middle bars shows irrigated land: 33.5 million hectares are irrigated in China.

Having reviewed China's principal crops and noted the gains in productivity that are attributable to the introduction of higher-yielding varieties of plants, it is now appropriate to examine other factors that have affected and will affect agricultural output. Perhaps foremost among these factors is an increase in the use of fertilizers. The amount of fertilizer applied in China today is small by American standards. By Chinese standards, however, recent increases in use of fertilizer have been substantial.

The value of organic fertilizers has been recognized in China for centuries. Crop wastes such as the straw from rice and wheat and the stalks from sorghum and maize are combined with animal feces (particularly pig and human feces) and urine in pits or compost piles and allowed to ferment and decompose. Before 1960 soil fertility in China was almost entirely dependent on the return of this organic manure to the fields, on the sowing of nitrogenous ("green manure") crops and, where conditions allowed, on spreading the fields with silt dredged from lakes and waterways.

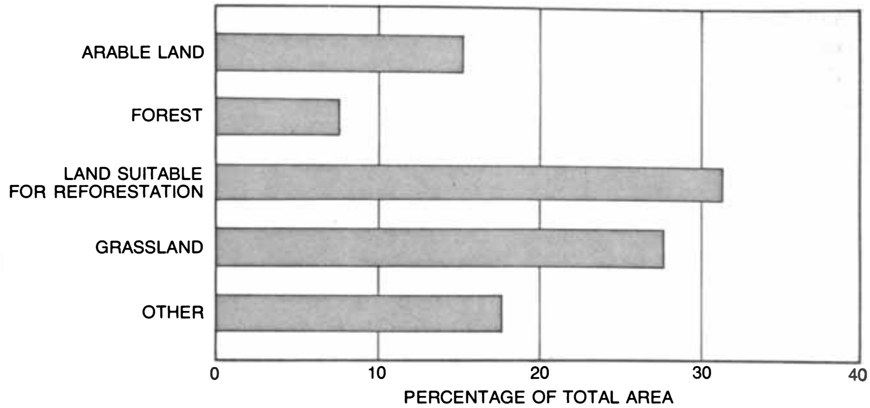
Over the past several years the Chinese have supplemented the use of traditional organic fertilizers with the application of increasing quantities of chemical fertilizers. Some of the chemical fertilizers have been produced in China and some have been imported. Indeed, China has recently become the world's largest importer of nitrogenous fertilizers: purchases abroad in 1973 came to 1.8 million tons. Of the nitrogenous fertilizer produced at home nearly 50 percent comes from small factories that make ammonium bicarbonate, which is 17 percent nitrogen, from lignite, or brown coal.

Ammonium bicarbonate is a compound that is rarely if ever used as fertilizer in the U.S. Nonetheless, Stavis estimates that in 1971 some 7.2 million metric tons out of a total of 16.9 million metric tons of fertilizer manufactured in China was ammonium bicarbonate. In 1972 the proportion was 9.9 million metric tons out of a total of 19.9 million.

Brady and Borlaug visited an ammonium bicarbonate plant near Shanghai. Operations had begun in 1959, and annual production was then some 800 tons of ammonia. At the time of their visit 15 years later the plant was producing 20,000 tons per year; 80 percent of the output was being converted to ammonium bicarbonate, a fine white powder. The remainder was distributed as a fluid, aqua ammonia, with a nitrogen content of 42.5 percent. The aqua ammonia is transported by river and canal in boats fitted with concrete tanks; it is applied directly to the soil by adding it to irrigation water.

Other fertilizers manufactured in China include phosphates (made in both large and small factories) and additional nitrogenous fertilizers made in plants much larger in scale than the one in Shanghai. Comparing the total domestic production from all sources in 1972 (just short of 20 million metric tons) with estimates of production in 1964 (5.9 million metric tons), it is evident that in the eight-year period the supply of domestic chemical fertilizers has increased more than threefold.

The Chinese government has recently signed contracts for the installation of 10 or more plants designed to produce anhydrous ammonia; each plant has a daily capacity of 1,000 tons. That amounts to a potential annual production of at least 3.5 million tons. Each of the ammonia plants will be tied into a urea plant, where most of the anhydrous ammonia will be converted into solid nitrogenous fertilizer. It is projected that by 1978 the complex will yield 2.7 million tons of nutrient nitrogen; the feedstocks will come primarily from China's own supply of crude oil and natural gas. Although the capital investment, including the cost of pipelines and rolling stock, may go as high as the equivalent of $1 billion, China will have effectively insulated itself from the rising cost of fertilizer imports.



LAND UTILIZATION in China includes endeavors other than agriculture. Grasslands alone are almost twice the size of the arable land. Only some 20 percent of the land is so dry or mountainous as to support no significant plant life. Forests cover some 8 percent of the land. The largest land category, more than 30 percent of the total, is old grassland or forest that was stripped of timber and is now eroded. Some reforestation is in progress. These figures, based on estimates for 1963, are taken from a 1972 study by Chao Shih-ying.

Another factor that has contributed to China's expanded agricultural output is the improvement and enlargement of irrigation systems. As I have mentioned, China has long been known for the high proportion of its arable land that is irrigated. Recent years have seen construction of many new diversion canals, catchment basins and wells equipped with mechanical pumps. We were given no official data on the extent of the increase in irrigated lands. One informal estimate, however, is that since 1965 five million hectares of irrigated land have been added to the previous total of some 33 million. Much of the effort to improve and enlarge existing systems is reportedly expended during winter months, when the demand for agricultural field labor is relatively slack.

One additional technological factor (which I have mentioned in passing with respect to individual crops) deserves summary here. That is the contribution made to increased productivity by the practice of multiple cropping. It is rare to see land lying fallow in any part of China where temperature and rainfall allow plant growth. In many areas that previously grew only two annual crops there are now three, for example rice-rice-wheat and rice-rice-barley. In other areas two crops are now raised where only one grew before, for example wheat-maize, wheat-sorghum and wheat-cotton. Furthermore, many combinations of intercropping and mixed cropping are under evaluation, and at the same time that larger amounts of fertilizer are being applied the density of plántings is being increased. The overall rise in intensive cropping has been made possible by the development and planting of high-yielding, quick-maturing varieties. With the increased application of fertilizers additional progress can be anticipated.

Another key element in China's agricultural endeavors is the human factor. It seemed to our group that, concerning both China's agricultural technicians and China's farmers, two separate questions arise. The first is just how effectively the government bureaucracy influences the activities of the 80 to 85 percent of the population who are engaged in agriculture. The second is a question that is quite as pressing elsewhere in the world as it is in China: Can the national birthrate be lowered before population growth cancels out the benefits of increased agricultural production?

On the question of bureaucratic control, in former times the lowest-level provincial officials directly answerable to the central government were the *hsien* magistrates (in our terms county administrators). Today the network of central control descends below the county level to the level of the traditional Chinese rural marketing community: the small town and its satellite villages. This unit, in current terminology, is the commune. Commune administrators are national cadres, that is, they are on the state payroll. Each commune includes a number of production brigades, which may be considered equivalent to groups of villages, each production brigade includes a number of production teams and each production team includes a number of individual households.

It is the leader of the production team who exercises the most direct influence over the individual farmer's life. The production team is the accounting unit that computes the "work points" earned by each individual and thus determines his share of annual income. At the same time provision of basic services at the brigade level, such as primary schooling and health care, means that each farmer and farm-family member interacts with this wider community almost on a daily basis. In the same way brigade leaders regularly interact with the local representative of central authority: the commune administrator.

Where agriculture is concerned this network of communication and accountability is an efficient instrument for the introduction and adoption of new technology. For example, each provincial agricultural research unit is integrated with the provincial farm population by means of what are called basic points. These are small agricultural experiment stations attached to selected production brigades throughout a province. Brigades selected under the system exemplify some typical environmental aspect of the province: a given kind of soil, a dearth or abundance of rainfall and so on. Each such station is manned by scientists and technicians "sent down" from their provincial research center and by farmers drafted from brigade production teams for training in experimental work; thus each province contains a network of basic points. We observed the network-and-points agricultural-experiment method at work in two provinces, Kirin and Shensi.

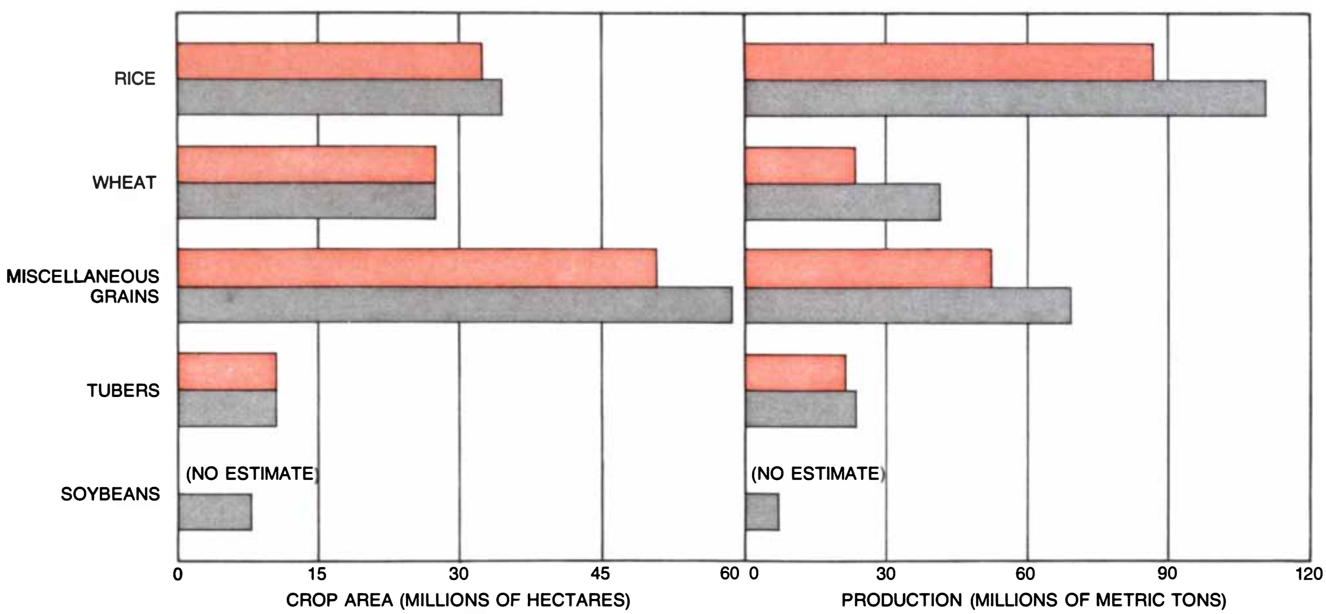
In Kirin the provincial Academy of Sciences of Agriculture and Forestry has a staff of 255 scientific and technical workers. Academy personnel work at experiment stations attached to 29 provincial production brigades. At any one time perhaps 80 members of the academy staff are manning the provincial network of farm-level experiment stations and another 80 are engaged in extension work involving communes that do not have experiment stations. Similarly, in Shensi personnel of the provincial agricultural institute reportedly work at 100 locations in 75 counties of the province.

The main theme of the experimental work at the brigade level is to develop improved systems of crop or livestock production suited to each particular environment. In addition some of the activity is not experimental but what we would call demonstrational. For example, improved practices are displayed in parallel with traditional ones.

The academicians who take turns participating in these brigade-level agricultural experiments are careful to stay in the background. (So are the other sent-down intellectuals, who devote the time they are required to spend in the countryside to promoting literacy among the farmers, posting wall newspapers and guiding political discussions.) Nonetheless, the academic agriculturists play a major role in introducing the farmers to progressive practices, just as American county agents did in the earlier decades of this century.

One cannot quarrel with the short-term practicality of this combination of applied research and demonstration work in China. It has obviously been effective in raising crop yields in all the areas we visited. The approach is not, however, free of long-term hazards. The interruption of basic research and the decentralization of experimental studies imposed by such a system cannot help but weaken the potential for future scientific advances in agriculture. Whether the Chinese will support their farm-level efforts with more basic research remains to be seen.

If China's present living standards are merely to be maintained, let alone improved further, the nation must avoid continuous population growth. An annual increase of no more than 2 percent means at least 18 million new mouths to feed, which requires an extra five million tons of food grains. There is good evidence that the birthrate in China's urban areas is beginning to decline sharply in response to government pressures [see "The Delivery of Health Care in China," by Victor W. and Ruth Sidel; SCIENTIFIC AMERICAN, April, 1974]. The size of China's urban population, however, is trivial compared with the size of the rural population. The key group that must be won over to a prompt and effective stabilization program is the farm labor force. That may already have been



SELF-SUFFICIENCY in agricultural production was announced by the Chinese with the harvest of 1971 (*gray*), when total grain production approached 250 million metric tons. Estimates of the area sown to various major crops and the amounts harvested are compared here with similar estimates for 1957 (*color*). Miscellaneous category includes millet, sorghum, maize, barley, buckwheat, oats and rye. Tubers are mainly potatoes. Most data are from the study by Benedict Stavis of Cornell University in 1974.

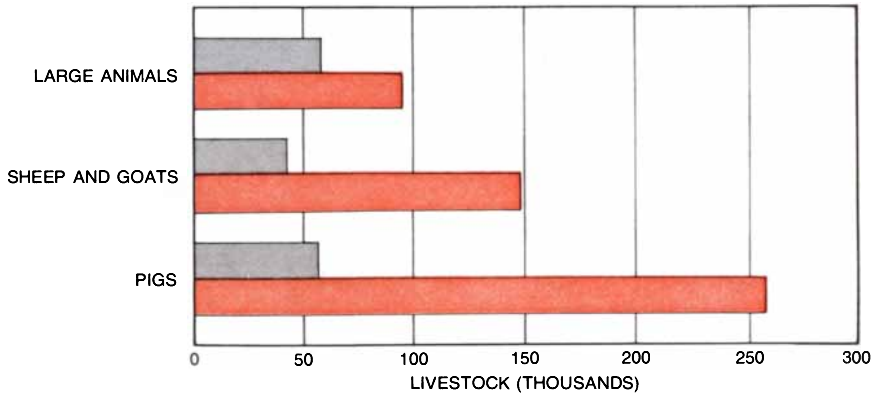
accomplished as a result of the establishment of certain new rural practices.

The first of these is related to household economics at the production-team level. Individual household members receive a share (in money and produce) of the annual earnings of their team; the share is proportional to the number of work points each has earned during the year. For example, a 70-year-old grandmother in a suburban Shanghai commune told me that she and nine of the other 10 members of the household worked regularly in the fields. The only nonproducer was a five-year-old child. The family received a monthly advance of the equivalent of $80 against its joint annual earnings. In 1973 its share of the team's income after operating expenses and taxes were deducted was $2,100, or about $200 per person.

Withholding for taxes and expenses and to provide funds for reinvestment absorbs about half of the gross income of a commune. For example, the manager of the suburban Rainbow Bridge Commune in Shanghai told us that 30 percent of the commune's gross income was required to meet the cost of its agricultural endeavors. Another 11 percent went into a fund for capital reinvestment, 4 percent was paid to the government as tax and 3 percent went into a welfare fund for the support of new mothers and the elderly. That left 52 percent of the gross income for distribution to individuals in proportion to the work points earned.

Such a work-point system simultaneously accomplishes two objectives. It provides a strong incentive for increased overall output in order to maximize the amount that is divided among the individuals in a household. At the same time it encourages householders and production teams alike to keep their numbers from increasing, so that the maximum earnings are divided into the fewest possible shares. The latter constraint, of course, places the disadvantages of population growth squarely before each individual. This economic consideration must generate substantial social pressure in favor of small families.

Other growth-limiting factors are at work. As one example, it is said that improved health care has greatly reduced infant mortality, thus undercutting the validity of the traditional view that many births are needed to ensure the survival of even one offspring to maturity. Another example is the current program of providing care for the aged, one of the objectives of the 3 percent welfare withholding reported by the Shanghai commune. We visited one such residence for the elderly, maintained by a commune near Peking. Normally, we were told, each family cared for its own elders. Nonetheless, some individuals inevitably lost their families or became separated from them; that was the case with the 75 residents we saw. They were encouraged to engage in handicrafts; meanwhile they were housed, clothed and fed at commune expense, and a clinic attached to the residence looked after their health needs. Custodial care of this kind, if it were widely practiced, would diminish the validity of another traditional view that favors multiple births: the view that one needs to have many children in order to ensure one's welfare in old age.



LIVESTOCK IN CHINA are compared according to categories; estimates, made by the U.S. Department of Agriculture, are for 1949 (*gray*) and 1972 (*color*). Sheep and goats are largely confined to grasslands and mountains in the west and northwest. That is also true of range cattle, lumped together here with large draft animals: horses, mules, donkeys and water buffaloes. Pigs are by far the most numerous of all the domestic mammals in China.

A third growth-limiting factor that may even now be undergoing a crucial test is the continued existence of private landholdings throughout rural China. We were told that when the commune system was established in the late 1950's, between 5 and 7 percent of the arable land in each commune was set aside for private use. The land was parceled out at a fixed rate: each adult in a household was assigned one 150th of a hectare. The household children received private land too, but only the two eldest children in each household were eligible. If a household had more than two children at the time of the land assignment, or if more children arrived later, those children received no share. In such a family there would simply be less private land per person. In its early days the allocation system probably had relatively little effect on rural family planning. Today, nearly a generation later, the system must act as a powerful social force favoring small families.

In summary, with respect to agricultural production in China now and in the near future our group was inclined to accept the government's assertion that self-sufficiency was achieved in 1971, when some 250 million metric tons of rice, wheat and other major foodstuffs were harvested. Furthermore, given a continuation of the present aggressive and coordinated effort it seems that China will be able to achieve substantial increases in agricultural production over the next decade. Whether the increase for any particular crop will be as little as 20 percent or as much as 50 percent will probably prove to be a function of present yields. For example, rice yields are already high, so that an increase of 50 percent in the annual harvest will be much harder to achieve than a 50 percent increase in the maize or sorghum harvest.

China nonetheless faces serious long-term problems. If the country's remarkable agricultural advance is to continue, two of these problems must soon be resolved. The first problem lies in the current Chinese policy that emphasizes applied and decentralized agricultural research for the sake of immediate increases in food production. Although the short-term benefits of this policy are obviously important, the policy must be complemented in the near future by similar emphasis on more basic scientific investigation.

The second problem has to do with the government's remarkably comprehensive efforts to retard the growth of China's population. Only if this policy meets with success will the Chinese people continue to receive the benefits of increased agricultural production and the accompanying advances in living standards that so many of China's people now seem to expect.

How the Liver Metabolizes Foreign Substances

Among the most significant of the liver's chemical transformations are the inactivation of drugs, the detoxification of environmental pollutants and the activation of chemicals that can cause cancer

by Attallah Kappas and Alvito P. Alvares

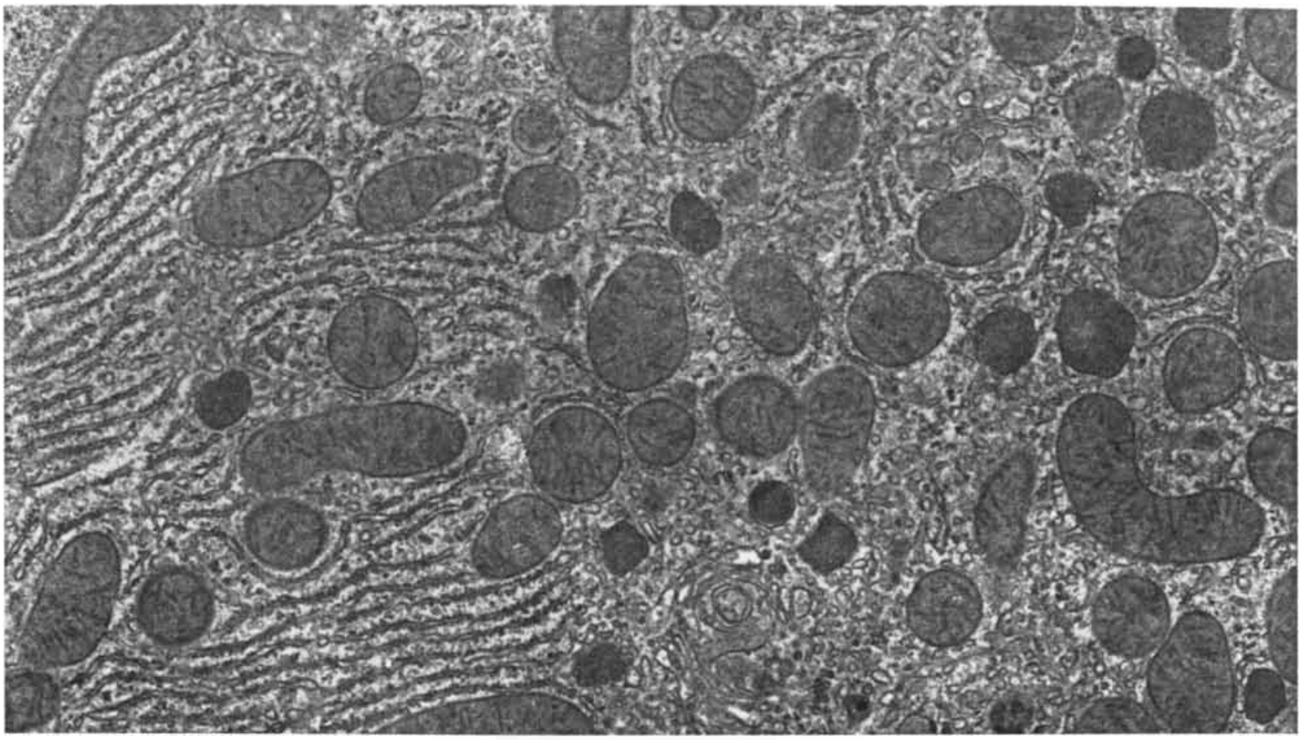
The intensity and duration of the action of most drugs is determined in large part by their rate of metabolism. If nothing else happened to a drug after it entered the body and reached its target organ, for example, it might continue to act indefinitely. Something does happen, however: most drugs are transformed into inactive substances and then excreted. The biotransformation can occur in any of several tissues and organs. Some drugs are transformed chemically in the intestine, some in the lung, the kidney or the skin. By far the greatest number of these chemical reactions are carried out in the liver, which metabolizes not only drugs but also most of the other foreign chemicals to which the body is exposed. Biotransformation in the liver is therefore a critical factor not only in drug therapy but also in defending the body against the toxic effects of a wide variety of environmental chemicals such as insecticides, herbicides, dyes, food preservatives and a number of substances that are suspected of inducing cancer. The central step in the metabolism of most of these agents involves an oxidation reaction mediated by a complex of enzymes that has come under intensive study in recent years in our laboratory at the Rockefeller University Hospital and in other laboratories.

The liver is the largest organ in the body (it weighs about three pounds in an adult) and has diverse functions. It serves, first of all, as the primary receiving depot, chemical-processing plant and distribution center for almost everything that enters the body through the walls of the alimentary canal. All the blood that has absorbed digested food and other substances from the intestines enters the liver through the large portal vein, which ramifies into fine channels through which the blood perfuses slowly among the liver cells. Here nutrients and other foreign substances are removed, metabolized, in some cases stored and then released into the general circulation. Amino acids, for example, are made into proteins and other nitrogenous compounds; glucose is converted into glycogen and stored, to be converted back into glucose and released as required. And drugs and other toxic substances are detoxified. Not everything is metabolized on the first passage of blood through the liver, of course; drugs, for example, are given in doses such that a sufficient amount of the drug moves through the liver to its site of action and is transformed later, on return visits to the liver [*see illustration on page 25*]. The liver also produces bile, which is a secretion that aids in the digestion of fats when it is released into the small intestine and is also a vehicle for the excretion of transformed substances and other waste products of metabolism.

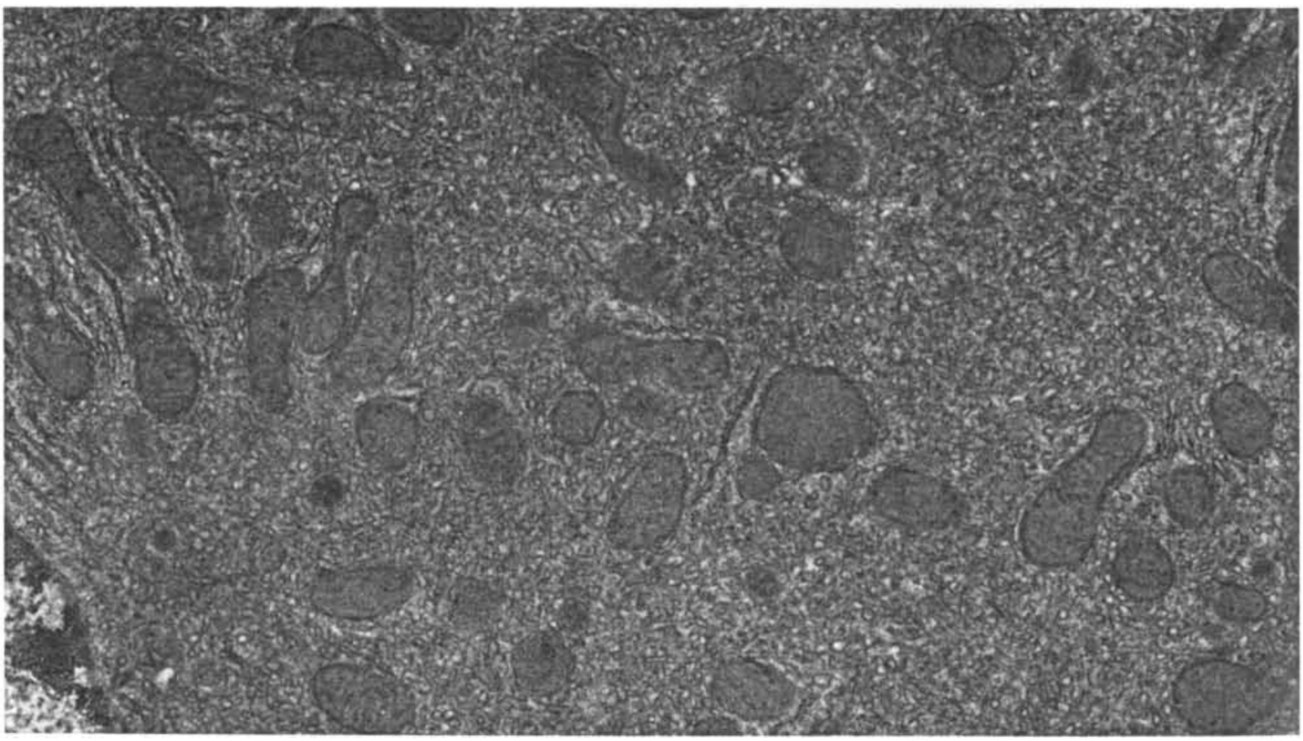
The biotransformation of drugs and other foreign compounds in the liver is accomplished by several remarkable enzyme systems that can metabolize a wide variety of structurally unrelated drugs, toxic agents and environmental pollutants, which enter the body primarily through ingestion but also through the lungs and the skin. The enzyme systems are built into the membranes of the endoplasmic reticulum of the liver cells, a network of interconnected channels that is present in the cytoplasm of most animal cells. There are two kinds of endoplasmic reticulum, rough and smooth, and they differ in both form and function. The surfaces of the rough membranes are studded with ribosomes, small granules that translate the genetic code into the sequences of amino acids that constitute proteins. The smooth membranes have no ribosomes. In the liver a major function of both kinds of membrane is to assemble the enzymatic complexes that transform foreign substances and then to serve as the site of those transformations. Unlike some other cellular subsystems, the endoplasmic reticulum cannot be separated from cells as an intact structure. If liver cells are homogenized and then centrifuged, the tubular reticulum breaks up and bits of the membranes are sealed off to form the tiny vesicles, or sacs, called microsomes. The microsomal fraction thus obtained from liver cells is a convenient natural source of enzymes for laboratory studies of liver-cell metabolism.

Drugs and other foreign compounds are metabolized in the liver by a rather small number of reactions: oxidations, reductions, hydrolyses and conjugations. Their essential effect is to convert lipophilic, or fat-soluble, compounds into hydrophilic, or water-soluble, ones. The hydrophilic compounds are the more readily removed from the blood by the kidneys and excreted.

Oxidation accounts for most of the transformations, largely because there are so many different ways in which a compound can be oxidized [*see illustration on pages 28 and 29*]. The alkyl side chains of barbiturates and some other drugs, for example, are oxidized to form alcohols. In the case of compounds incorporating aromatic rings, including polycyclic hydrocarbons (such as those in cigarette smoke) and many drugs, a hydroxyl group is inserted into the ring.



ENDOPLASMIC RETICULUM of a rat liver cell is enlarged 18,000 diameters in this electron micrograph made by Edward S. Reynolds of the Harvard Medical School. Most of the biotransformations of foreign substances take place in this system of membranous tubules. The "rough" endoplasmic reticulum is covered with ribosomes, the structures in which proteins are synthesized; it appears here as fairly linear double membranes studded with black dots. The "smooth" endoplasmic reticulum lacks ribosomes and forms a more branching, tubular network; there are patches of it between the mitochondria, the fine-grained gray objects, at right.

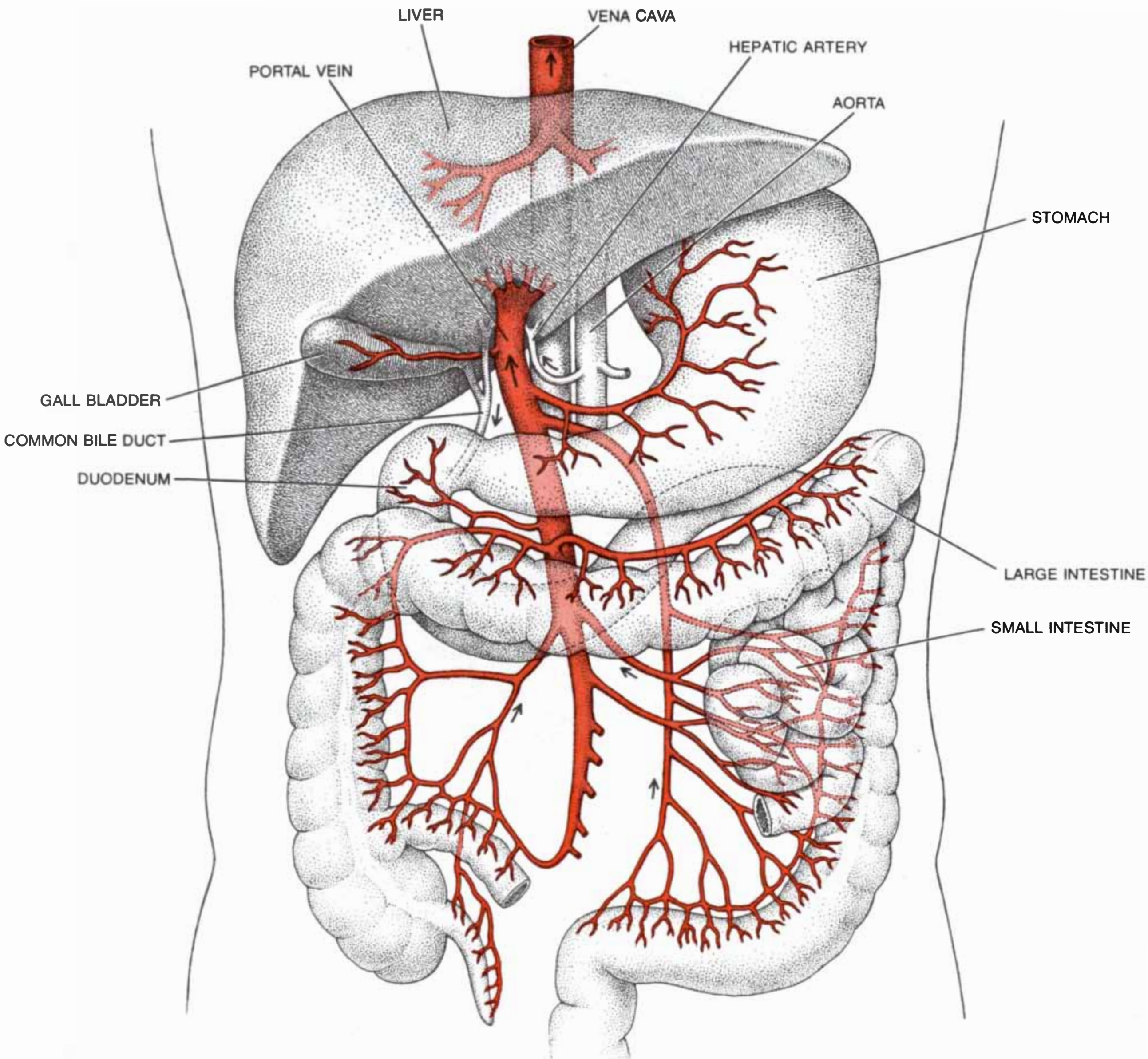


SMOOTH ENDOPLASMIC RETICULUM proliferates when the synthesis of the enzyme systems it contains is stimulated by drugs. This micrograph, also made by Reynolds and reproduced at the same scale as the one at the top of the page, is of a cell from the liver of a rat that was treated with phenobarbital. The drug stimulates the synthesis of cytochrome *P*-450, an enzyme located in the endoplasmic reticulum membranes. The cytoplasm of the cell has expanded because of a striking increase in the smooth endoplasmic reticulum membranes, which now fill most of the space between the more dispersed mitochondria. Enzyme induction by one drug can significantly affect the metabolism, and thus the activity, of other drugs that are metabolized by the same enzymes.

In other cases alkyl groups are removed from either nitrogen or oxygen atoms, amino groups are removed or sulfoxides are formed. Reduction and hydrolysis are also catalyzed by liver enzymes, but these reactions are less common than oxidation.

Conjugation of a chemical is combination with some natural constituent of the body such as the glucose derivative glucuronic acid, the amino acid glycine or the tripeptide glutathione. In the presence of the appropriate enzyme these natural agents can combine readily with compounds that have carboxyl (COOH), sulfhydryl (SH), amino (NH_2) or hydroxyl (OH) groups. Some drugs have these groups when they are in their active form and are handled in the liver by conjugation; for most drugs, however, conjugation is a second step that comes after metabolism by oxidation, reduction or hydrolysis. Almost without exception the conjugated compound is devoid of pharmacological (or any biological) activity.

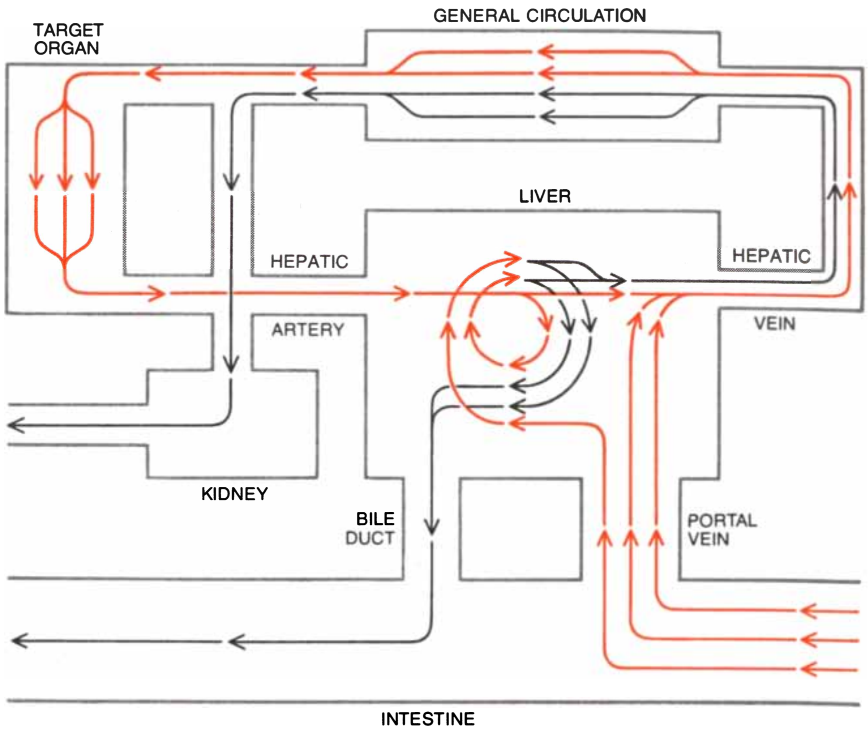
Clearly the enzyme systems that catalyze these reactions were not invented by the mammalian organism in order to cope specifically with drugs or novel pollutants. Their basic physiological role has presumably been to metabolize endogenous substrates: substances normally present in the body. For example, liver microsomal enzymes oxidize steroid hormones, cholesterol and fatty acids. The products of these oxidations may then be conjugated (with glucuronic acid, for instance) and excreted. Bili-



LIVER is a primary site of metabolism of substances entering the body through the alimentary tract. All the fine blood vessels that absorb nutrients and other substances through the wall of the intestines come together and enter the liver through the portal vein. The liver's supply of oxygenated blood from the heart enters through the hepatic artery. After passing through the sinusoids of the liver (*see illustration on page 26*) the blood is collected by the hepatic veins, which feed into the vena cava. The liver secretes bile, which is collected by the bile duct, stored in the gall bladder and emptied into the duodenum, the first segment of the small intestine.

rubin, a product of the oxidation of heme, the red pigment of hemoglobin, is an example. It is normally prepared for excretion by glucuronide conjugation. The rate of formation of glucuronides is generally low in newborn infants, however, because of a deficiency of the enzyme glucuronyl transferase. As a result bilirubin may not be conjugated and excreted at an adequate rate. Excessive amounts of it can then accumulate and cause grave damage to the brain, a condition called kernicterus.

It has been apparent for some time that the human fetus and the newborn infant are far more sensitive than adults to many drugs. A number of drugs can pass across the placenta, so that obstetricians need to exert care in administering them to an expectant mother. Barbiturates or morphine given to a woman during childbirth can be stored in the infant's tissues and cause respiratory depression and occasionally death. The explanation for the sensitivity of infants to drugs has emerged from a number of reports in recent years on the maturation of the capacity to metabolize and conjugate drugs. These studies make it clear that the capacity to oxidize and conjugate is negligible in the mammalian fetus and newborn animal and increases after birth at a rate that varies with the species, the type of reaction and the drug. Impaired drug metabolism increases the intensity and duration of drug action. For example, newborn mice treated with a dose of the hypnotic drug hexobarbital equivalent to 10 milligrams per kilogram of body weight sleep more than six hours, whereas adult mice given 10 times as large a dose sleep less than an hour.



PATHWAY of a drug that is transformed in the liver is shown schematically. The drug (*colored arrows*) enters the liver through the large portal vein, passes into the general circulation, has its effect on the target organ and eventually returns to the liver. On each passage through the liver a fraction of the drug is converted, usually into inactive metabolites (*black arrows*). The metabolites may be carried by the bile into the intestines for excretion or may pass through the circulation to the kidneys, to be excreted into the urine.

The adverse effect of drugs on newborn infants as a result of inefficient metabolic conversion is illustrated by the striking "gray-baby syndrome" in infants that may come a few days after treatment with the antibiotic chloramphenicol: abdominal distention, respiratory difficulty, cyanosis (blue skin color as a result of insufficient oxygenation of the blood) and shock. The condition is apparently the result of deficient metabolism of the antibiotic and deficient conjugation with glucuronic acid. In adults about 90 percent of the antibiotic is excreted in the urine in the form of conjugated metabolites within 24 hours after it has been orally administered; only a small amount is excreted unchanged. In comparison a 10-day-old infant in one study excreted less than 50 percent of the drug in 24 hours.

Rates of drug metabolism are very different in different species, and the effective dosage varies accordingly. The anti-inflammatory agent phenylbutazone is metabolized slowly in man; its half-life in the plasma averages about three days. In the horse, the dog, the rabbit, the rat and the guinea pig, however, the drug is metabolized much more quickly and the half-life ranges from three to six hours. (Knowledge of the rapid rate of metabolism of phenylbutazone in the horse is of practical importance because the drug is administered to treat arthritic conditions in racehorses.) A dose of hexobarbital (adjusted for the body weight of the animal) that makes mice sleep for an average of 12 minutes puts rabbits to sleep for 49 minutes, rats for 90 minutes and dogs for 315 minutes. When the enzymatic oxidation of hexobarbital by microsomes from the liver of these animals was measured, the fastest oxidation was carried out, as expected, by microsomes from the mouse; the rates of oxidation by microsomes from the rabbit, the rat and the dog were proportionately lower. Differences in the rates and patterns of drug metabolism in the various animals seem to explain most of the species differences in effect, but there may also be differences in the distribution of the drug, the response of the target tissues and excretion.

Even among human patients there are marked individual variations in the metabolism of drugs that are handled primarily by microsomal enzymes. The variability causes some patients to metabolize a drug so quickly that therapeutically effective blood and tissue levels are difficult to achieve and others to metabolize the drug so slowly that they suffer toxic effects. It can therefore be difficult for the physician to predict just what dosage of some drugs will provide a safe and therapeutic effect in an individual patient. Very large individual differences have been noted in the metabolism of the coumarin-derivative anticoagulant drug Dicumarol; the half-life can vary from seven to 74 hours. That makes it hard to predict how much of the drug will provide the desired anticoagulant effect in a patient. Marked variations have also been observed in the metabolism in man of phenylbutazone and of diphenylhydantoin, a drug that is administered to control epileptic seizures.

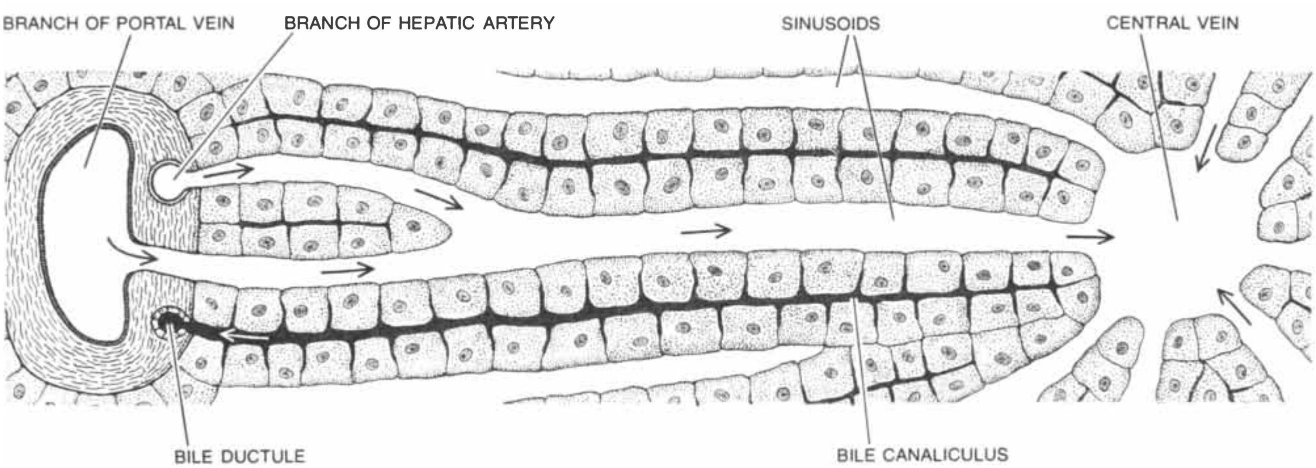
Genetic factors may play an important role in the metabolism of drugs, as Elliot S. Vesell of the Pennsylvania State University College of Medicine has shown. The large individual differences in Dicumarol half-life persist to some extent when fraternal twins are compared but almost disappear in identical twins, and similar results have been reported for other drugs studied in twins. The antituberculosis drug isoniazid has been the subject of a number of investigations covering different populations. Isoniazid is metabolized in man primarily by an acetylation reaction catalyzed by the enzyme *N*-acetyl transferase. Soon after the drug was introduced it became apparent that individuals vary in their ability to acetylate it. The distribution of the acetylation rates for a group of subjects tends to be bimodal, that is, plotting the rate against the number of patients exhibiting each rate yields a curve with two peaks. Apparently there are two classes of individuals: those who acetylate isoniazid rapidly and excrete the drug primarily as acetylisoniazid and those who acetylate and excrete the drug more slowly and excrete more of it in an unchanged form. The frequency of rapid inactivators is about 90 percent in a population of Eskimos or of Japanese, whereas among both whites and blacks in North America there are about as many slow inactivators as there are rapid ones.

A clear relation has now been established between this variation in acetylation rate and the incidence of isoniazid toxicity. Several studies have shown that the slow inactivators are more susceptible to a toxic effect of isoniazid: a disorder of the peripheral nerves caused by a specific vitamin B_6 deficiency resulting from an interaction of the drug and the vitamin.

The first indication that microsomal enzymes were special kinds of complexes was developed some 20 years ago by Gerald C. Mueller and James A. Miller at the University of Wisconsin. They discovered that compounds known as aminoazo dyes could be oxidized (specifically, *N*-demethylated) by liver microsomes and that the oxidation required molecular oxygen (O_2) and the coenzyme NADPH (reduced nicotinamide-adenine dinucleotide phosphate). Soon afterward Bernard B. Brodie's group at the National Heart and Lung Institute showed that there was a similar requirement for the oxidative metabolism of a number of drugs. In 1957 Howard S. Mason of the University of Oregon Medical School proposed that such oxidations are catalyzed by a class of "mixed-function oxidases": enzyme complexes that require oxygen and NADPH, are nonspecific (they catalyze the oxidation of different kinds of compounds) and catalyze the consumption of a molecule of oxygen for each molecule of the drug or other substrate, with one atom of oxygen appearing in the metabolized substrate and the other atom usually combining with two hydrogen atoms to form water.

The key enzyme of these oxidases is cytochrome *P*-450. A cytochrome is a complex of protein and heme, the iron-containing ring structure that is the oxygen-binding component of hemoglobin. Like hemoglobin, the various cytochromes serve to bind oxygen, which they deliver to their substrates in such processes as cell respiration. Cytochrome *P*-450 gets its designation from the fact that in the reduced form it binds carbon monoxide and then absorbs light most intensely at a wavelength of 450 nanometers. The amplitude of the absorbance peak is the basis of quantitative studies of the enzyme. In the mixed-function oxidases cytochrome *P*-450 serves as the terminal oxidase: it accepts electrons passed along by several intermediates, binds oxygen and then delivers the oxygen to oxidize its substrate and (usually) produce water [*see illustration on page 30*]. In addition to NADPH the system includes the enzyme cytochrome *P*-450 reductase. Another heme protein, cytochrome b_5, is present in liver microsomes and may participate in drug oxidations, but its precise function is not clear.

Nor is it known just how the various components of the mixed-function oxidase complex are arrayed in the tubular membranes of the endoplasmic reticulum. It appears that the reductase and cytochrome b_5 are on the outside of the membranes, with cytochrome *P*-450 in the deeper layers. The enzyme glucuronyl transferase, which catalyzes the most important conjugation reaction, is also present within the membranes. The drug or other fat-soluble compound to be transformed is presumably bound and



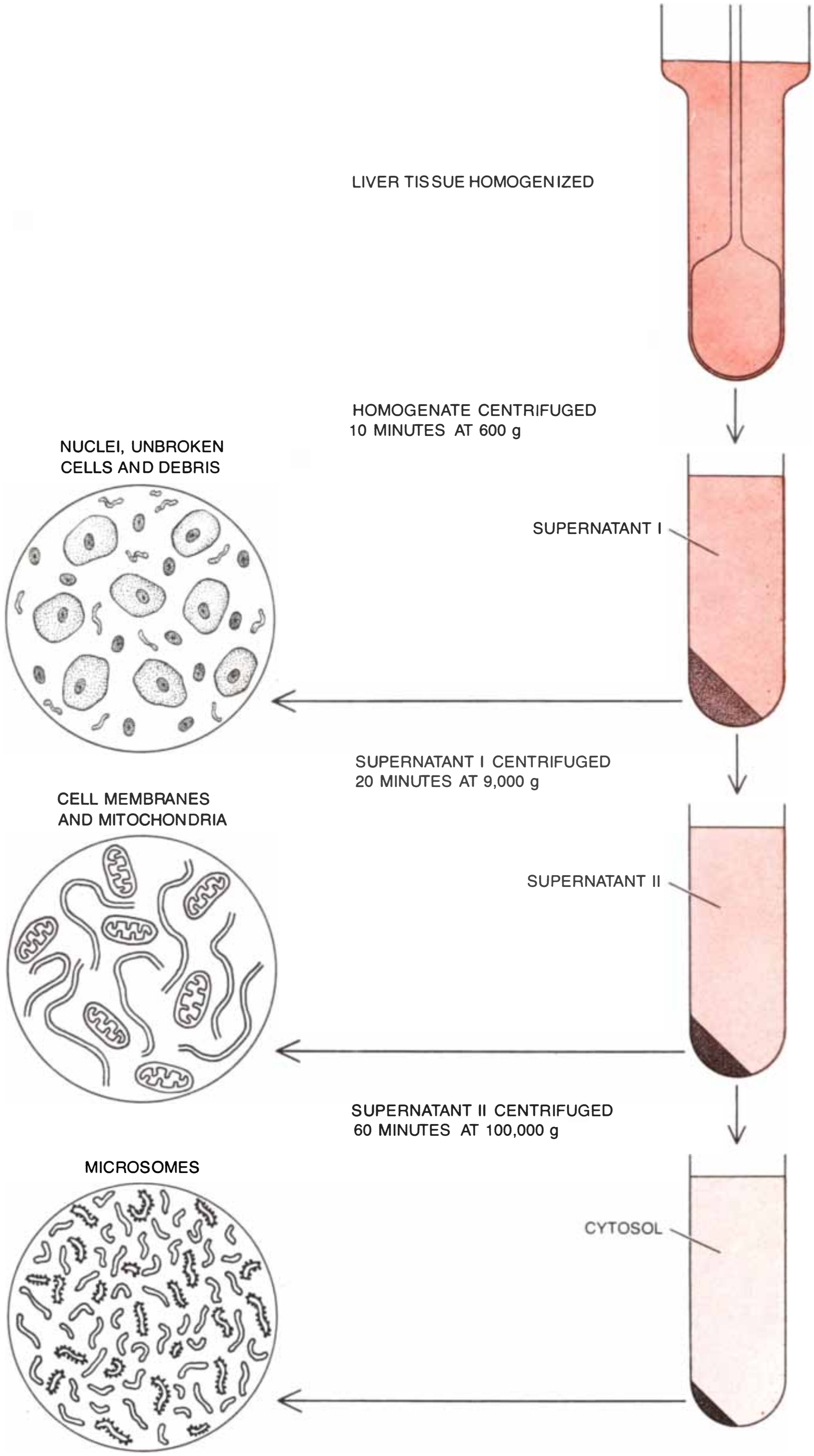
LIVER LOBULE, the functional unit of liver tissue, is defined by branches of the portal vein, the hepatic artery, the bile duct and lymphatic vessels, which run together through the tissue, outlining the lobules. (The lymphatics are not shown here.) The venous blood, with its nutrients and other ingested substances, and the oxygenated arterial blood enter the fine sinusoids and perfuse the liver cells, which carry out the metabolic functions discussed in the text. The blood drains into a central vein in each lobule and thence to sublobular and hepatic veins. The liver cells also secrete bile, which is collected by bile canaliculi that feed into bile ductules and eventually into the main bile duct. This two-dimensional view of part of a lobule is highly simplified and diagrammatic.

metabolized by the mixed-function oxidase system that contains cytochrome *P*-450 and then is often converted into a highly water-soluble compound by conjugation with glucuronic acid; the transformed product passes into the lumen, or central channel, of the tubular membranes and is excreted from the cell into the bile or the bloodstream.

The cytochrome *P*-450 mixed-function oxidase system has come to be recognized as having a central role in the body's defense against chemical agents, whether they are normal body constituents or are introduced from the environment. It is now clear that the system is responsible for the detoxification of many of the potentially harmful environmental pollutants. The system is highly inducible, that is, its activity can be greatly increased by exposure to a wide variety of environmental agents and drugs that act as substrates for the system. Such chemicals stimulate the synthesis of cytochrome *P*-450 and other components of the complex. More than 200 steroid hormones, drugs, insecticides, carcinogens and other foreign chemicals are now known to stimulate drug metabolism in experimental animals, and many of them have been shown to do the same thing in man. On the other hand, some substances (such as lead and other heavy metals) inhibit the mixed-function oxidase system, although they are fewer in number and less diverse than the inducers. The *P*-450 system is also the site of much competitive interaction among drugs and other chemicals that are undergoing transformation.

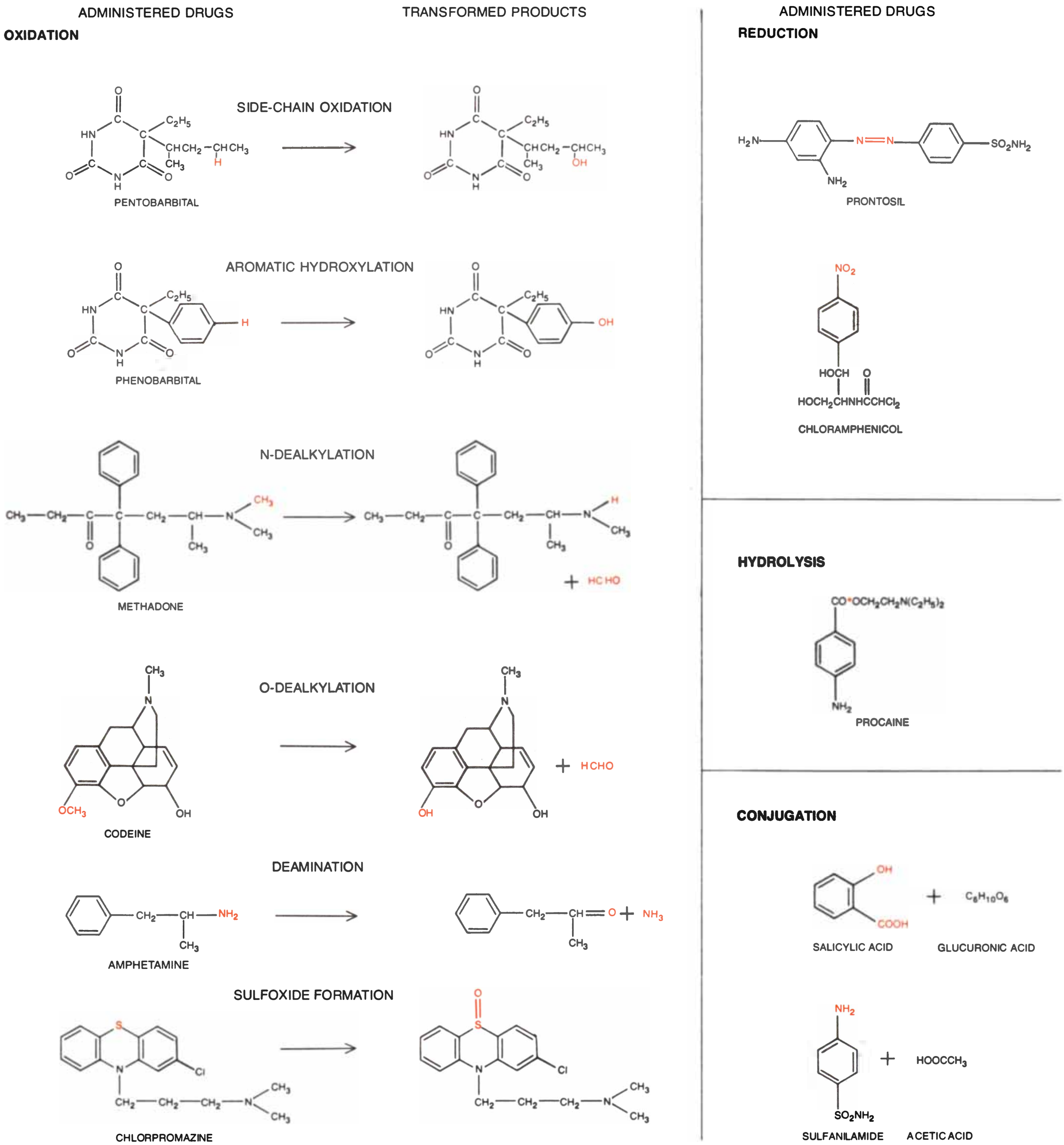
The consequences of this inducibility, inhibition and competition have important implications in drug therapy. Patients are often given several drugs at the same time. Certain combinations can have unpredictable and often undesirable effects if one drug inhibits or stimulates the metabolism of another or competes with it. For example, phenylbutazone, the coumarin anticoagulants and chloramphenicol compete with the metabolic inactivation of tolbutamide, a drug given to reduce the blood-sugar level in diabetics; the competition can lead to excessive tolbutamide activity and thus to serious hypoglycemia, or low blood sugar.

Phenobarbital is a prime example of a drug that has a different effect: enzyme induction. When rats are treated with phenobarbital, there can be a three- to fourfold increase in the microsomal content of cytochrome *P*-450 and a two-



LIVER MICROSOMES, the membrane structures that contain most of the liver enzymes engaged in detoxification, are isolated by spinning homogenized liver tissue at successively higher speeds, which produce successively higher gravity (*g*) forces, in a centrifuge. Homogenized liver tissue is centrifuged for 10 minutes at 600 *g*; a pellet of dense material, primarily whole cells, cell debris and cell nuclei, collects at the bottom of the tube. The supernatant, or liquid portion, is centrifuged more strongly, isolating less dense structures such as mitochondria and pieces of membrane. When supernatant II is spun at very high *g* forces in an ultracentrifuge, the microsomes are separated from the cytosol, or cell fluid.

fold increase in the reductase, and drugs such as methadone and ethylmorphine are metabolized three or four times as fast. In accordance with these experimental observations chronic administration of phenobarbital to patients decreases the effects of many drugs by hastening their inactivation. Sedative doses of the drug reduce the concentration in the blood plasma of phenylbutazone, the analgesic antipyrine and the coumarin anticoagulants; it also decreases their pharmacological actions. Anticoagulant therapy is particularly sensitive to dosage. A patient who is satisfactorily maintained on a given coumarin-drug dosage while he is being given phenobarbital as a sedative may hemorrhage when the phenobarbital is withdrawn because the P-450 system is no longer being induced and the anticoagu-



BIOTRANSFORMATIONS carried out in the liver include oxidation, reduction, hydrolysis and conjugation, examples of which are given here. The molecular sites of each reaction are indicated in color. In a few cases transformation converts an inactive form of a drug (such as Prontosil) into an active form (sulfanilamide), but most of the reactions lead to inactivation. Conjugation with a nat-

lant is therefore being inactivated more slowly.

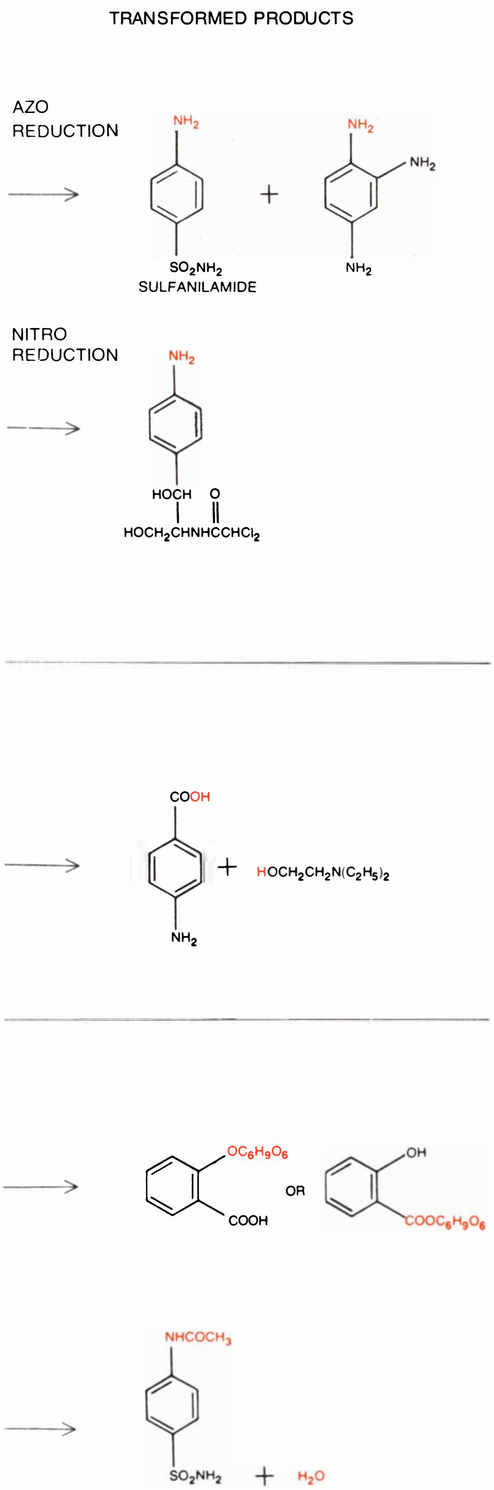
Alcohol is converted into acetaldehyde largely in the liver, perhaps to some extent by the mixed-function oxidase system: heavy drinkers are found to have an increased concentration of cytochrome *P*-450. The habitual consumption of alcohol therefore stimulates the metabolism of a wide variety of drugs. This helps to explain why heavy drinkers are less affected than other people by barbiturates and other sedatives—when they are sober. A single very large dose of alcohol taken together with another drug, on the other hand, inhibits the drug's metabolism, presumably by competing with the drug for the appropriate enzymes. This effect, in addition to the depressant effect of alcohol on the central nervous system, helps to explain the enhanced sensitivity to barbiturates and other sedatives of a person who has been drinking heavily. The synergistic actions of alcohol and sedatives in the brain can cause death.

The inducibility of microsomal enzymes by drugs suggests a form of therapy for certain conditions in which normal body constituents ordinarily metabolized by such enzymes are present in excessive amounts. Long-term administration of phenobarbital, for example, can lower the concentration of bilirubin, the pigment that produces jaundice, in the blood of patients with chronic obstruction of bile flow in the liver. The excessive bilirubin levels that are normally observed in infants after birth can also be markedly reduced if the mother is given a small dose of phenobarbital for a number of days before delivery. Presumably the drug crosses the placenta and stimulates the conjugating enzyme system that is ordinarily slow to develop in the fetus and the newborn infant.

Halogenated hydrocarbon insecticides such as DDT are potent stimulators of drug and steroid sex-hormone metabolism in mammals and in birds; the breakdown of sex hormones explains in part the devastating effects of DDT on reproduction in some bird populations. The minimum exposure to DDT that will stimulate the metabolism of pentobarbital and decrease its hypnotic action in experimental animals is one that results in concentrations of from 10 to 15 micrograms of DDT per gram of fat; that is a level commonly found in human fat tissues. Among the other insecticides that induce microsomal enzymes in experimental animals are chlordane, aldrin and dieldrin. It is interesting that piperonyl butoxide, a synergist that was added to insecticides to inhibit the enzymatic defenses insects had developed against DDT and its chemical relatives, also inhibits the activity of the microsomal enzymes in the mammalian liver.

The polychlorinated biphenyls (PCB's) constitute another class of environmental pollutants that have been shown to induce microsomal enzymes. The PCB's are lubricants, heat-exchange fluids, insulators, plasticizers for paints and plastic compounds and a major component of the lens-immersion oil used in microscopy. Whereas some of the consumer-product applications have recently been curtailed, the immersion oils are still handled daily by many laboratory workers. PCB's have been found in the tissues of numerous bird and fish species and in human fat and milk, although the route of entry into the human body has not been accurately determined. In recent experiments we have been able to show that the application of pure PCB's or of microscope immersion oil to the skin of experimental animals in very small amounts (one microliter) causes a marked increase in mixed-function oxidase activity and reduces the pharmacological effect of zoxazolamine, a muscle relaxant, and of hexobarbital in the live animal. These findings suggest that trivial skin exposure to chemicals can have significant and perhaps harmful biological effects in man.

A number of chemicals to which human beings are regularly exposed have been identified as chemical carcinogens, that is, they cause cancers when they are applied to the skin of experimental animals or otherwise administered to them. Clearly a factor that inhibits or stimulates the metabolism of such compounds may affect the development of human cancers. Benzpyrene, benzanthracene and similar polycyclic aromatic hydrocarbons are among the most ubiquitous carcinogens: they are present in tobacco smoke, in polluted city air and in charcoal-broiled and smoked foods. The polycyclic hydrocarbons are metabolized by a mixed-function oxidase enzyme that in this case is called aryl hydrocarbon hydroxylase because of the particular oxidation it catalyzes. Like other microsomal oxidations, this one requires NADPH and molecular oxygen, but the terminal oxidase of the system induced by the polycyclic hydrocarbons is somewhat different from the oxidase induced by drugs. The catalytic properties of the cytochrome are changed, as are its spectral properties: the absorbance maximum of the complex of carbon monoxide with the reduced cytochrome is at 448 nanometers rather than at 450, and so the enzyme is designated cytochrome *P*-448. Apart from the polycyclic hydrocarbons, only one other class of compounds has so far been noted to induce the formation of cytochrome *P*-448: we have found that the PCB's induce



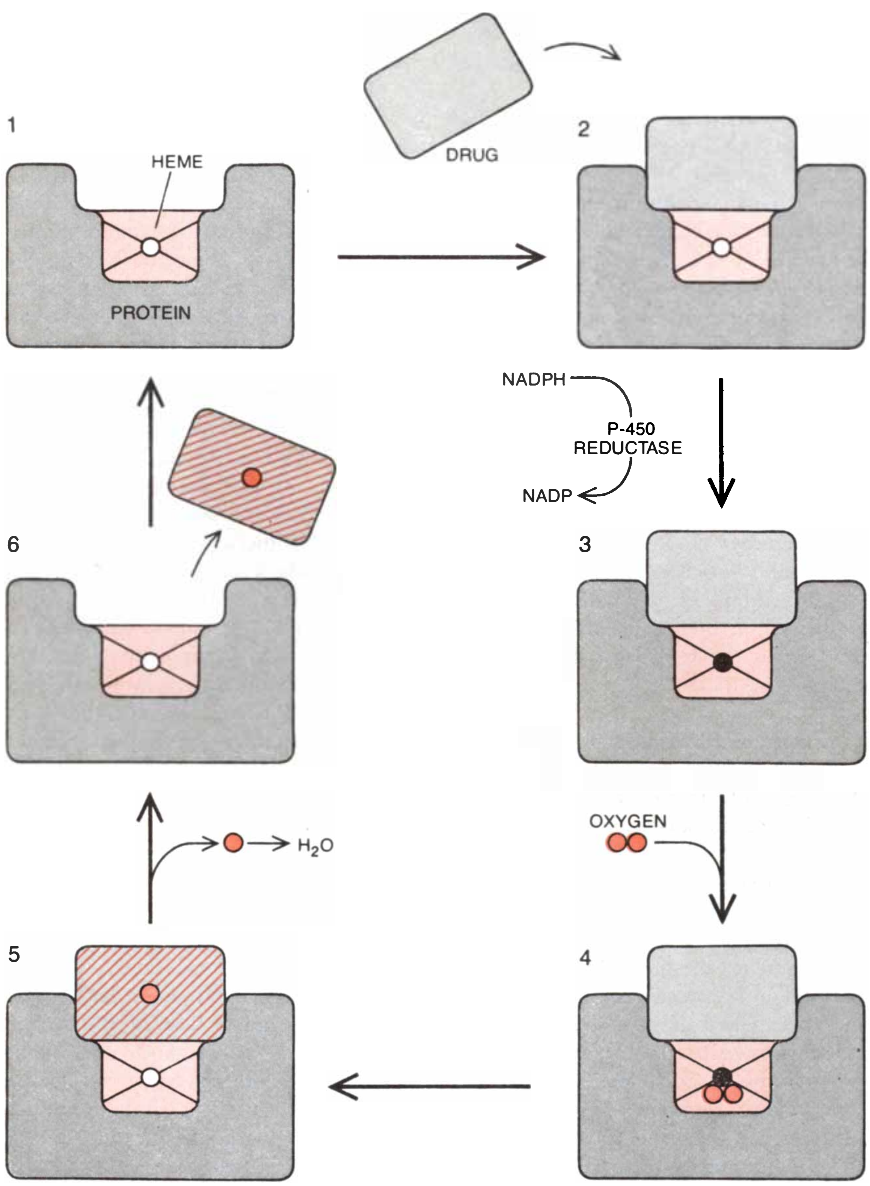
ural substance such as glucuronic acid often follows metabolism of drugs and facilitates excretion. Not all the by-products are shown.

some of the newly identified cytochrome along with cytochrome *P*-450. This suggests the possibility, for which there is developing evidence in animals, that the PCB's too may have carcinogenic properties.

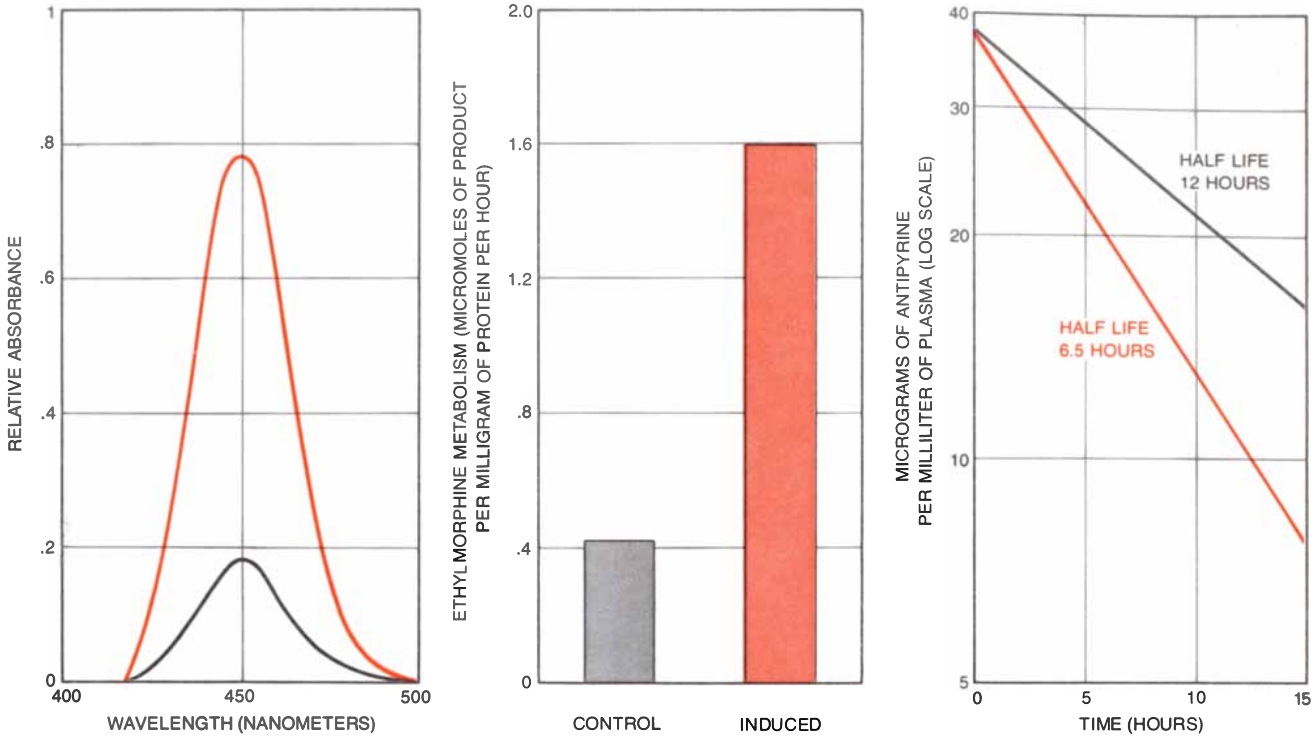
The aryl hydrocarbon hydroxylase system has been the subject of intensive investigation as a possible link in the causation of some cancers. The reason is that rather than detoxifying its polycyclic-hydrocarbon substrates it seems to make some of them more toxic: intermediates of polycyclic-hydrocarbon metabolism such as epoxides are more active in the malignant transformation of tissue-cultured cells than the parent products are. Moreover, the enzyme system for carcinogen metabolism is available at many sites in the body that are exposed to polycyclic hydrocarbons and is induced not only in the liver but also in the gastrointestinal tract, the kidneys, the skin and the lungs. Cigarette smoking markedly induces aryl hydrocarbon hydroxylase activity even in the human placenta, as Allan H. Conney, Richard M. Welch and their colleagues showed at the Wellcome Research Laboratories. Little or no such enzyme activity was found in placentas from nonsmokers.

We have investigated the cytochrome *P*-448 system in human skin and have found that there is marked variability in the hydroxylase activity of different skin samples and that incubation of skin in tissue culture with the polycyclic hydrocarbon benzanthracene induces more enzyme activity. A skin biopsy is easy to do, so that assaying skin aryl hydroxylase in the presence or absence of polycyclic hydrocarbons may provide a convenient test of individual differences in the capacity to metabolize certain environmental carcinogens. The possibility that induction of aryl hydrocarbon hydroxylase is of considerable significance in chemical carcinogenesis is suggested by the results of some recent experiments performed by Gottfried Kellermann and his associates at the University of Texas. They found that the inducibility of such hydroxylase activity in lymphocytes is significantly greater in cigarette smokers who have lung cancer than it is in healthy nonsmokers.

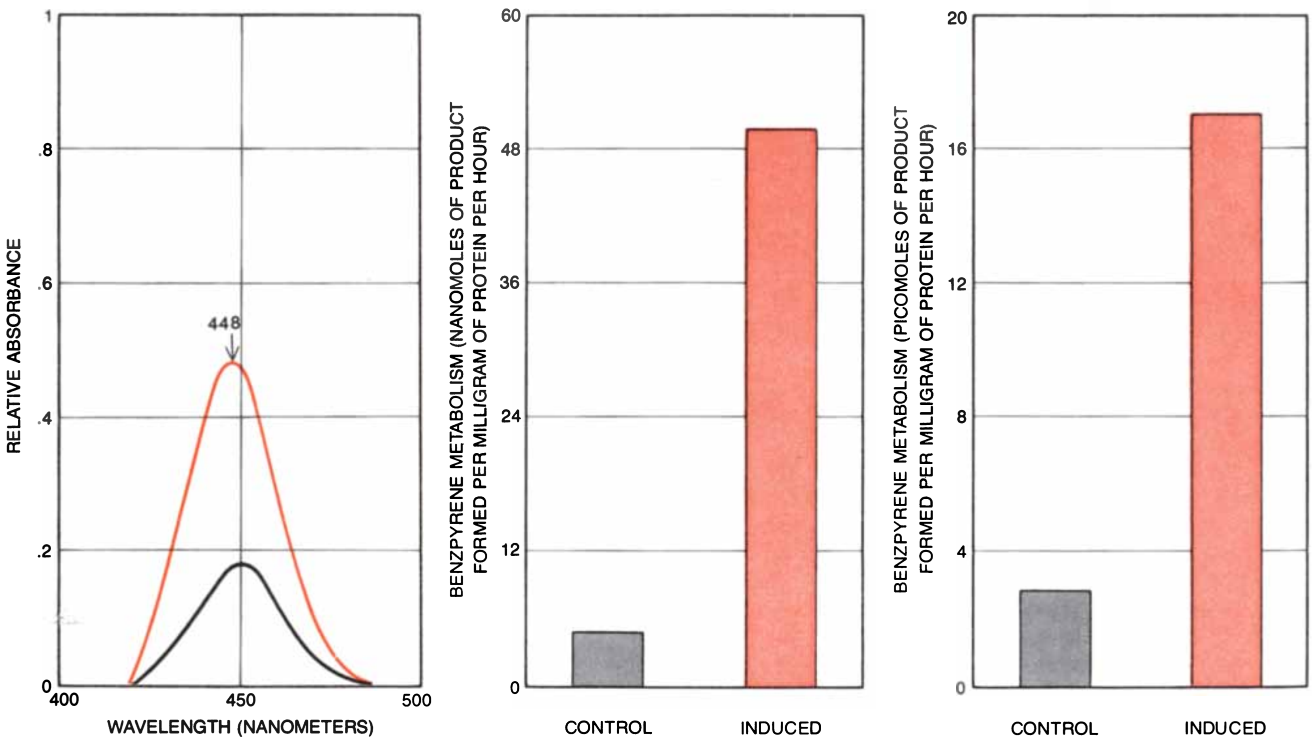
It is clear that the liver's mixed-function oxidase system is implicated in a number of processes affecting human health. The ability of a drug or other foreign substance to stimulate the metabolism of another drug by the system may explain some of the adverse drug reactions observed in clinical practice. Drug interactions have been well documented in anticoagulant therapy; more research is required to explore other drug interactions, since most patients are given several drugs at the same time. The ability of environmental pollutants to modify drug action is now under active investigation. It is clear that insecticides, for example, stimulate drug-metabolizing enzymes and that heavy-metal substances such as lead and methyl mercury inhibit the enzymes; the clinical significance of these effects may be appreciable in populations that are exposed occupationally to such agents. Finally, recently acquired evidence seems to indicate that the induction of aryl hydrocarbon hydroxylase in human tissues by polycyclic hydrocarbons may play a significant role in chemical carcinogenesis. Research on this biological action of carcinogens may provide important leads toward predicting the susceptibility of individuals to certain kinds of chemically induced cancer.



OXIDATION OF A DRUG by the enzyme cytochrome *P*-450 is visualized here as a sequential process. The enzyme (*1*) is a complex of a protein and the oxygen-binding compound heme, which contains an iron atom that is initially in the ferric (Fe^{+++}) form (*open circle*). The cytochrome binds the drug (*2*). Then (*3*) the enzyme cytochrome *P*-450 reductase, utilizing the coenzyme NADPH, reduces the iron of the heme to the ferrous (Fe^{++}) form (*black dot*), in which it can bind a molecule of oxygen (*4*). It supplies one atom of oxygen to oxidize the drug and one generally to form water, in the process reverting to its oxidized form (*5*). The drug, oxidized and in most cases inactive, is thereupon released (*6*).



PHENOBARBITAL induces cytochrome *P*-450 and thus stimulates the metabolism of drugs oxidized by that enzyme. The curves (*left*) show the light-absorbance spectrum, with a peak at 450 nanometers, characteristic of reduced cytochrome *P*-450 complexed with carbon monoxide, the usual means of identifying and quantifying the enzyme. In a rat treated with phenobarbital the cytochrome *P*-450 content of microsomes (*colored curve*) is substantially higher than that of microsomes from an untreated rat (*black curve*). More of the test drug ethylmorphine is therefore metabolized (*center*) by liver tissue from a phenobarbital-treated rat (*colored bar*) than by the same amount of tissue from an untreated rat (*gray*). Similarly, phenobarbital speeds up metabolism of the drug antipyrine in man (*right*): antipyrine level in plasma falls faster after phenobarbital treatment (*colored curve*) than ordinarily (*black*).



CARCINOGENS induce a different species of cytochrome, cytochrome *P*-448. The curves (*left*) compare the absorbance of microsomes from the liver of a rat treated with the carcinogen methylcholanthrene (*color*) with the absorbance of control microsomes (*black*); the induced cytochrome has a peak at 448 nanometers. Methylcholanthrene enhances metabolism of another carcinogen, benzpyrene (*center*): the microsomes of treated rats produce more metabolite (*color*) than control microsomes (*gray*) do. Similarly, treating human skin in tissue culture with the carcinogen benzanthracene stimulates skin-cell metabolism of benzpyrene (*right*).

SLAVERY IN ANTS

Certain species of ants raid the nests of other species for ants to work in their own nest. Some raiding species have become so specialized that they are no longer capable of feeding themselves

by Edward O. Wilson

The institution of slavery is not unique to human societies. No fewer than 35 species of ants, constituting six independently evolved groups, depend at least to some extent on slave labor for their existence. The techniques by which they raid other ant colonies to strengthen their labor force rank among the most sophisticated behavior patterns found anywhere in the insect world. Most of the slave-making ant species are so specialized as raiders that they starve to death if they are deprived of their slaves. Together they display an evolutionary descent that begins with casual raiding by otherwise free-living colonies, passes through the development of full-blown warrior societies and ends with a degeneration so advanced that the workers can no longer even conduct raids.

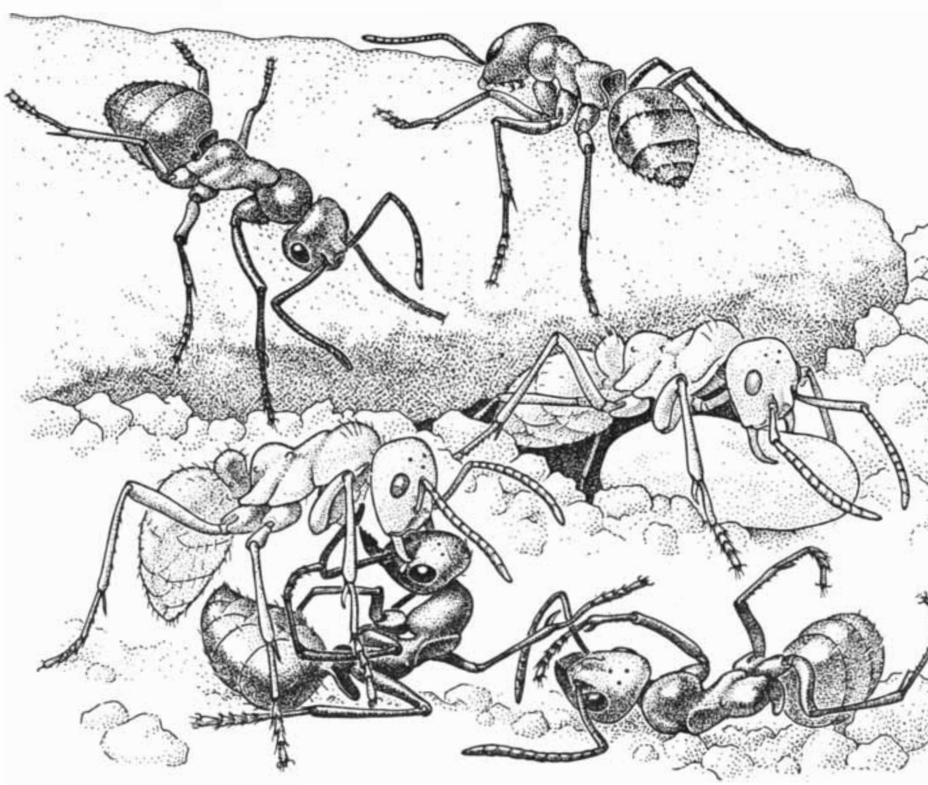
Slavery in ants differs from slavery in human societies in one key respect: the ant slaves are always members of other completely free-living species that themselves do not take slaves. In this regard the ant slaves perhaps more closely resemble domestic animals—except that the slaves are not allowed to reproduce and they are equal or superior to their captors in social organization.

The famous Amazon ants of the genus *Polyergus* are excellent examples of advanced slave makers. The workers are strongly specialized for fighting. Their mandibles, which are shaped like miniature sabers, are ideally suited for puncturing the bodies of other ants but are poorly suited for any of the routine tasks that occupy ordinary ant workers. Indeed, when *Polyergus* ants are in their home nest their only activities are begging food from their slaves and cleaning themselves ("burnishing their ruddy armor," as the entomologist William Morton Wheeler once put it).

When *Polyergus* ants launch a raid, however, they are completely transformed. They swarm out of the nest in a solid phalanx and march swiftly and directly to a nest of the slave species. They destroy the resisting defenders by puncturing their bodies and then seize and carry off the cocoons containing the pupae of worker ants.

When the captured pupae hatch, the workers that emerge accept their captors as sisters; they make no distinction between their genetic siblings and the *Polyergus* ants. The workers launch into the round of tasks for which they have been genetically programmed, with the slave makers being the incidental beneficiaries. Since the slaves are members of the worker caste, they cannot reproduce. In order to maintain an adequate labor force, the slave-making ants must periodically conduct additional raids.

It is a remarkable fact that ants of slave-making species are found only in cold climates. Although the vast majority of ants live in the Tropics and the



RAID BY SLAVE-MAKING AMAZON ANTS of the species ***Polyergus rufescens*** **(*light color*) against a colony of the slave species *Formica fusca* (*dark color*) is depicted. The *fusca* ants make their nest in dry soil under a stone. The raiding Amazon ants kill resisting**

warm Temperate zones, not a single species of those regions has been implicated in any activity remotely approaching slavery. Among the ants of the colder regions this form of parasitism is surprisingly common. The colonies of many slave-making species abound in the forests of the northern U.S., and ant-slave raids can be observed in such unlikely places as the campus of Harvard University.

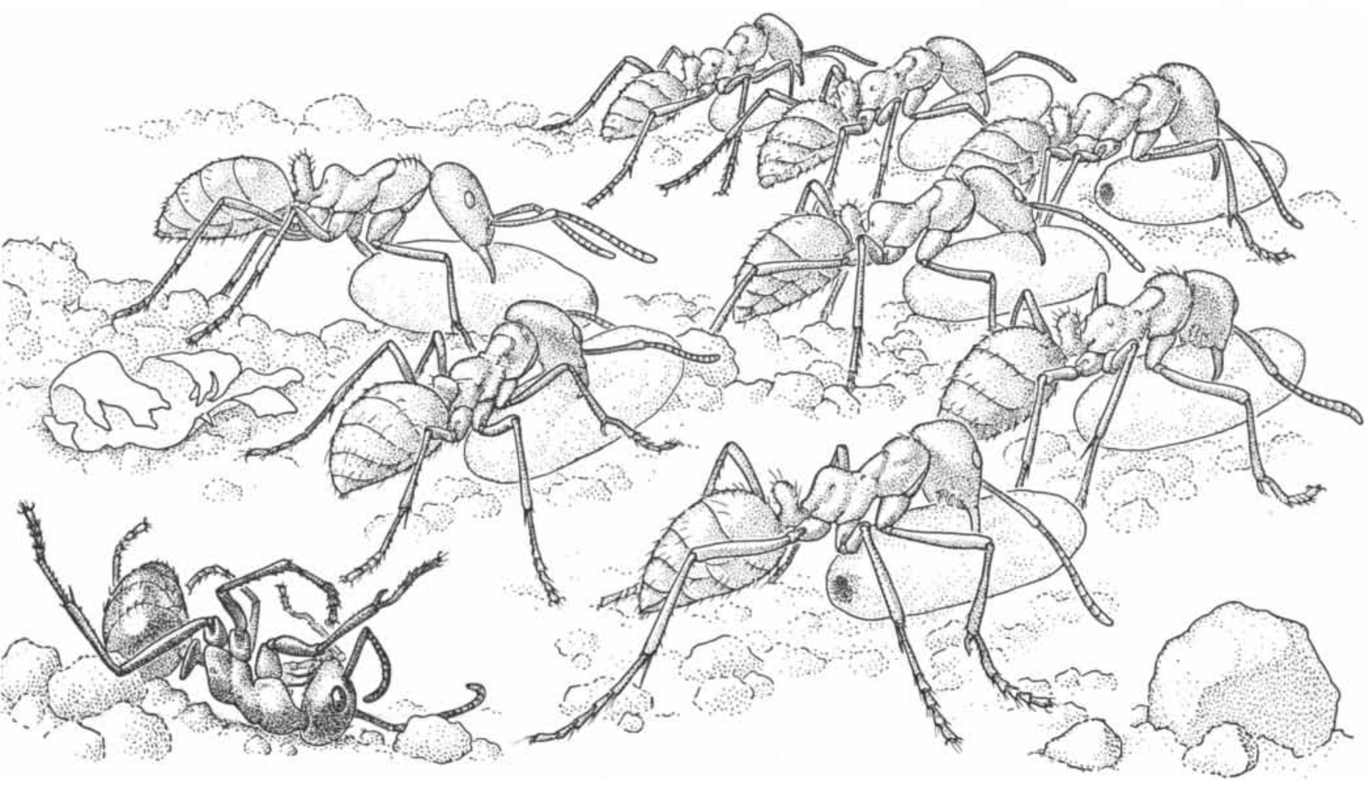
The slave raiders obey what is often called Emery's rule. In 1909 Carlo Emery, an Italian myrmecologist, noted that each species of parasitic ant is genetically relatively close to the species it victimizes. This relation can be profitably explored for the clues it provides to the origin of slave making in the evolution of ants. Charles Darwin, who was fascinated by ant slavery, suggested that the first step was simple predation: the ancestral species began by raiding other kinds of ants for food, carrying away their immature forms in order to be able to devour them in the home nest. If a few pupae could escape that fate long enough to emerge as workers, they might be accepted as nestmates and thus join the labor force. In cases where the captives subsequently proved to be more valuable as workers than as food, the raiding species would tend to evolve into a slave maker.

Although Darwin's hypothesis is attractive, I recently obtained evidence that territorial defense rather than food is the evolutionary prime mover. I brought together in the Harvard Museum of Comparative Zoology different species of *Leptothorax* ants that normally do not depend on slave labor. When colonies were placed closer together than they are found in nature, the larger colonies attacked the smaller ones and drove away or killed the queens and workers. The attackers carried captured pupae back to their own nests. The pupae were then allowed by their captors to develop into workers. In the cases where the newly emerged workers belonged to the same species, they were allowed to remain as active members of the colony. When they belonged to a different *Leptothorax* species, however, they were executed in a matter of hours. One can easily imagine the origin of slave making by the simple extension of this territorial behavior to include tolerance of the workers of related species. The more closely related the raiders and their captives are, the more likely they are to be compatible. The result would be in agreement with Emery's rule.

One species that appears to have just crossed the threshold to slave making is *Leptothorax duloticus,* a rare ant that so far has been found only in certain localities in Ohio, Michigan and Ontario. The anatomy of the worker caste is only slightly modified for slave-making behavior, suggesting that in evolutionary terms the species may have taken up its parasitic way of life rather recently.

In experiments with laboratory colonies I was able to measure the degree of behavioral degeneration that has taken place in *L. duloticus.* Like the Amazon ants, the *duloticus* workers are highly efficient at raiding and fighting. When colonies of other *Leptothorax* species were placed near a *duloticus* nest, the workers launched intense attacks until all the pupae of the other species had been captured.

In the home nest the *duloticus* workers were inactive, leaving almost all the ordinary work to their captives. When the slaves were temporarily taken away from them, the workers displayed a dramatic expansion in activity, rapidly taking over most of the tasks formerly carried out by the slaves. The *duloticus* workers thus retain a latent capacity for working, a capacity that is totally lack-

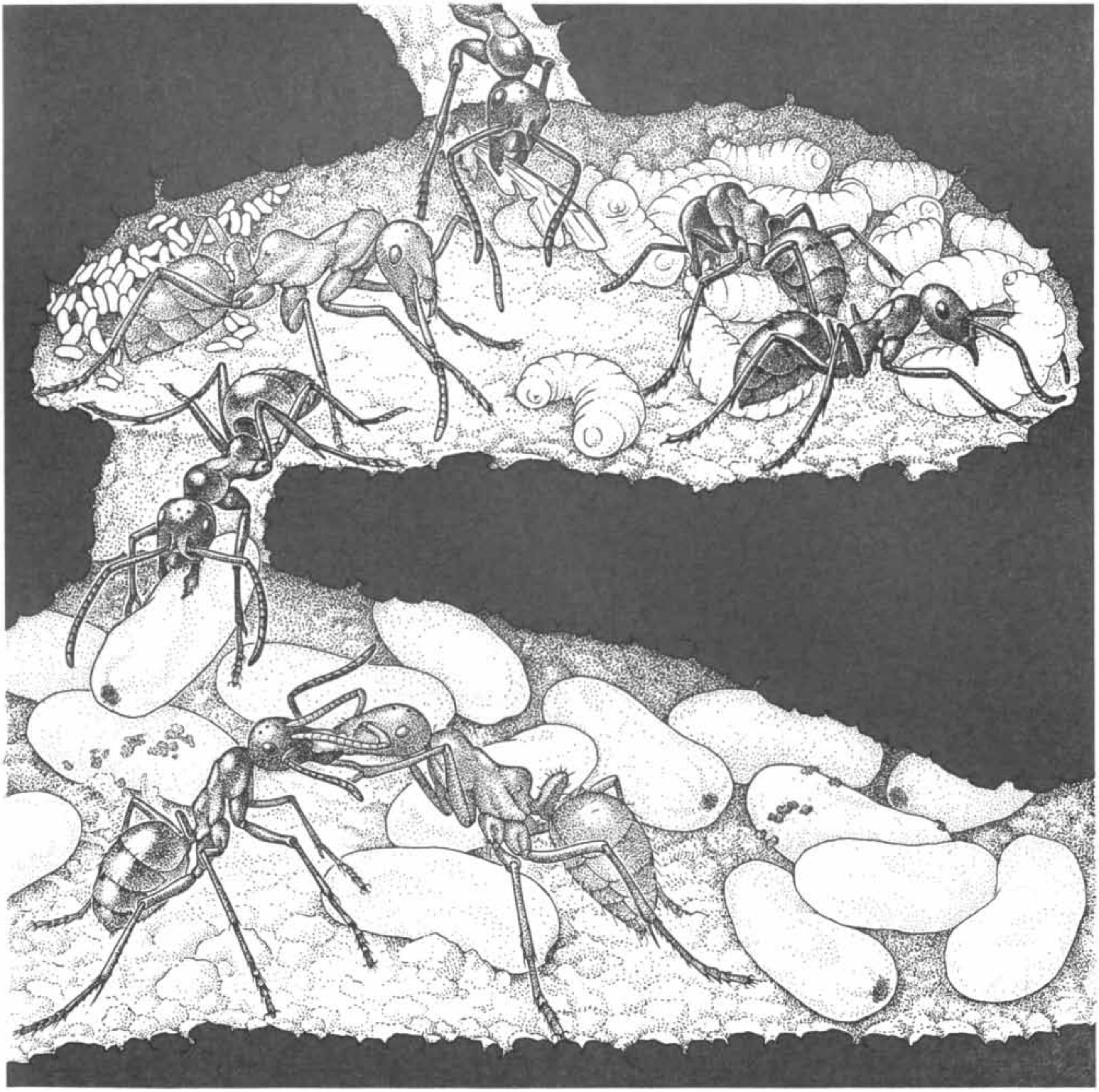


fusca workers by piercing them with their saberlike mandibles. Most of the Amazon ants are transporting cocoons containing the pupae of _fusca_ workers back to their own nest. When the workers emerge from the cocoons, they serve as slaves. Two dead _fusca_ workers that resisted lie on the ground. Two other workers have retreated to upper surface of the rock over the nest's entrance.

ing in more advanced species of slave-making ants.

The *duloticus* workers that had lost their slaves did not, however, perform their tasks well. Their larvae were fed at infrequent intervals and were not groomed properly, nest materials were carried about aimlessly and were never placed in the correct positions, and an inordinate amount of time was spent collecting and sharing diluted honey. More important, the slaveless ants lacked one behavior pattern that is essential for the survival of the colony: foraging for dead insects and other solid food. They even ignored food placed in their path. When the colony began to display signs of starvation and deterioration, I returned to them some slaves of the species *Leptothorax curvispinosus.* The bustling slave workers soon put the nest back in good order, and the slave makers just as quickly lapsed into their usual indolent ways.

Not all slave-making ants depend on brute force to overpower their victims. Quite by accident Fred E. Regnier of Purdue University and I discovered that some species have a subtler strategy. While surveying chemical substances used by ants to communicate alarm and to defend their nest, we encountered two slave-making species whose substances differ drastically from those of all other ants examined so far.· These ants, *Formica subintegra* and *Formica pergandei,* produce remarkably large quantities of decyl, dodecyl and tetradecyl acetates. Further investigation of *F. subintegra*



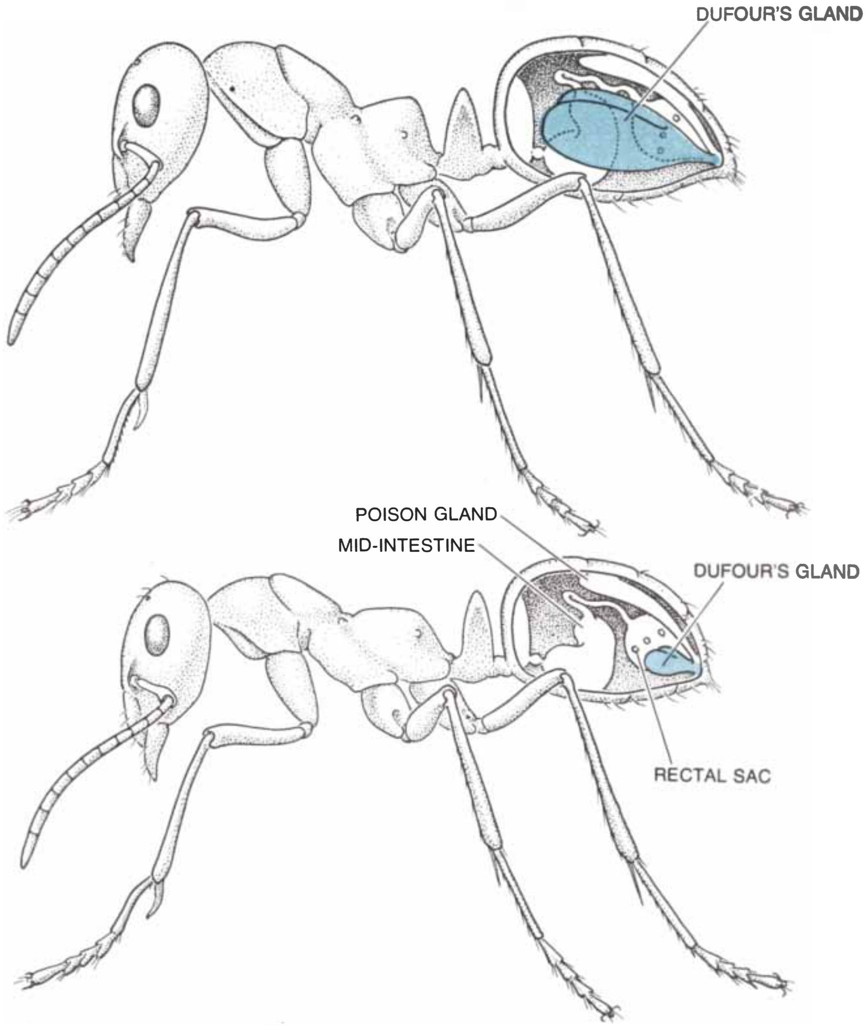
INTERIOR VIEW OF THE HOME NEST of a colony of Amazon ants shows *Formica fusca* slaves (*dark color*) performing all the housekeeping labor. At top center one of the slaves brings a fly wing into the nest for food. Other slave workers care for the small eggs, grublike larvae and cocoon-enclosed pupae of their captors. During the raiding season some of the pupae are likely to be those of *fusca* workers. The slave makers (*light color*) can do nothing more than groom themselves (*upper left*). In order to eat, the Amazon ants must beg slave workers to regurgitate liquid droplets for them (*lower left*). These ant species are found in Europe.

revealed that the substances are sprayed at resisting ants during slave-making raids. The acetates attract more invading slave makers, thereby serving to assemble these ants in places where fighting breaks out. Simultaneously the sprayed acetates throw the resisting ants into a panic. Indeed, the acetates are exceptionally powerful and persistent alarm substances. They imitate the compound undecane and other scents found in slave species of *Formica,* which release these substances in order to alert their nestmates to danger. The acetates broadcast by the slave makers are so much stronger, however, that they have a long-lasting disruptive effect. For this reason Regnier and I named them "propaganda substances."

We believe we have explained an odd fact first noted by Pierre Huber 165 years ago in his pioneering study of the European slave-making ant *Formica sanguinea.* He found that when a colony was attacked by these slave makers, the survivors of the attacked colony were reluctant to stay in the same neighborhood even when suitable alternative nest sites were scarce. Huber observed that the "ants never return to their besieged capital, even when the oppressors have retired to their own garrison; perhaps they realize that they could never remain there in safety, being continually liable to the attacks of their unwelcome visitors."

Regnier and I were further able to gain a strong clue to the initial organization of slave-making raids. We had made a guess, based on knowledge of the foraging techniques of other kinds of ants, that scout workers direct their nestmates to newly discovered slave colonies by means of odor trails laid from the target back to the home nest. In order to test this hypothesis we made extracts of the bodies of *F. subintegra* and of *Formica rubicunda,* a second species that conducts frequent, well-organized raids through much of the summer. Then at the time of day when raids are normally made we laid artificial odor trails, using a narrow paintbrush dipped in the extracts we had obtained from the ants. The trails were traced from the entrances of the nest to arbitrarily selected points one or two meters away.

The results were dramatic. Many of the slave-making workers rushed forth, ran the length of the trails and then milled around in confusion at the end. When we placed portions of colonies of the slave species *Formica subsericea* at the end of some of the trails, the slave makers proceeded to conduct the raid in a manner that was apparently the same

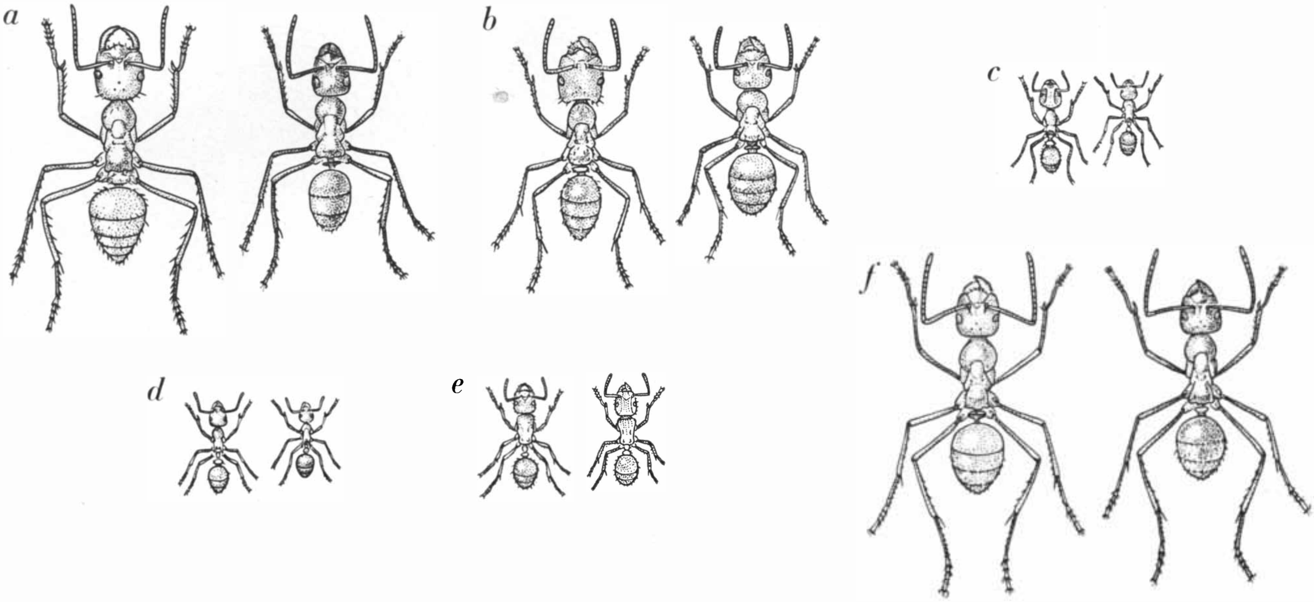


DUFOUR'S GLAND, which produces substances that serve as communication scents among ants, is much larger in the slave-making species *Formica subintegra* (*top*) than in the slave species *F. subsericea* (*bottom*). The *subsericea* ant releases its scent to alert its nestmates to the presence of danger. The *subintegra* sprays its secretions at resisting ants during slave raids. The secretions are so strong that they create panic in the colony being attacked.

in every respect as the raids initiated by trails laid by their own scouts. Studies of the slave-making species *Polyergus lucidus* and *Harpagoxenus americanus* by Mary Talbot and her colleagues at Lindenwood College provide independent evidence that raids are organized by the laying of odor trails to target nests; indeed, this form of communication may be widespread among slave-making ants.

The evolution of social parasitism in ants works like a ratchet, allowing a species to slip further down in parasitic dependence but not back up toward its original free-living existence. An example of nearly complete behavioral degeneration is found in one species of the genus *Strongylognathus,* which is found in Asia and Europe. Most species in this genus conduct aggressive slave-making raids. They have characteristic saber-shaped mandibles for killing other ants. The species *Strongylognathus testaceus,* however, has lost its warrior habits. Although these ants still have the distinctive mandibles of their genus, they do not conduct slave-making raids. Instead an *S. testaceus* queen moves into the nest of a slave-ant species and lives alongside the queen of the slave species. Each queen lays eggs that develop into workers, but the *S. testaceus* offspring do no work. They are fed by workers of the slave species. We do not know how the union of the two queens is formed in the first place, but it is likely that the parasitic queen simply induces the host colony to adopt her after her solitary dispersal flight from the nest of her birth.

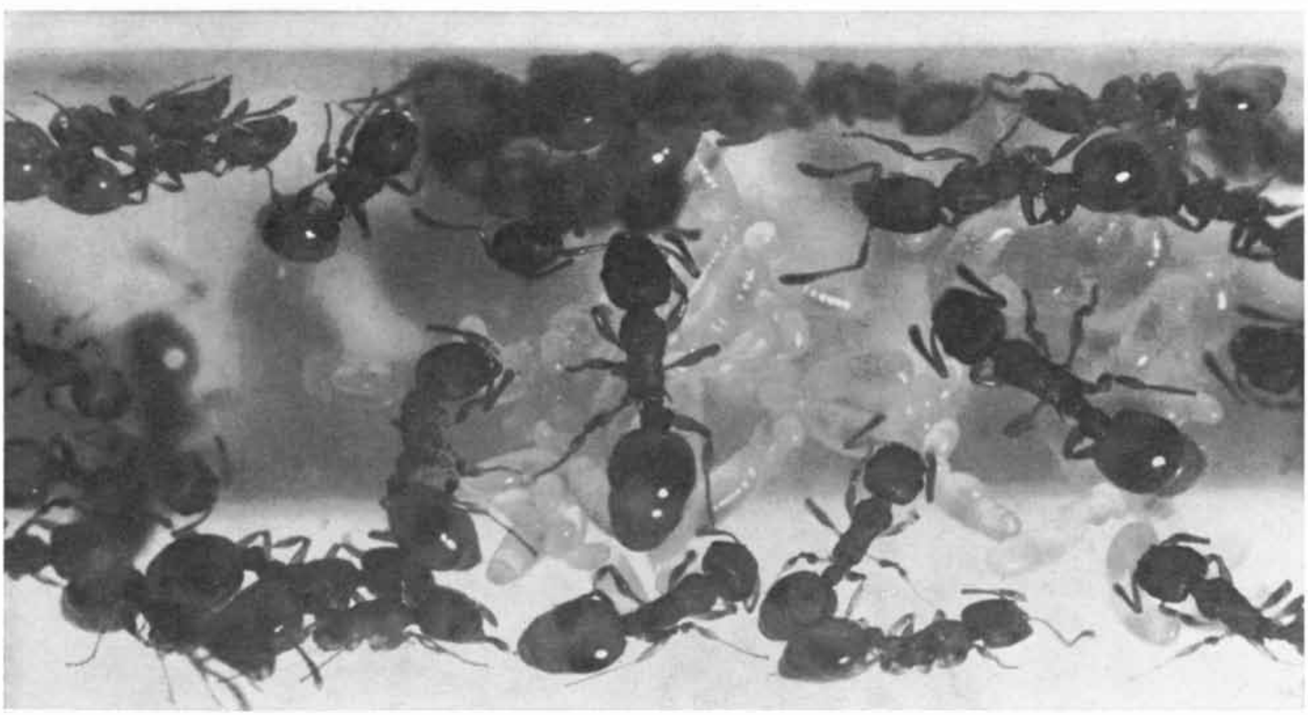
Thus *S. testaceus* is no longer a real slave maker. It has become an advanced social parasite of a kind that commonly infests other ant groups. For example,



RESEMBLANCE of slave maker and slave was noted by an Italian myrmecologist, Carlo Emery, in 1909. In each pair of ants shown here the slave maker is on the left and the slave on the right. The species depicted are (*a*) *Polyergus rufescens* and *Formica fusca,* (*b*) *Rossomyrmex proformicarum* and *Proformica nasutum,* (*c*) *Harpagoxenus americanus* and *Leptothorax curvispinosus,* (*d*) *L. duloticus* and *L. curvispinosus,* (*e*) *Strongylognathus alpinus* and *Tetramorium caespitum* and (*f*) *F. subintegra* and *F. subsericea.*

many species of ant play host to parasites such as beetles, wasps and flies, feeding them and sheltering them [see "Communication between Ants and Their Guests," by Bert Hölldobler; SCIENTIFIC AMERICAN, March, 1971].

Does ant slavery hold any lesson for our own species? Probably not. Human slavery is an unstable social institution that runs strongly counter to the moral systems of the great majority of human societies. Ant slavery is a genetic adaptation found in particular species that cannot be judged to be more or less successful than their non-slave-making counterparts. The slave-making ants offer a clear and interesting case of behavioral evolution, but the analogies with human behavior are much too remote to allow us to find in them any moral or political lesson.



COLONY OF ANTS housed in a glass tube consists of the rare species *Leptothorax duloticus* and a slave species, *L. curvispinosus.* The *duloticus* ant, found in Ohio, Michigan and Ontario, has only recently become a slave maker. One of the *duloticus* workers can be seen in the center of the photograph; below it are three slave workers. The white objects are immature forms of both species. When the slave workers are removed, the *duloticus* workers attempt to carry out necessary housekeeping tasks but do so poorly.

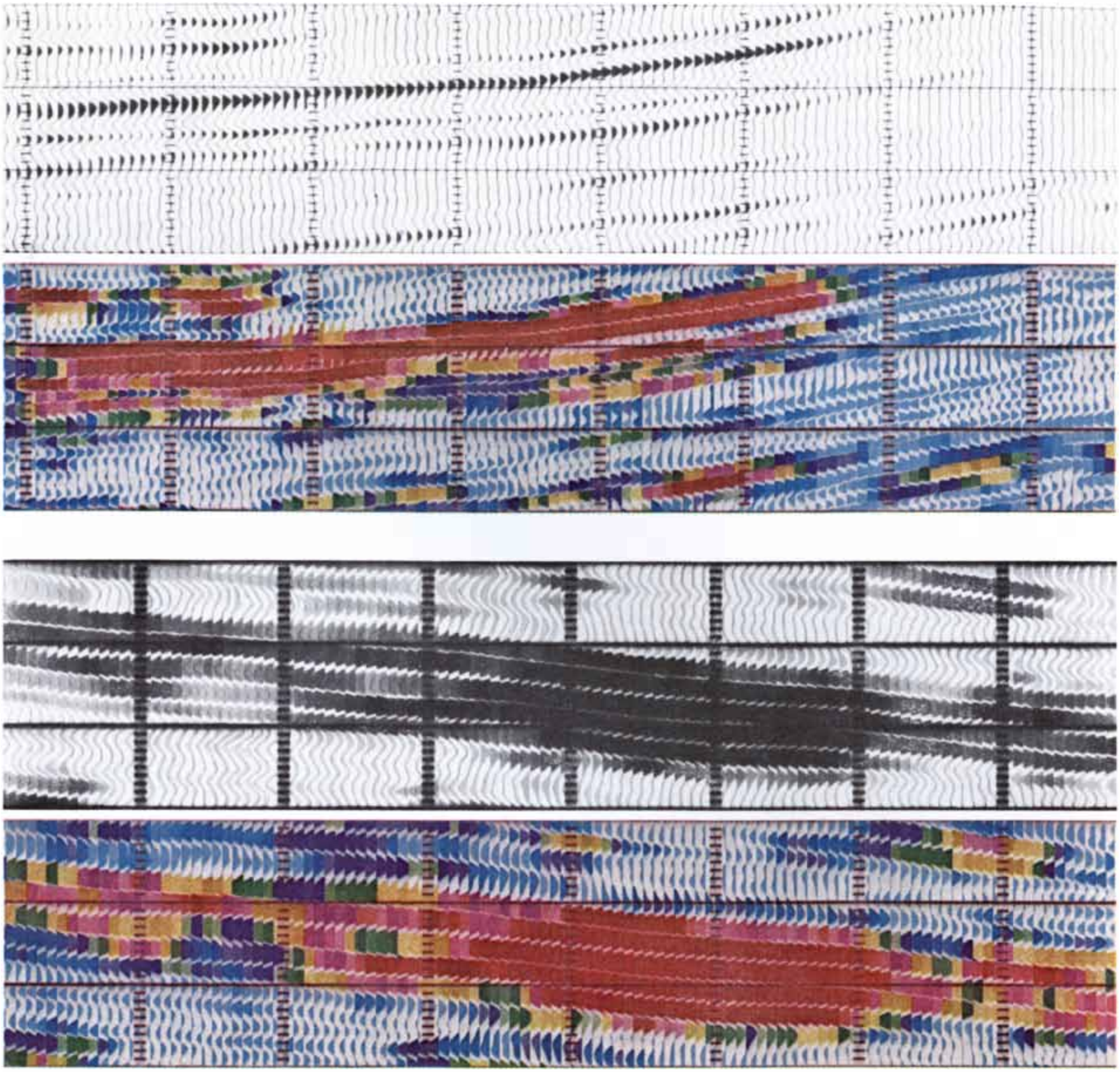
We are in the pretty picture business ...yes indeed

A

B

C

D



These are fragments of a geophysical "playback" recording from an oil company's exploration department. A pattern of sensors responded to a shock. The signals were taped. A computer "massaged" the tapes. But need remains for human judgment. The program is not quite capable of typing out a letter to the Finance Committee: "Invest $[ ] at Latitude [ ], Longitude [ ]. You won't be sorry."

Human judgment draws assurance from the human eye. "I see, I see," someone must say.

The company that was kind enough to permit reproduction of the above pictures was not kind enough to toss their costly secrets to the four winds. So you are not going to learn here how to read these pictures. You might, however, learn something applicable to your own work.

Note that **B** is **A** in color, and **D** is **C** in color. **A** uses the two dimensions of the paper and variable area of the blips. In **C** it's variable density instead of variable area. *In* **B** *and* **D** *the added dimension of color has allowed more information to be brought together for judgment.* See? Unless someone does see, the data is just a lot of tape.

We just make the color photographic materials. Others make the instruments which exploit human color sensation to make sense out of complicated data. The data need not be geophysical, of course. A. L. Sullivan can give you the names of companies in the business. His address is Kodak, Graphic Data Markets, Rochester, N.Y. 14650. Phone him at 716-724-4484.

Capability. Quality. Value.

The technological achievement under the keyboard is still the reason TI's professional calculators offer so much quality and math power for the money.

Engineer. Scientist. Businessman. Student. If you're doing more than basic mathematics in your field, consider an SR-50 or SR-51 from Texas Instruments.

Both deliver answers you can trust – quickly and efficiently – to problems ranging from simple arithmetic to highly complex calculations. You don't have to learn special entry methods or difficult-to-master key sequences. There's a better way – TI's algebraic entry system lets you key your problem just the way you would say it. Naturally. You don't worry about losing data in stacks, or keeping track of what is in each stack, or remembering if the stack is full. The way you learned math is the way it's done. On both the SR-50 and SR-51 – you can command tremendous math power with confidence – from the beginning. Power and accuracy you can really put to work.

Answers are calculated to 13 significant digits, rounded off and displayed to 10. And for maximum accuracy, all 13 are held inside for subsequent calculations.

Scientific notation is automatic when you need it. For numbers as large as $\pm 9.999999999 \times 10^{99}$. Or as small as $\pm 1. \times 10^{-99}$.

For a closer look at real math power, delve into the list. You'll see for yourself the SR-50 is a lot of calculator for the dollar. And the SR-51 does everything the 50 does and a lot more: Mean, variance and standard deviation. Linear regression. Percent and percent difference. Has a random number generator. And, offers 20 preprogrammed conversions and inverses.

Quality – it's built in right from the start. Texas Instruments designs and manufactures every critical component. From high-purity silicon semiconductor materials to integrated circuits to light-emitting-diode displays to circuit boards to keyboards. So, we design-in and control quality – not just monitor it – at every level: Materials. Components. The complete system.

FUNCTION	SR-51	SR-50
Log, lnx	yes	yes
Trig (sin, cos, tan, INV)	yes	yes
Hyperbolic (sinh, cosh, tanh, INV)	yes	yes
Degree-radian conversion	yes	yes
Deg/rad mode selection switch	yes	yes
Decimal degrees to deg.min.sec.	yes	no
Polar-rectangular conversion	yes	no
y^x	yes	yes
e^x	yes	yes
10^x	yes	no
x^2	yes	yes
$\sqrt{x}$	yes	yes
$\sqrt[x]{y}$	yes	yes
1/x	yes	yes
x!	yes	yes
Exchange x with y	yes	yes
Exchange x with memory	yes	no
% and Δ%	yes	no
Mean, variance and standard deviation	yes	no
Linear regression	yes	no
Trend line analysis	yes	no
Slope and intercept	yes	no
Store and sum to memory	yes	yes
Recall from memory	yes	yes
Product to memory	yes	no
Random number generator	yes	no
Automatic permutation	yes	no
Preprogrammed conversions	20	1
Digits accuracy	13	13
Algebraic notation (sum of products)	yes	yes
Memories	3	1
Fixed decimal option	yes	no
Keys	40	40
Second function key	yes	no
Constant mode operation	yes	no

To assure you reliable performance, every calculator is subjected to severe environmental and reliability testing prior to release to production. In production, every one is thoroughly tested, then "burned-in", then thoroughly tested again. If there's any problem, we want to find it before it gets to you.

Inside, steel machine screws anchor all important structural elements – plastic welds and glue fastenings aren't good enough. A double-tough Mylar* barrier keeps dust and moisture from getting under the keyboard. The case is high-strength, injection-molded plastic designed to take a beating. It's a quality calculator. And you know it as soon as you get your hands on one. The heft and solid feel tells you it's a fine-quality instrument even before you press a key.

The SR-50 and SR-51 are human engineered, too, for maximum comfort and efficiency. For a hand or a desktop. Keys have positive-action, tactile feedback. And the bright, wide-angle displays are easy to read at your desk or on the go. Slim. Compact. Light. In your briefcase or on your belt, you'll hardly notice just 8.3 ounces.

Technological leadership and quality craftsmanship are why Texas Instruments can offer so much value at such low prices: $109.95 for the SR-50. $179.95 for the SR-51.

20 Preprogrammed Conversions

FROM	TO
mils	microns
inches	centimeters
feet	meters
yards	meters
miles	kilometers
miles	nautical miles
acres	square feet
fluid ounces	cubic centimeters
fluid ounces	liters
gallons	liters
ounces	grams
pounds	kilograms
short ton	metric ton
BTU	calories, gram
degrees	gradients
degrees	radians
°Fahrenheit	°Celsius
deg.min.sec.	decimal degrees
polar	rectangular
voltage ratio	decibels

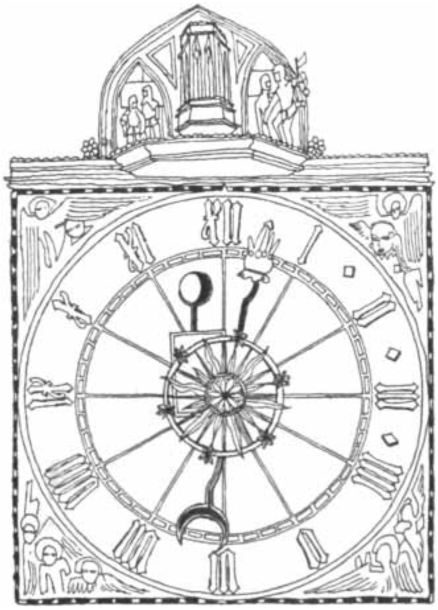
See them at your nearest TI calculator retailer. Or, send for our new fact-filled color brochure. It details the outstanding capability of both the SR-50 and SR-51 with full feature descriptions, sample problems, entry-method considerations and more. Write, Texas Instruments, M/S358, P.O. Box 22013, Dallas, Texas 75222

TEXAS INSTRUMENTS
INCORPORATED

82051

*Trademark of DuPont

SCIENCE AND THE CITIZEN



Buzz Bombs

Recent technological advances, mainly in the microminiaturization of electronic circuits, are leading to the emergence of an entirely new class of ultra-accurate guided missiles that could "drastically alter the conduct of tactical and strategic warfare," according to a recent report by Kosta Tsipis in *The Bulletin of the Atomic Scientists.* In an effort to "explore the emerging weapons systems, the underlying technologies and some far-ranging implications of their deployment" Tsipis, a physicist and arms-control expert at the Massachusetts Institute of Technology, focuses on one new weapon now being developed jointly by the U.S. Navy and the U.S. Air Force: the long-range "cruise" missile.

Until recently, the report points out, available electronics technology did not provide the means for continuous long-range missile guidance. "Deployed guided weapons, such as the Soviet surface-to-surface naval cruise missiles, have ranges generally under 50 kilometers and are usually of the remotely guided type requiring line-of-sight contact between the target and the human operator or the guidance radar.... An exception of course are the intercontinental ballistic missiles (ICBM's), which are guided by inertial-guidance systems during the powered first few minutes of flight and then follow a ballistic trajectory completely determined by the local gravitational field and their velocity at the instant of thrust termination."

In contrast, the strategic version of the cruise missile currently under development in the U.S. would be continuously powered by a small turbofan jet engine and could fly at altitudes ranging from treetop level to 10 kilometers (about six miles) or more. It would have a range of some 2,500 kilometers (1,500 miles) and a miniaturized guidance system capable of delivering a nuclear warhead to a target with a probable error radius of only a few tens of meters. An extraordinarily versatile weapon, it could be launched from a fixed site, a mobile land vehicle, an airplane, a surface ship or the torpedo tube of a submerged submarine.

Actually, Tsipis writes, the particular long-range cruise missile in question is designed with two guidance systems: "an inertial navigation system aided by mid-course updating derived from terrain-matching and a secondary microwave radiometry guidance system. This combination of guidance systems can guide the missile to the target, recognize it and land on it.... Further improvements in all the guidance components are planned as the missile enters the testing stage sometime in 1977."

The prospective U.S. strategic cruise missiles are, in Tsipis' view, "undoubtedly only the first examples of future consecutive generations of increasingly sophisticated guided cruise missiles. The new electronics technology incorporated in these systems will certainly advance and with it will come new applications in weapons systems.... There are no technical obstacles to new versions of guided cruise missiles that may have longer ranges and supersonic speeds, that may be retargetable while in flight and that may incorporate elaborate countermeasures to frustrate countermissile defenses." Because a cruise missile has aerodynamic surfaces, Tsipis continues, it is "inherently a much more efficient device than a ballistic missile to carry large payloads.... Given the ability to guide these vehicles precisely to their target, it can be expected that their intrinsic advantages will prove attractive to the military and that their development will be pursued vigorously."

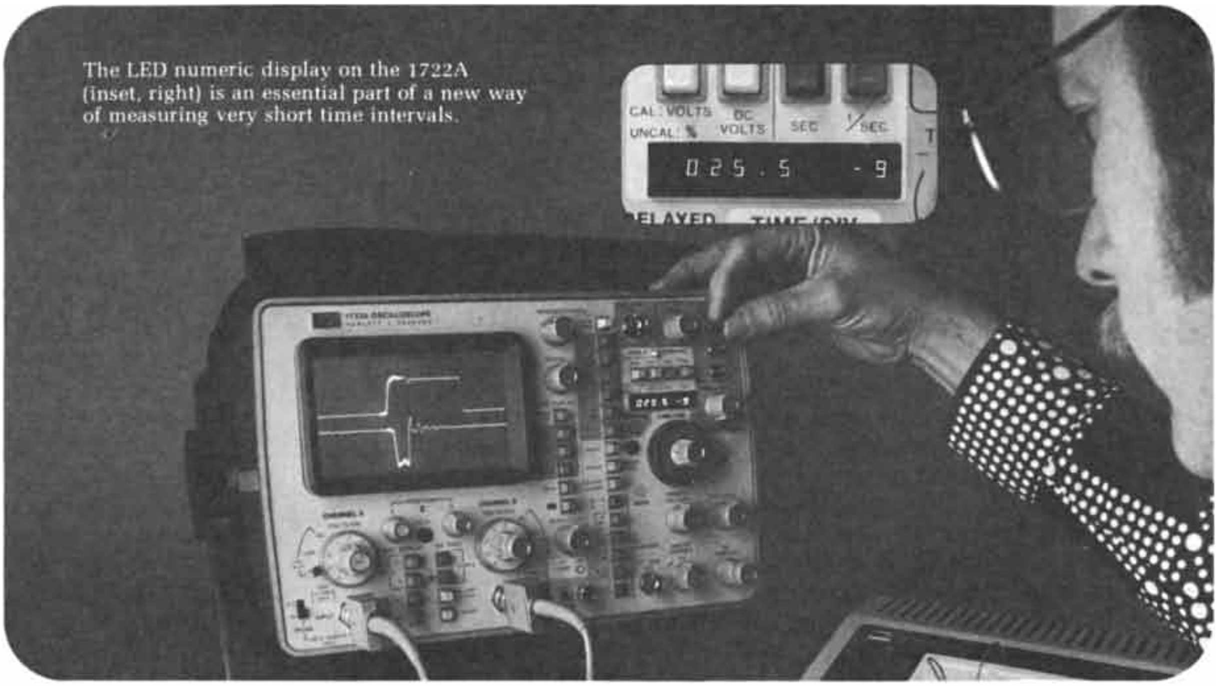
In spite of the many perceived advantages of the new weapons systems from the military viewpoint (the chief being their comparatively low cost of acquisition and operation), Tsipis believes "the total cost of a new weapons system should not be calculated only on the basis of the monies expended for its development, procurement and maintenance in operational readiness, but must be extended to include the overall economic, political and security costs induced by its deployment.... When this broader balance sheet is drawn up for the high-accuracy long-range cruise missile, it is seen that its development and procurement cost is only a small fraction of what the U.S. or any other nation would have to pay for its deployment."

The deployment of long-range cruise missiles by the U.S. would, Tsipis maintains, virtually force the U.S.S.R. to follow suit. Thus the same escalatory cycles that have characterized the arms race in ballistic missiles would be duplicated with cruise missiles. In the latter case, however, the escalation would be much more widespread, since many countries have the "technological infrastructure" to produce long-range cruise missiles. The end result of this emulative process would, as before, be to diminish the security of all countries.

In the final analysis, Tsipis states, "by far the most serious disadvantage of the long-range cruise missile, and especially the submarine-launched version, is its impact on arms-limitation efforts," since it would be "impossible" to verify an arms-control agreement aimed at limiting the proliferation of long-range submarine-launched cruise missiles once they are deployed. "In the future," he concludes, "the failure of the Vladivostok SALT II agreement to control the deployment of long-range strategic cruise missiles by the United States will have the most profound effect on the security of the two countries involved, and the rest of the world.... The accurate inexpensive cruise missile may prove to be the great equalizer among nations, but it may also prove to be the ultimate leveler of their cities."

The Fossil Reactors of Gabon

Two, and probably three, nuclear-fission reactors were assembled by geological processes about 1.7 billion years ago in a region of Africa that is now a part of the Gabon Republic. With enriched uranium 235 as a fuel and light water as a moderator, the natural reactors operated at power levels of several kilowatts more or less continuously for at least 100,000 years and possibly for



The LED numeric display on the 1722A (inset, right) is an essential part of a new way of measuring very short time intervals.

The first smart oscilloscope: an introduction to the "new measurement" technology.

Hewlett-Packard's "new measurement" technology radically changes the traditional relationship between man and machine. It does so by giving the machine some of the intelligence previously supplied by the human operator. It creates "smart" instruments that can monitor their own operations, detect and avoid procedural errors, perform all the necessary computations, and directly produce the desired final answer.

As a leading manufacturer of both measurement and computation instruments, we know that it's possible to make many kinds of "smart" instruments—right now—through a marriage of separate measurement and computation instruments. And we know that such a marriage results in unique advantages that far outweigh its cost.

The new HP 1722A oscilloscope is a recent case in point. Its development started, typically, with a choice of candidates for the marriage. For *measurement*, we chose our 1720A scope, a 1.3 nanosecond rise time, 275 MHz bandwidth, dual-channel instrument; and for *computation*, the digital microprocessor originally developed for the HP-35 pocket-sized calculator.

It was clear from the outset that the proposed instrument could not realize the full potential of microprocessor control if the scope were limited to state-of-the-art *single* delayed sweep capability. So the marriage was put off while our designers developed the technology for *dual* delayed sweep. In the 1722, this new two-dot system operates under microprocessor control to keep track of any two events automatically, whether they originate on the same or different channels.

With its combined measurement and computation capabilities, the 1722 is clearly in a class by itself, the first of a new generation of "smart" oscilloscopes. It can make time interval measurements more accurately than has ever before been possible with an oscilloscope. The 1722 also avoids the numerous errors that can creep into conventional scope measurements: it never misreads control settings, never misses events, and its automatic lock-out systems prevent most wrong interpretations.

The 1722's built-in computation provides final answers—directly, digitally, and automatically—for frequency, instantaneous voltage, and relative amplitude measurements as well as for dc voltage and time interval.

Priced at $4750*, the 1722A is ideally suited for clock phasing measurements in large computer systems. It also easily qualifies for less demanding applications where its speed, convenience, and automation more than justify its moderately higher cost.

The scientific calculator that "talks back" to you.

For some scientists and engineers, a hard-copy record of their calculations can save hours of recalculating and rechecking.

The HP-46 not only presents your calculations on its LED display, it also prints all your entries and labels them clearly with alphanumeric symbols—and then "talks back" to you as it performs the calculations. For example, when you calculate the mean and standard deviation of a series of numbers, the printer lists each entry with $\Sigma+$ symbol. After you push the $\bar{x},\Delta$ key, it prints the results in order, each calculation clearly labeled: the number of entries (#), the standard deviation (Δ), and the mean ($\bar{x}$).

Whenever you ask for it, the HP-46 prints the contents of its operational stack or its 9 addressable memories, each also clearly labeled. Should you make a logical error in data entry or call for an improper calculation, the HP-46 will print an error message and refer you to an explanation in the operator's manual.

The HP-46 performs arithmetic, statistical, transcendental, and trigonometric calculations, the latter in any of three angular modes. It instantly converts angles to degrees, polar coordinates to rectangular, and U.S. units of length, weight, or volume to metric. It also performs two-dimensional accumulations simultaneously for vector calculations, and adds or subtracts vector components.

The HP-46 features characteristics that you have come to expect from HP calculators: 200-decade range, floating decimal, 10 significant digits, and computer-based logic with 4-register operational stack.

The HP-46 costs $675* including owner's manual, printer paper, and carrying case.

For more information on these products, write to us. Hewlett-Packard, 1505 Page Mill Road, Palo Alto, California 94304

Mail to: Hewlett-Packard, 1505 Page Mill Road, Palo Alto, CA 94304.

Please send me further information on

( ) HP 1722A Oscilloscope
( ) HP-46 Scientific Desk-Top Calculator

Name ______________________ Title ______________
Company ______________________________________
Address ______________________________________
City ________________ State __________ Zip __________

*Domestic USA prices only.

00545

HEWLETT hp PACKARD

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

more than a million. At least as surprising as the existence of the "fossil" reactors is the fact that the geological site has been so little disturbed by tectonic processes in the intervening aeons that French investigators have been able to establish the precise dimensions of the reaction centers (*foyers de reactions*) and the general physicochemical factors that brought the reactors to criticality and controlled their power levels for millenniums on end.

The existence of the natural reactors first came to light three years ago at the French gaseous-diffusion plant at Pierrelatte, when a laboratory preparation of uranium hexafluoride, to be used as a secondary control standard, showed a slight deficit in the expected content of uranium 235: a value of .7171 percent instead of the normal and universal value of .7202. Workers tracking down the discrepancy eventually found that more than 700 tons of uranium shipped from a mine north of Oklo in Gabon was depleted in U-235, representing a total deficit of 200 kilograms of the desirable isotope. A reanalysis of core samples taken some years earlier revealed one sample with a U-235 content as low as .440 percent. The only plausible hypothesis was that some of the U-235 at the site of the mine had been consumed as fuel in a natural nuclear reaction.

The hypothesis was quickly confirmed by identifying the zones of greatest U-235 depletion and examining the ratios of particular isotopes that would have appeared as fission fragments if the nuclear reaction had actually taken place. The ratios carried the proper "signature." In the course of this study one sample was found in which the U-235 reached a record low value of .296 percent. In general the two principal reaction centers showed an impoverishment of U-235 proportional to the total uranium content up to a value of about 30 percent. In still richer samples there was less depletion of U-235, indicating that the optimum value for a sustained chain reaction was a uranium content of between 20 and 25 percent. Overall nearly 500 tons of uranium in the two fossil reactors was strongly depleted in U-235. Taking account of the fact that a certain fraction of the abundant isotope U-238 would have been converted into fissionable plutonium 239, the total production of energy in the two fossil reactors is estimated to have been about 10,000 megawatt-years, equivalent to the output of a 10-kilowatt reactor running for a million years.

Two things were essential for the chain reaction to have occurred: a sufficient enrichment of U-235 above the normal value and a moderator, such as water, to slow down the fast neutrons released by spontaneous fission so that they would support a chain reaction among other nuclei of U-235. Since U-235 spontaneously disintegrates with a half-life of .7 billion years (giving rise to an isotope of thorium), one can calculate that 1.7 billion years ago the Gabon deposits contained about 3 percent U-235, or about the level of enrichment commonly found in the fuel of commercial nuclear reactors. There is also geological evidence that the uranium, in the form of its oxide, was well mixed with a variety of hydrated minerals, so that the reaction centers contained about 15 percent water, at least some of it in liquid form, saturating the mineral bed. The ratio of water to uranium oxide was about the same as it is in a typical man-made light-water reactor. Finally, the reaction centers had to be reasonably free of any well-known nuclear-reaction "poisons," such as boron, which have a voracious appetite for neutrons and would have made a chain reaction impossible.

A volume of something less than a cubic meter is needed to sustain a chain reaction in a water-moderated reactor with a fuel containing 3 percent U-235. Each of the two fossil reaction centers contains at least 300 times the minimum volume. The French investigators conjecture, therefore, that each center consisted of a number of weakly coupled reactors, probably not all functioning at the same time. The details of the control mechanism are still a matter of speculation, but it is clear that variations both in the water content and in the concentration of nuclear-reaction poisons, as the chain reaction proceeded, must have played a major role. A full description of the fossil reactors appears in *Bulletin d'Informations Scientifiques et Techniques*, published by the French Atomic Energy Commission.

Topography of the Cell

The plasma membrane that surrounds all cells contains a large and diverse population of protein molecules. These proteins have been considered free to migrate laterally in the membrane, so that one would expect them to be distributed randomly over the surface of the cell. It now appears that in the membranes of some cells the distribution of proteins is not random at all. Under some circumstances the molecules assume distinctly ordered arrangements, and they can be shepherded into those patterns by agents both inside the cell and outside it. The proteins may even be segregated by type.

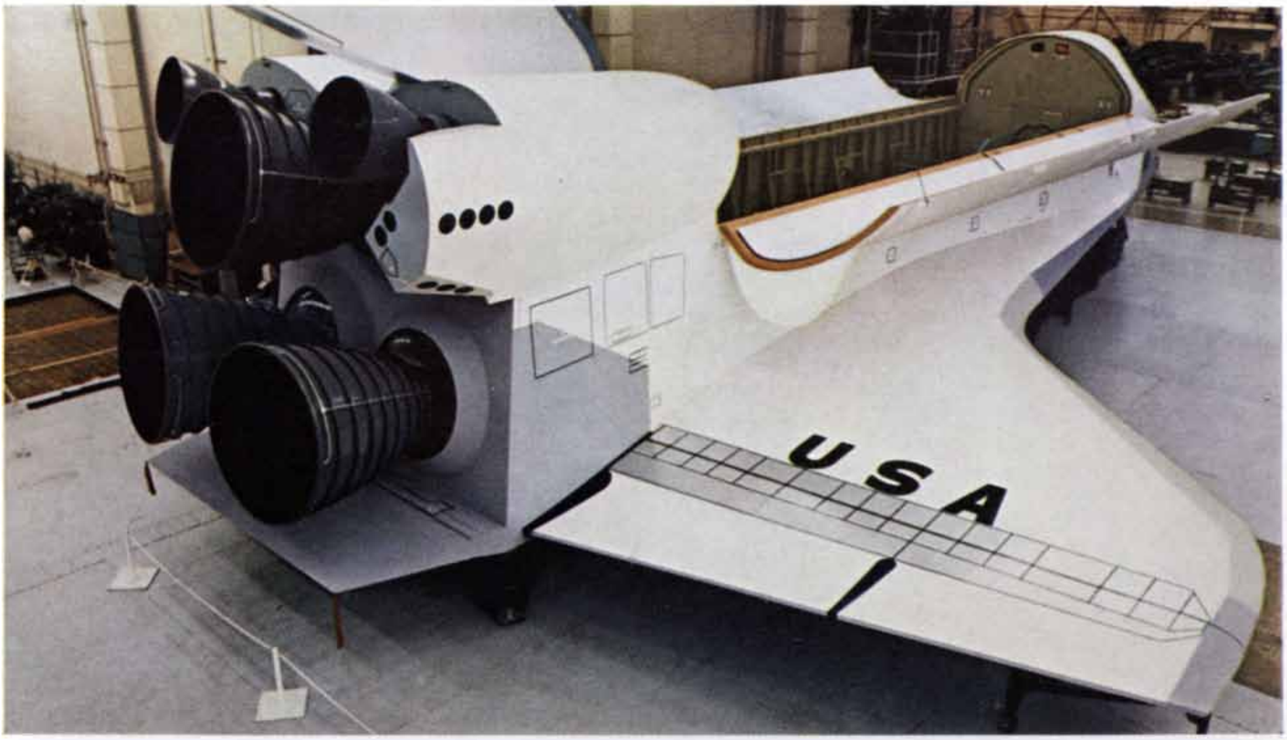
The plasma membrane consists of a double layer of phospholipids: molecules that have a hydrophilic "head" and a hydrophobic "tail." The tails extend from each side into the interior of the membrane; the heads are immersed in the aqueous medium. Proteins and glycoproteins (which have sugar units attached to them) are interspersed throughout this matrix; they are usually arrayed at the surface of the outer layer, but some apparently penetrate both layers to connect the cytoplasm of the cell with the external medium. For the protein molecules the phospholipid matrix should behave as a fluid, that is, the proteins should be able to diffuse freely through it. Recent experiments performed by Richard D. Berlin and his colleagues at the University of Connecticut School of Medicine and by other investigators indicate that the movement of the proteins is in some way restrained or directed.

Berlin's investigation of membrane proteins relied on two classes of substances derived from plants: the lectins and the alkaloids. Lectins bind to the carbohydrate portion of certain glycoproteins on the cell surface. By incubating a cell with a lectin such as concanavalin A and then binding a labeling molecule to the lectin, Berlin observed the distribution of the lectin-receptor molecules. In some cells the pattern does appear to be random. In others, however, including lymphocytes and cells that have been transformed, or made cancerous, by a carcinogenic virus, the receptors quickly aggregate into small clusters. Alkaloids such as colchicine further modify the distribution: the clusters themselves aggregate, retreating from the edges of the cell and forming a large central patch.

The lectin-receptor proteins also assume a nonrandom distribution during phagocytosis, the process by which a cell engulfs foreign particles. In phagocytosis a portion of the plasma membrane surrounds the particle and is eventually pinched off to form a vacuole inside the cell. Berlin and his colleagues have found that a disproportionate number of lectin-binding sites are removed with the excised area of membrane. Surprisingly, they observed exactly the opposite effect when they investigated a different group of surface proteins, those concerned with the transport of substances across the

DP Science Dialogue

Notes and observations from IBM which may prove of interest to the scientific and engineering communities.



A full scale model of the Space Shuttle orbiter, measuring 122 feet end to end, developed by Rockwell International in Downey, California. Computer simulations played a major role in the Shuttle's design.

From Space Shuttles to Microelectronics

Determining the best re-entry path for a 200,000 pound vehicle approaching earth at more than 12,000 miles per hour . . . Modifying the layout of a tiny microelectronic device to allow for changes in logic capability.

Those are a sample of the kinds of design challenges encountered by engineers at various divisions of Rockwell International Corporation.

To help solve their design questions, engineers at Rockwell are using an online computer capability called the Time Sharing Option. With TSO, many engineers can use terminals located in their own divisions and communicate directly with Rockwell's four IBM System/370 Model 168 computers in Downey, California. An engineer can receive a response to his inquiry within a few seconds, almost as if the computer were dedicated totally to him.

Simulating Flight Conditions

"Interactive computing under TSO has significantly reduced the time needed to develop and debug many of the computer programs needed to simulate the U.S. Space Shuttle's in-flight conditions," says Dr. Joseph F. Gloudeman, director of Management Systems Engineering and Computing Services for the company's Space Division.

"We have also been able to take better advantage of the Continuous System Modeling Program (CSMP), which allows us to simulate a wide range of flight characteristics," adds Ray Brown, manager of engineering applications. "It has proved to be an extremely valuable tool.

"For example, CSMP helps us determine the best angle for the Shuttle's re-entry. That calculation involves a wide variety of data such as the heat caused by atmospheric drag, orbital velocity, and aerodynamic characteris-

(Continued on next page)

Solving Musical Mysteries by Computer

Plagiarism didn't concern sixteenth-century Italian song writers. They didn't think twice about borrowing each other's best tunes. And nobody minded.

That's the way it was back then. But today, musicologists want to trace precisely who borrowed what from whom. And the computer is helping them do this, just as it is helping them in many other kinds of musical research.

The two scores below show how an obscure musician, Nicola Broca, borrowed a melody from a better-known composer, Josquin des Prés. Nicola even went so far as to twist Josquin's words, and turn a sacred song into a ditty of disappointed love. Out of some 40,000 different tunes, the IBM System/370 Model 158 at the State University of New York in Binghamton selected these two, because they had such similar melodic form. Dr. Harry Lincoln, Chairman of the Department of Music, was then able to compare the printouts of the opening themes, scrutinize publication dates and trace the borrowing.

Musicologists like Dr. Lincoln have to cope with such a vast repertory they just couldn't tackle much research of this kind without computer help. Of course, they must have a way to put a musical score into computer-readable form. And that's why so many musicologists today are using a coding system called DARMS.

DARMS, Digital Alternate Representation of Musical Scores, was developed by Stefan Bauer-Mengelberg, a visiting professor at Binghamton who is also a staff member at the IBM Systems Research Institute. He says, "Now a musicologist can take any piece of music in standard notation and transcribe it into a code for entry into a computer."

Since musicologists have a way to tell the computer precisely what a composer has scored, they can now process a formidable volume of data. In fact, in many universities today music departments are among the biggest computer users.

With DARMS as their tool, musicologists can develop programs to analyze a composer's use of harmony, rhythm and counterpoint. With this knowledge they can develop a theory about his style, and study how it evolved. They can even attempt to determine when Bach, for example, composed a particular work.

"Once you know enough about composers' stylistic techniques," says Bauer-Mengelberg, "much music that was once dubbed 'anonymous', or was wrongly attributed, can be ascribed to the right composer. This is especially important in early music, where title pages from folios are often lost."

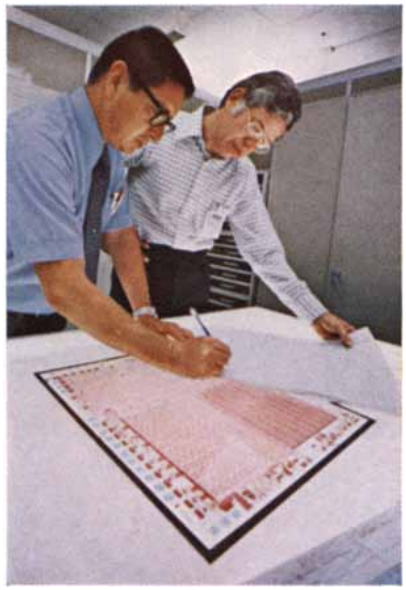
Looking to the future, Bauer-Mengelberg speaks of how the computer could be used to print musical scores: "Now that a way has been found to make music machine-readable, we hope the day is not far off when we will be able to use the computer in the preparation of master plates for music printing."

Two songs. Two composers. Similar melodies. The computer helps solve the mystery: who borrowed from whom.

Microelectronics...

(Continued from preceding page)

tics," explains Brown. "An engineer sitting at a terminal can change any one of these variables and determine almost immediately the impact of the change on vehicle performance."



At the Microelectronic Device Division, Bob Larsen (right) and Gerry Lozano verify a set of color masks which represent photoplates used to manufacture microelectronic devices.

Completed designs are stored online in a different data base organized under the Information Management System (IMS). Other vital information, such as material and parts in inventory, production in progress and costs, is also stored online.

IMS, TSO and CSMP are all IBM program offerings.

"If we have to modify a design for a part which happens to be in production, we can go to a terminal, locate the part by typing in its identification number, and integrate the new specifications immediately," explains Al Barnett, manager of systems development.

Designing Microelectronics

At the Microelectronic Device Division, Bob Larsen, manager of computer-aided design, tells a similar story: "Rockwell programs, written in Assembly and PL/1 languages and developed under TSO, simulate device logic, perform circuit analysis, evaluate nodal speed-noise problems, and even produce our preliminary design layouts.

"Interactive computing keeps our engineers completely involved in the design process, allows them to correct errors on the spot, and evaluate more design alternatives in greater detail."

A New Tool for Telescopic Mirror Designers

Predicting the stresses that a telescope mirror undergoes when it swivels about in an observatory has long been a challenge for optical designers. Today the problem of stress and the resulting deformation is more crucial than ever before in the design of mirrors for large-scale earth orbiting telescopes. These mirrors must be able to sustain incredible structural and thermal demands. They operate under extreme space conditions, they must survive severe launch forces, yet their weight must be kept to a minimum.

At Perkin-Elmer, the company's IBM System/370 Model 158 is playing an ever-increasing role in the mathematical modeling of mirrors. George Hefferon, Systems Analysis Manager at the company's Optical Technology Division, explains: "With our new computer modeling capability we can tackle more design problems than ever before, especially in the critical area of deformation prediction."

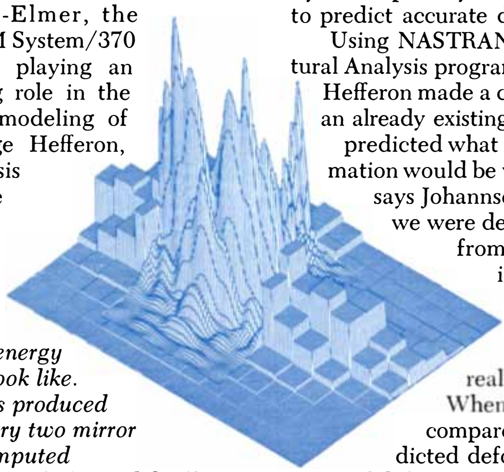
The company has been using the computer extensively in preliminary design for optical, structural and thermal analysis. But computer modeling of large space telescope mirrors is a relatively new computer application.

Recently Hefferon and structural analyst Peter Johannsen set out to verify the capability of computer models to predict accurate deformation.

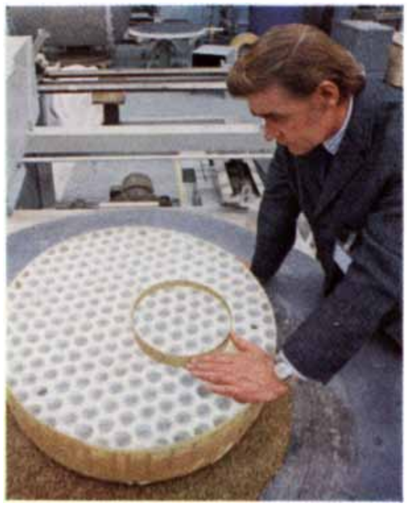
Using NASTRAN (a NASA Structural Analysis program) Johannsen and Hefferon made a computer model of an already existing mirror. "We then predicted what the mirror's deformation would be when under load," says Johannsen, "just as though we were designing the mirror from scratch. Using an interferometer, we later measured the actual deformation of the real mirror."

When both results were compared, computer-predicted deformation matched actual deformation within a millionth of an inch. "This proved conclusively the remarkable accuracy of computer modeling," says Johannsen.

Now Perkin-Elmer is using computer models in the design of large-scale telescope mirrors. Says Hefferon, "It is no exaggeration to say the computer has become a new design tool that has expanded our whole capability."



What an uncorrected star energy system might look like. This image was produced by an elementary two mirror system and computed on an IBM System/370 Model 158.



George Hefferon evaluates mirror optical quality. The honeycomb-like coring greatly reduces mirror weight by eliminating unnecessary glass.

Describing Molecular Structures with New Alchemy

Like their medieval counterparts, a group of modern chemists is using "Alchemy" to explain some very elusive transformations. They are doing something quite different from changing lead into gold. The new "Alchemy" is a computer program, developed at IBM's San Jose Research Laboratory, which enables chemists to describe changes in the electronic structure of small molecules during chemical reactions.

Dr. Douglas McLean (left) and Dr. Megumu Yoshimine (right) discuss their theoretical determinations of the structure of HO_2 with experimentalist Dr. Heinrich Hunziker (center) who, with Mr. Russell Wendt, first detected it.

"Computational chemists determine the properties of molecules in different electronic states (energy levels) by solving differential equations which describe the forces between the electrons and nuclei in the molecule," explains Dr. Douglas McLean, a researcher in the Molecular Dynamics Group at San Jose.

"By solving the equations for different configurations of the nuclei, we can predict molecular geometries," he adds. "These complex calculations, which require the execution of billions of computer instructions, make us heavy users of our Model 195 computer."

Theoretical descriptions produced by computations are adding a new dimension to chemistry. While experimental chemists can observe the initial reactants and final products of a reaction, often they cannot observe or explain the intermediary stages during which transient, highly reactive molecules exist only briefly. Using computation, however, all stages of a reaction are equally accessible.

Dr. McLean is currently using "Alchemy" to aid experimental detection of methylperoxyl (CH_3O_2) by making a prediction of how its lowest electronic excitation differs from the corresponding one in the related, already observed, hydroperoxyl (HO_2). These short-lived molecules are important in combustion reactions, in the smog cycle, and in biological oxidation processes.

"Computational chemistry complements the work of the experimentalist by adding data on molecules inaccessible to laboratory study, and in aiding interpretation of observations," adds Dr. McLean. "With it, our ability to understand such processes as combustion and pollution will be greatly enhanced."

Computers Essential in Nuclear Fuel Management

Fifty miles north of the University of Chicago's Stagg Field—where 33 years ago Enrico Fermi achieved the world's first controlled nuclear chain reaction—is Zion Station, one of the world's largest nuclear power plants. Capable of producing 2.1 million kilowatts of electricity, Zion is the newest of three nuclear stations operated by the Commonwealth Edison Company.

"Nuclear power accounted for about 30 percent of all the electricity generated by Commonwealth last year. Over the next decade, that figure is expected to increase to one-half," says Paul Fenoglio, manager of computer services.

The nuclear fission process—splitting apart Uranium 235 isotopes to release tremendous amounts of energy—is carefully controlled to produce the desired amount of power with the smallest quantity of fuel. Refueling, which requires plant shutdown, must be scheduled when peak-load demand is lowest.

Literally thousands of calculations are needed every day to run a nuclear plant efficiently. The computer is a useful tool for a safe and economical operation.

To help evaluate fuel utilization and to monitor the safety parameters within the reactor, Commonwealth uses an IBM System/370 Model 158 Multiprocessor. The Model 158 MP, two computers operating as one system, is linked via leased telephone lines to computers monitoring data in each nuclear plant.

Unlike fossil-fueled units in which fuel consumption can be easily measured in terms of weight and volume, nuclear fuel burn-up must be computed from data taken from sensor devices located inside the reactor. The nuclear fuel configuration at Zion consists of thousands of 12-foot long, half-inch thick rods, each filled with slightly enriched uranium pellets. The rods are bundled into 193 fuel assemblies arranged in a core shape approximating a cylinder. The core consists of several regions.

The greatest fission activity occurs in the core center. Consequently, fuel in this region is depleted most rapidly. To insure uniform usage, fuel assemblies from outer regions are periodically shifted closer to the center. New assemblies, filled with U-235 pellets of a slightly higher isotopic concentration than those remaining in the core, are added to the outermost regions.

"To find out the best time for refueling, and determine which assemblies must be replaced, we must have detailed data on core power distributions, isotopic concentrations and safety parameters for each region in the core," explains Dr. William Naughton, nuclear engineer in Commonwealth's reactor analysis department.

That information is gathered by six moving sensors which make axial traverses through the centers of 58 assemblies distributed throughout the core. The final result is a full-core flux map, or a "scientific picture" of the power distribution of the reactor, which is digitized by a computer at the plant and telephoned via a data-link to the Model 158 MP for analysis.

"The flux map also helps us determine peaking factors, which are measures of local to average power," notes John Rawley, principal engineer in charge of analysis. "Periodic determinations of peaking factors are compared to license limits for assurance of safe operation.

"Data from the flux map can be analyzed on the Model 158 MP within a half-hour compared to 3 or 4 man-weeks without the computer. The results can be stored in our data base for future use by fuel specialists, design engineers and economists."

With the computer, power distribution and performance parameters can be compared against safety guidelines, costs, and future demand for electricity. This means greater accuracy and economy in planning nuclear power for the future.

In the control room of Zion Station, an operator monitors the voltage regulator on a 1,040 megawatt generator.

Giant steam turbines at Zion Station.

DP Science Dialogue is concerned with topics which may prove of interest to the scientific and engineering communities. Your comments and suggestions are welcome. Just write: Editor, DP Science Dialogue, IBM Data Processing Division, 1133 Westchester Ave., White Plains, N.Y. 10604.

IBM
Data Processing Division

membrane. During phagocytosis fewer transport proteins are engulfed than would be expected if the molecules were distributed uniformly over the cell surface. The effect of colchicine on phagocytic cells was also a surprise. Whereas it enhances the movement of lectin receptors in cells treated with concanavalin A, it apparently suppresses the movement of both lectin-binding sites and transport proteins during phagocytosis.

It is the seemingly paradoxical action of colchicine that provides the crucial evidence for a theory explaining these observations. Colchicine is known to disrupt microtubules, hollow fibers constructed as polymers of a soluble protein called tubulin. Microtubules have been identified as fundamental units in a number of subcellular structures, particularly those that are motile. They power cilia and flagella, for example, and they are instrumental in separating chromosomes during mitosis. Berlin and his colleagues have proposed that microtubules may be responsible as well for the directed movement of proteins in the plasma membrane.

The hypothesis that the membrane proteins are linked to an underlying network of microtubules can account for the general features of the experimental results. Concanavalin A, by binding many molecules together, tends to aggregate the lectin-receptor proteins; the microtubules, however, tend to anchor them in place. Colchicine, by destroying the microtubules, increases the mobility of the proteins and makes possible larger aggregations. In phagocytosis, on the other hand, the microtubules serve not simply to restrain the protein molecules but actually to move them. The lectin receptors are propelled toward the area of engulfment and the transport proteins away from it. Once again colchicine disrupts the process by destroying the locomotive mechanism.

That microtubules are the active elements in the transport of membrane proteins has not been proved, but the hypothesis is supported by several lines of evidence. That colchicine impedes the directed movement is in itself strongly suggestive, since it and certain other alkaloids are known to prevent the polymerization of tubulin. It has also been shown that restriction of the cell's metabolism of cyclic nucleotides, a procedure that prevents the polymerization of microtubules, eliminates the restraints on the movement of surface proteins. Finally, microtubules have been observed by electron microscopy in the region immediately underlying the plasma membrane, and Berlin has recently demonstrated by a technique in which membranes are labeled with one fluorescent dye and tubulin with another that membranes and microtubules can make intimate contact.

The Tornado Plague

During the spring and early summer months tornadoes are likely to occur in the region between the Gulf of Mexico and Lake Ontario. On April 3 of last year a mass of warm, moist air moved into that region. It encountered a center of extreme low pressure that was coming from the west; conditions were right for severe thunderstorms and possibly tornadoes. The first tornado struck at 1:10 P.M. By 5:20 A.M. the next day a total of 148 tornadoes had swept across the countryside and through towns and cities from Laurel, Miss., to Windsor, Ont., killing 315 people and injuring 5,484. In terms of sheer scope and the number of storms it was the largest tornado outbreak on record.

Immediately after the outbreak T. Theodore Fujita of the University of Chicago and his colleagues organized aerial and ground surveys. They mapped the intensity and the path of each tornado and compiled data on the damage they did.

At the height of the outbreak 15 tornadoes were on the ground simultaneously. Fujita and his co-workers found that 74 percent of the fatalities occurred in houses and buildings, 17 percent in mobile homes, 6 percent in automobiles and 3 percent among people en route to shelter. There were more deaths in rural areas than in cities, and more women and children were killed than men.

Fujita points out that most of the deaths were due to head injuries. "When you know a tornado is coming," he says, "your first action should be to find shelter. The basement is the first choice, but avoid locations near cinderblock walls, which may collapse. Even in a basement one should take shelter under a table for added safety." He also recommends wearing a hard helmet if one is available. A good second choice for refuge is a bathroom or any small interior room.

In the outbreak of April 3 and 4 the tornado with the longest track went through Monticello, Ind. Its damage path was 121 miles long; it killed 19 people and injured 362. The largest human toll was in Xenia, Ohio, where 34 were killed and 1,150 injured. The Windsor tornado crossed the international boundary twice, from the U.S. to Canada and back to the U.S. An unexpected finding was that the tornadoes produced continuous damage paths up and down steep slopes, across mountaintops and through deep gorges.

The English Galileo

At about the same time that Galileo began to investigate the law of the parabolic trajectory and to turn his telescope toward the heavens an unknown contemporary of his in England was doing precisely the same thing. The English mathematician and astronomer was Thomas Harriot, who among other contributions gave algebra its modern notation and discovered the law of refraction. In a recent issue of *History of Science* D. T. Whiteside of the University of Cambridge, reviewing the book *Thomas Harriot: Renaissance Scientist*, edited by J. W. Shirley, describes Harriot as "Britain's greatest mathematical scientist before Newton."

Harriot was born in Oxford in 1560, four years before Galileo, and received his bachelor of arts at the University of Oxford in 1580. He became scientific adviser to Sir Walter Raleigh, who sent him to the newly founded colony of Roanoke Island in what is now North Carolina. As one result of his visits Harriot brought tobacco plants back to England. He also wrote a promotional booklet titled ***A Briefe and True Report of the New Found Land of Virginia*** in an effort to attract settlers. It was the first report by an Englishman describing the Indians and the natural history of the region around the settlement.

Among his contributions to algebra Harriot introduced the signs for greater than (>), less than (<) and the raised dot (·) to signify multiplication. In optics he discovered the sine law of refraction known as Snell's law, which was later attributed to the Dutch mathematician Willebrod van Roijen Snell. Harriot also made telescopes in the same years Galileo did, and he used them to observe the moon, sunspots, comets and the satellites of Jupiter. In mathematical physics Harriot investigated the ballistic path of a projectile under the influence of gravity a decade before Galileo did. He concluded that the path was a parabola.

Why, then, is Harriot comparatively obscure? He did not publish his work during his lifetime, and it was not until a century and a half after his death that some 7,000 pages surviving from his notes were discovered. Harriot, as Whiteside puts it, "chose, for a complex of possible reasons at whose relative weight we can now only guess, not to communicate... any account [of his work] to his contemporaries at large."

Electron-Positron Annihilation and the New Particles

Energetic collisions between electrons and positrons give rise to the unexpected particles discovered last November. They may help to elucidate the structure of more familiar particles

by Sidney D. Drell

When matter and antimatter are brought together, they can annihilate each other to form a state of pure energy. A fundamental principle of physics demands that the reverse of that process also be possible: A state of pure energy can (quite literally) materialize to form particles of ponderable mass. When the matter and antimatter are an electron and a positron, the state formed by their annihilation consists of electromagnetic energy. It is a particularly simple state, since electromagnetism is described by a well-tested theory and is believed to be understood. For some time physicists have been eager to learn just what kinds of particles are created when an electron and a positron collide at high energy. During the past two years several experiments have provided a preliminary view of the annihilation process; the results have annihilated expectations. It is as the British poet Gerald Bullett described the arrival of spring:

> *Like a lovely woman late for her appointment*
> *She's suddenly here, taking us unawares,*
> *So beautifully annihilating expectation.*

The discoveries are the most startling and exciting to emerge from high-energy physics in a decade or more.

One reason for the great interest in these experiments is that they provide a means of testing a central concept of modern particle physics: the notion that the "herd" of supposedly elementary particles discovered during the past 25 years may actually be assemblages of only a few structureless entities that are truly fundamental. These constituent particles have been named quarks. Different versions of the quark theory make different predictions about what is to be expected in the aftermath of an electron-positron annihilation, and it was hoped that the experiments would help to determine which version is the correct one.

As it turned out, the results of an initial series of experiments in 1973 and early 1974 were not in accord with any of the predictions. Then, last November, as the measurements were being repeated and refined, two very massive particles were unexpectedly discovered. By coincidence the discovery of the first of the particles was announced simultaneously by physicists at two laboratories studying quite different reactions.

Four Forces

The existence of the new particles is in itself a surprise, but even more remarkable is their extraordinary stability. Although they decay to more familiar, less massive particles in a period that by conventional standards is very brief, their lifetime is about 1,000 times longer than that of other particles of comparable mass. This exceptional stability suggests that the new particles are fundamentally different from other kinds of matter. As yet their nature has not been satisfactorily explained, and their significance remains a subject of lively speculation. Theories abound and physics is in a state of great ferment, but we cannot be sure where the particles fit into the scheme of things.

Subatomic particles can be classified in broad families according to the kinds of interactions they participate in, or, as it is often put, according to the kinds of forces they "feel." The forces considered are the four fundamental ones that are believed to account for all observed interactions of matter: gravitation, electromagnetism, the strong force and the weak force.

In the everyday world gravitation is the most obvious of the four forces; it influences all matter, and the range over which it acts extends to infinity. For the infinitesimal masses involved in subatomic events, however, its effects are vanishingly small and can be ignored.

The electromagnetic force also has infinite range, but it acts only on matter that carries an electric charge or current. The photon is the quantum, or carrier, of the electromagnetic force, and when two particles interact electromagnetically, they can be considered to exchange a photon or photons. In the classification of particles the photon is in a category by itself: it has no mass and no charge and it does not participate in either strong or weak interactions.

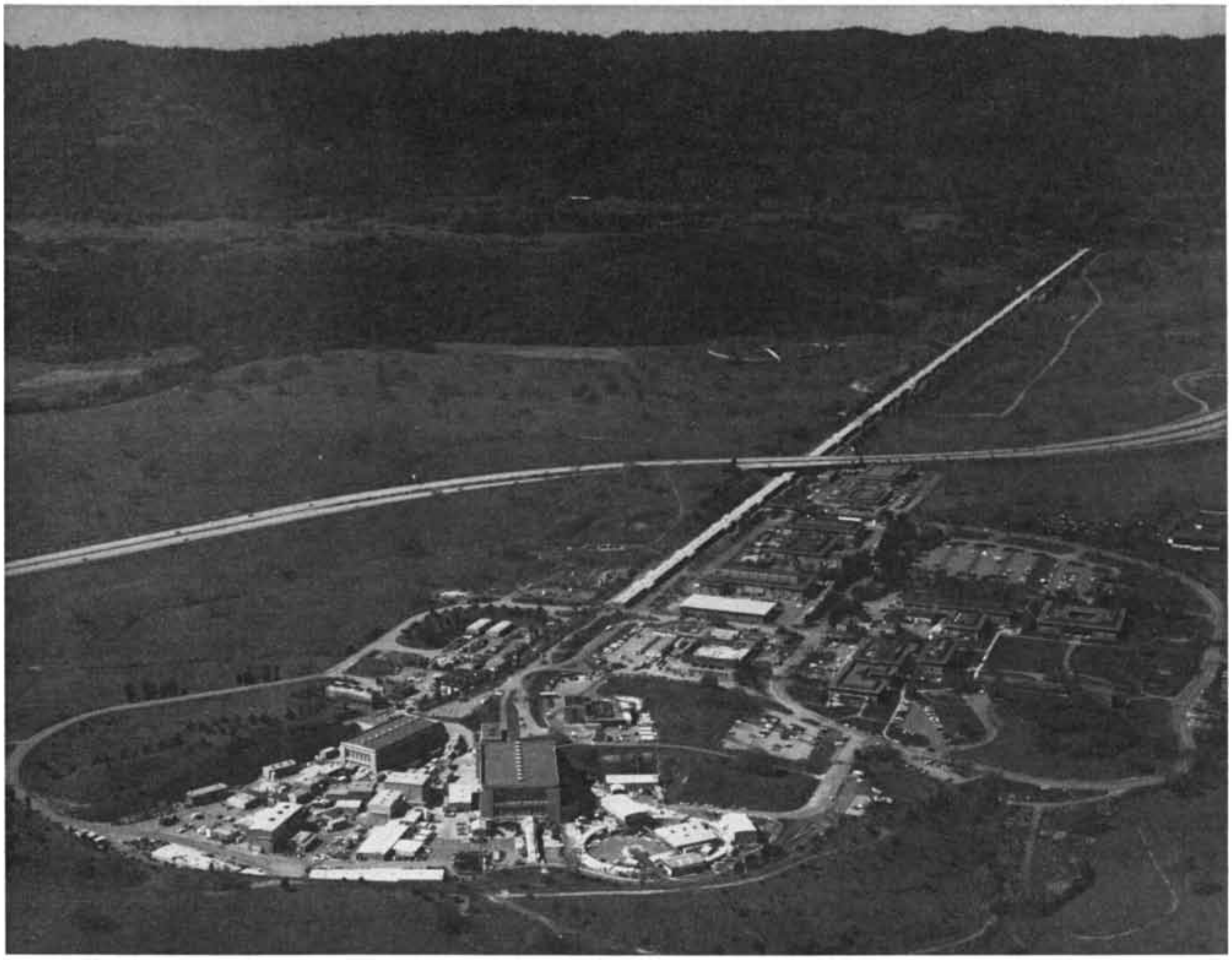
The strength referred to in the names of the strong and the weak interactions is related to the rate at which the interactions take place. The strong force has a short range: its effects extend only about 10^{-13} centimeter, or approximately the diameter of a subatomic particle such as a proton. When two particles that feel the strong force approach to within this distance, the probability is very high that they will interact, that is, they will either be deflected or they will produce other particles. In contrast, particles that interact electromagnetically are 10,000 times less likely to interact under the same circumstances. If the strongly interacting particles pass each other at nearly the speed of light (3×10^{10} centimeters per second), then they must interact during the 10^{-23} second they are within range of each other. That is the characteristic time scale of the strong interactions.

Compared with the strong force, the weak force is feeble indeed: for collisions at low energy it is weaker by a factor of about 10^{13}. At higher collision energies the strength of the weak force increases, but even at the highest energies yet studied experimentally it is weaker than the strong force by a factor of about 10^{10}. Moreover, the range of the weak force is at most a hundredth that of the strong force. Two particles must approach to within 10^{-15} centimeter in order to feel the weak force, and even at that range the probability that they will interact is less than one in 10^{10}.

All particles except the photon are classified according to their response to these two forces. Those that feel the strong force are called hadrons; those that do not feel the strong force but do respond to the weak force are called leptons. Particles belonging to these two families have quite different properties.

The hadrons are subdivided into two classes called baryons and mesons. The baryons include the familiar proton and neutron (and it is the strong force that binds these particles together in atomic nuclei). The mesons include such particles as the pion and the kaon; they are generally less massive than the baryons, but they are all more massive than the leptons.

As a result of discoveries made over the past two decades the baryons and the mesons have become very large families of particles; there are in all more than 100 known hadrons, most of them massive and unstable. It was in an effort to explain this great proliferation of particles that the quark hypothesis was introduced independently in 1963 by Murray Gell-Mann and by George Zweig, both of the California Institute of Technology. The quark model posits that the unstable particles are excited states of the stable ones. Baryons are thought to consist of three quarks bound together and mesons to consist of a quark and an antiquark. Just as an atom can enter an



TWO-MILE-LONG ACCELERATOR at the Stanford Linear Accelerator Center (SLAC) was employed to generate high-energy electrons and positrons in the experiments that led to the discovery of the new particles. The main beam of the accelerator (extending under the highway) propels electrons into a target of tungsten, a third of the way down the accelerator. Some of the electrons collide with tungsten nuclei, creating electron-positron pairs. The positrons and the remaining electrons in the main beam are then further accelerated and injected into a storage ring, where the particles and antiparticles interact. The ring, called SPEAR, is to the right of the largest buildings. In it beams of electrons and positrons are confined in a single evacuated chamber. The beams circulate in opposite directions, guided by a magnetic field, and pass through each other twice each revolution. The two buildings straddling the ring enclose detectors that surround the regions where particle-antiparticle annihilations take place. The ring is about 250 feet in diameter.

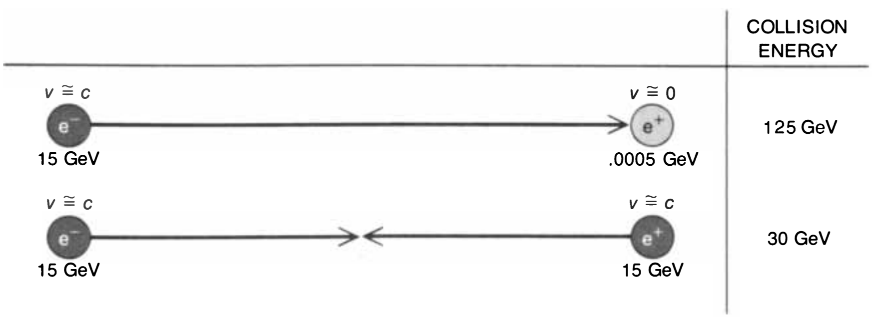
		PARTICLE	SYMBOL	MASS	CHARGE	SPIN	LEPTON NUMBER	MU-NESS	BARYON NUMBER	LIFETIME
		PHOTON	γ	0	0	1	0	0	0	STABLE
	LEPTONS	ELECTRON	e^-	.5	−1	$\frac{1}{2}$	+1	0	0	STABLE
	LEPTONS	POSITRON	e^+	.5	+1	$\frac{1}{2}$	−1	0	0	STABLE
	LEPTONS	ELECTRON NEUTRINO	ν_e	0	0	$\frac{1}{2}$	+1	0	0	STABLE
	LEPTONS	ELECTRON ANTINEUTRINO	$\bar{\nu}_e$	0	0	$\frac{1}{2}$	−1	0	0	STABLE
	LEPTONS	MUON	μ^-	106	−1	$\frac{1}{2}$	+1	+1	0	10^{-6}
	LEPTONS	ANTIMUON	μ^+	106	+1	$\frac{1}{2}$	−1	−1	0	10^{-6}
	LEPTONS	MUON NEUTRINO	ν_μ	0	0	$\frac{1}{2}$	+1	+1	0	STABLE
	LEPTONS	MUON ANTINEUTRINO	$\bar{\nu}_\mu$	0	0	$\frac{1}{2}$	−1	−1	0	STABLE
HADRONS	BARYONS	PROTON	p	939	+1	$\frac{1}{2}$	0	0	+1	STABLE
HADRONS	BARYONS	ANTIPROTON	$\bar{p}$	939	−1	$\frac{1}{2}$	0	0	−1	STABLE
HADRONS	BARYONS	NEUTRON	n	939	0	$\frac{1}{2}$	0	0	+1	10^3
HADRONS	BARYONS	ANTINEUTRON	$\bar{n}$	939	0	$\frac{1}{2}$	0	0	−1	10^3
HADRONS	MESONS	PION	π^+	137	+1	0	0	0	0	10^{-8}
HADRONS	MESONS	PION	π^-	137	−1	0	0	0	0	10^{-8}
HADRONS	MESONS	PION	π^0	137	0	0	0	0	0	10^{-15}
HADRONS	MESONS	RHO MESON	ρ^+	750	+1	1	0	0	0	10^{-23}
HADRONS	MESONS	RHO MESON	ρ^-	750	−1	1	0	0	0	10^{-23}
HADRONS	MESONS	RHO MESON	ρ^0	750	0	1	0	0	0	10^{-23}
HADRONS	MESONS	PSI (3095)	ψ	3095	0	1	0	0	0	10^{-20}
HADRONS	MESONS	PSI (3684)	ψ	3684	0	1	0	0	0	10^{-20}

SUBATOMIC PARTICLES are classified according to the kinds of interactions in which they participate. The hadrons take part in "strong" interactions; the leptons do not; the photon interacts only electromagnetically. The hadrons are divided into mesons and baryons, which differ in their spin angular momentum and in other properties. The classification is reflected in quantum numbers such as lepton number, mu-ness and baryon number. The newly discovered particles, psi(3095) and psi(3684), are mesons. Their most perplexing property is their lifetime, which is 1,000 times longer than that of other particles of comparable mass, such as the rho meson.

excited state when the orbital configuration of its electrons is changed, so a hadron, by analogy, enters an excited state of higher energy (or mass) when the configuration of its constituent quarks is altered.

A further insight into the nature of hadrons was provided in the late 1960's by experiments in which high-energy electrons were scattered by protons and neutrons. The experiments were performed at the Stanford Linear Accelerator Center (SLAC) by a group of investigators from the Massachusetts Institute of Technology and SLAC under the direction of Jerome I. Friedman, Henry W. Kendall and Richard E. Taylor. Once again the results inspired an analogy with events on the atomic scale. In 1911 Ernest Rutherford investigated the scattering of alpha particles by atoms and was led to predict the existence of a massive nucleus within the atom. Similarly, the SLAC experiments revealed an internal structure within the individual hadron; the pattern of electron scattering suggested the existence of pointlike substructures, which were named partons. Partons and quarks are believed to be equivalent [see "The Structure of the Proton and the Neutron," by Henry W. Kendall and Wolfgang K. H. Panofsky; SCIENTIFIC AMERICAN, June, 1971].

The leptons are a much smaller class of particles than the hadrons. There are just four: the electron and its neutrino and the muon and its neutrino (and their four antiparticles). The electron has a mass of about .5 MeV (million electron



COLLISION ENERGY in electron-positron annihilations depends on the initial energy and momentum of both particles. When one particle is stationary and the other has a velocity near c (the speed of light in vacuum), the collision energy is only a small fraction of the energy expended in accelerating the moving particle. When two particles collide with equal and opposite velocity, all their energy is made available for the creation of new particles.

volts) and an electric charge of −1 unit. The muon has the same charge, but it is 207 times as massive as the electron. Both kinds of neutrino are without mass and charge. Because the electron and the muon are charged they can interact electromagnetically as well as by the weak force; the neutrinos feel only the weak force.

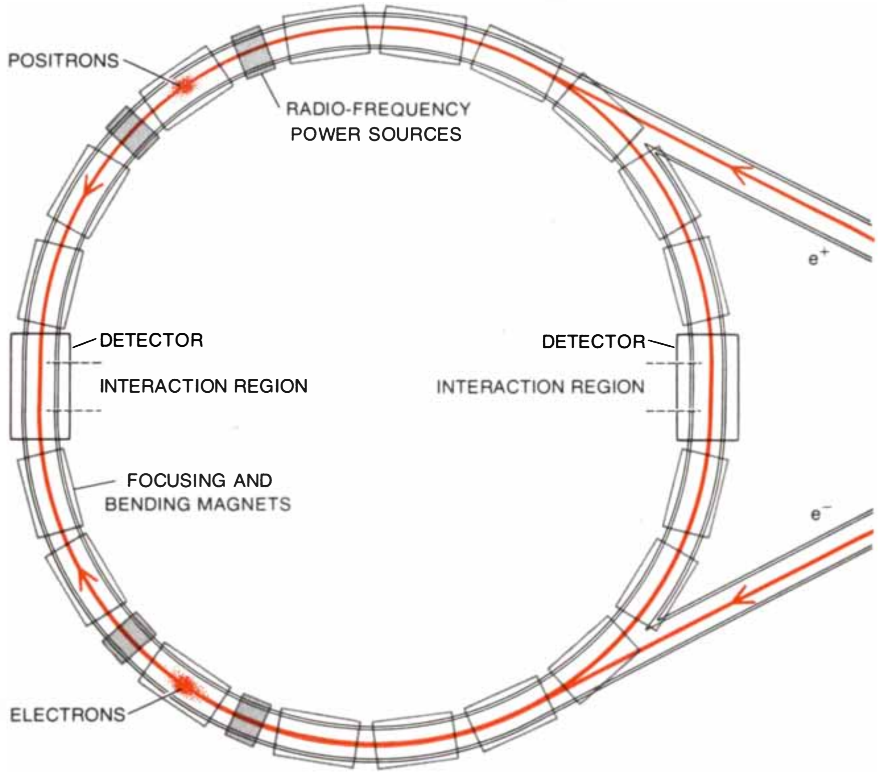
Among the leptons no spectrum of excited states comparable to that of the hadrons has been discovered. Nor have scattering experiments yielded any suggestion of an internal structure. The leptons are thus fundamentally different from the hadrons. They are not composed of quarks or partons but are themselves apparently pointlike. It is possible they are analogues of the pointlike constituents of hadrons.

Quantum Numbers

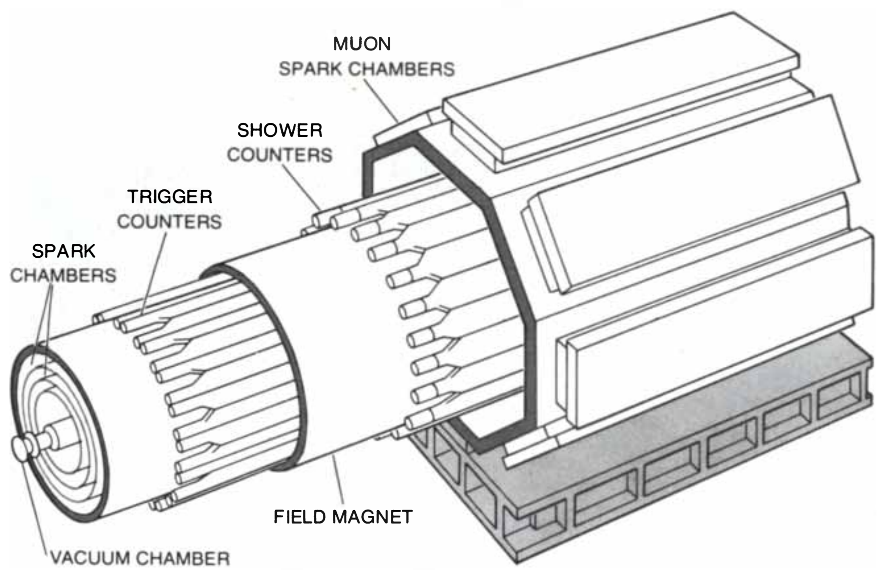
Physicists identify and describe subatomic particles by assigning them quantum numbers. Each number designates a property that is conserved, or left unchanged, when particles interact. Some quantum numbers, such as electric charge and spin angular momentum, refer to physical, measurable attributes of the particle. Others are more abstract; they denote family resemblances among particles and provide a valuable bookkeeping system for classifying particles and their interactions in algebraic form. Baryons and mesons, for example, are distinguished by a property called baryon number. Baryons are assigned a value of +1, antibaryons −1 and mesons 0; to say that baryon number is conserved is merely to say that baryons never turn into mesons. The least massive baryon, the proton, cannot decay into the less massive mesons or leptons because in such a process baryon number would be changed. The proton is therefore stable.

There are also conserved properties, or quantum numbers, of the leptons that ensure the stability of those particles. Lepton number is a quantity handled in the same way as baryon number. Its conservation prohibits the transformation of individual leptons into pure energy or into hadrons, just as hadrons are forbidden to become leptons. For instance, a positron has the same charge as a proton, and it can be accelerated until its mass is equal to that of a proton, but no positron has ever been observed to change into a proton.

In the same way a quantum number named mu-ness divides the leptons into two groups. The muon and the muon neutrino have mu-ness of +1, their anti-



COUNTERROTATING BEAMS of electrons and positrons in the SPEAR storage ring are each confined to a small "bunch" a few centimeters long. Their trajectories are determined by the magnetic field and their energy is maintained by the input of radio-frequency power; a single field guides both electrons and positrons, since the oppositely charged particles respond oppositely to it. Moving at nearly the speed of light, the beams pass through each other several hundred thousand times per second. Each beam can have an energy of up to 4 GeV (billion electron volts), so that total collision energies of up to 8 GeV can be achieved.



MAGNETIC DETECTOR completely surrounds one of the interaction regions at SPEAR. It consists of several layers of scintillation counters and spark chambers, which are here extended along their axis for clarity. Both kinds of detector produce an electrical impulse when a charged particle passes through them. Neutral particles cannot be detected. Information from the detectors is recorded electronically and employed to identify the particles, to reconstruct their trajectories and to determine which events are the result of an electron-positron annihilation and which are derived from other, extraneous sources. The momentum of a particle is determined by measuring its deflection in magnetic field of detector.

particles have mu-ness of −1 and the electron, the positron, the electron neutrino and the electron antineutrino all have zero mu-ness. An electron can transfer its mass and charge to more massive particles and become an electron neutrino, but because mu-ness is conserved it can never change simply into a muon or a muon neutrino.

Baryon number, lepton number and mu-ness are always conserved; certain other quantum numbers, however, are conserved only in some interactions. They are said to describe "approximate" conservation laws. Several properties, for example, are conserved in strong interactions but not in weak or electromagnetic ones.

In addition to the several conserved quantum numbers, energy and momentum are always conserved. Energy and momentum can be transferred from one particle to another, but the sum of the energies and momenta before an interaction must be exactly equal to the sum afterward.

In the annihilation of matter and antimatter the arithmetic of quantum numbers is particularly simple. The concept of antimatter was introduced in 1928 by P. A. M. Dirac, and it is now a firmly established principle that for every particle there exists an antiparticle. All the quantum numbers of an antiparticle are "opposite" those of the corresponding particle, that is, the sum of the quantum numbers of the particle and antiparticle is zero. The electron has lepton number +1, mu-ness 0 and charge −1; for the positron the corresponding values are −1, 0 and +1. For each quantum number the sum is zero.

When matter and antimatter meet, their quantum numbers, being opposite, simply cancel. Properties such as charge and lepton number are conserved, but they are also eliminated! Furthermore, from the state of energy thus created, particles having virtually any combination of quantum numbers can, at least in principle, be formed, as long as the sum of the quantum numbers of the products remains zero. In particular, a pair or more than one pair of particles and antiparticles can be created. There is only one constraint on the process: the energies of a colliding particle and antiparticle do not cancel, and in the annihilation energy must be conserved just as it would be in any other interaction. Momentum must also be conserved.

From these considerations it can be seen that some processes that are forbidden in interactions of ordinary particles are permissible in the annihilation of particle-antiparticle pairs. For example, although an electron is prohibited by the conservation of mu-ness from becoming a muon (except with the emission of an electron neutrino and a muon antineutrino, a rare process), it is entirely possible for an electron and a positron to annihilate each other and for a muon and an antimuon to materialize from the energy state formed. It is necessary only that momentum be conserved and that the initial particles have sufficient energy to account for the rest mass of the muon-antimuon pair [see "Electron-Positron Collisions," by Alan M. Litke and Richard Wilson; SCIENTIFIC AMERICAN, October, 1973].

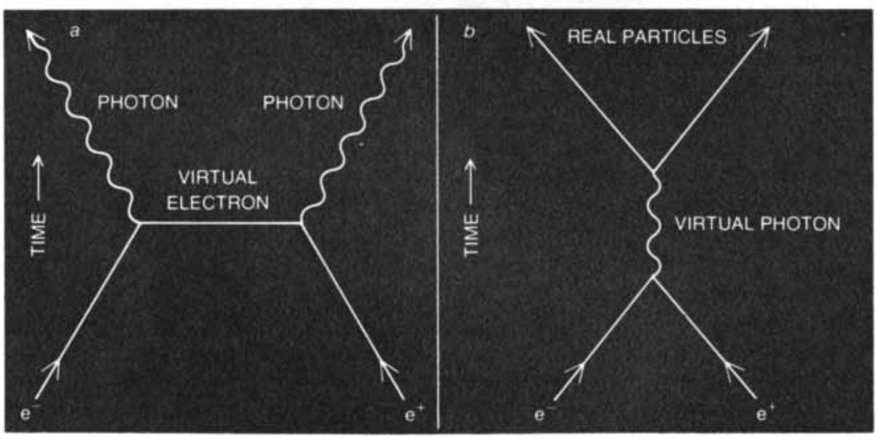
It is also possible for the annihilation of an electron-positron pair to give rise to hadrons. The reaction is written $e^-e^+ \rightarrow$ hadrons, where e^-e^+ represents the annihilating pair. It was during an investigation of this process at SLAC that the new particles were discovered.

The particles were found by a group of 35 physicists from the Lawrence Berkeley Laboratory and SLAC, led by Burton Richter, William Chinowsky, Gerson Goldhaber, Martin L. Perl and George H. Trilling. The less massive of the two particles was discovered simultaneously at the Brookhaven National Laboratory in experiments investigating a process that is the inverse of the reaction studied at SLAC. The Brookhaven experiments were performed by Samuel C. C. Ting and his colleagues from M.I.T. and Brookhaven; they studied the production of electron-positron pairs in collisions of hadrons. The detection of the particle by Ting's group was a technical tour de force: it is produced by their technique only once in 10^8 events.

Particle Storage Rings

Because hadrons are much more massive than electrons, hadrons can be created only in annihilations at high energy. In order to achieve the required collision energies it has been necessary to build machines known as particle storage rings. In these rings particles and antiparticles are made to circulate and then to collide head on [see "Particle Storage Rings," by Gerard K. O'Neill; SCIENTIFIC AMERICAN, November, 1966].

Ordinary particle accelerators can produce electrons or positrons of very high energy, but when such a particle strikes an electron in a stationary target, little of that energy is liberated in the collision. This effect is a consequence of the special theory of relativity, which states that when a particle acquires a large total energy, it also acquires a large total mass (according to the equation $E = mc^2$, where c is the speed of light in vacuum). Once an electron or a positron has been accelerated to about 1 GeV (billion electron volts) its velocity is within a few centimeters per second of the speed of light. Any additional energy imparted to it has negligible effect on its velocity and goes almost entirely toward increasing its mass. The two-mile linear accelerator at SLAC can produce positrons with an energy, or mass, of 15 GeV. The collision of such a particle with an electron at rest (which therefore has a mass of about .0005 GeV, 30,000 times



ANNIHILATION of an electron and a positron produces electromagnetic energy, or photons. The diagram depicting the event is called a Feynman graph (after Richard P. Feynman); distance is represented on one axis and time on the other. If momentum is to be conserved, two photons must be emitted (*a*); the photons, like the initial particles, have equal and opposite momentum and thus net momentum of zero. When only one photon is created (*b*), it must have large energy but zero momentum, a condition that is impossible for a real photon. The particle formed is called a virtual photon; it can never be detected and it quickly decays into real particles with zero total momentum. The production of two photons is said to take place through the exchange of another virtual particle: a virtual electron.

smaller than the mass of the positron) is like the collision of a charging elephant with a mouse. The mouse may bounce back or it may be crushed, but in either case the elephant's kinetic energy will be little changed by the collision. In the case of the stationary electron and the 15-GeV positron only about .125 GeV would be made available to create new particles. That is less than 1 percent of the energy supplied to the accelerated particle, and it is too little to produce even a single pion, the least massive hadron.

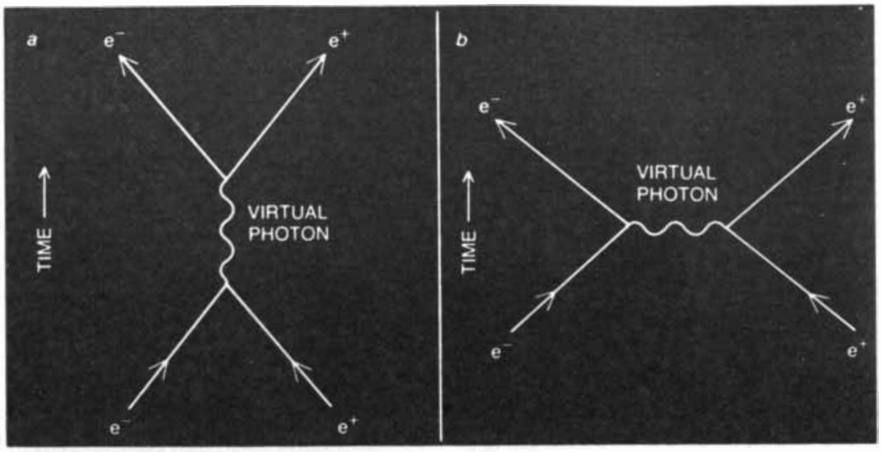
The result is quite different when an electron and a positron moving with equal velocity in opposite directions collide head on. In this case the sum of the momenta of the two particles is zero, and all their energy appears in the products of the annihilation. The collision is analogous to the head-on encounter of two charging elephants. In the storage-ring facility at SLAC, which is called SPEAR, electrons and positrons can be made to collide with energies of up to about 4 GeV each, for a total energy of 8 GeV, enough to generate a rich spectrum of hadrons. To produce such collision energies with stationary targets would call for a particle beam having an energy of 64,000 GeV; the accelerator required would be some 6,000 miles long.

The SPEAR storage ring was built under the direction of John R. Rees and Richter. In it "bunches" of electrons and positrons circulate in opposite directions in a single toroidal chamber about 250 feet in diameter. They are confined by a magnetic field, and their energy is maintained by the input of radio-frequency energy. The same magnetic field sustains both counterrotating bunches; because the particles are oppositely charged they react oppositely to it [*see top illustration on page 53*].

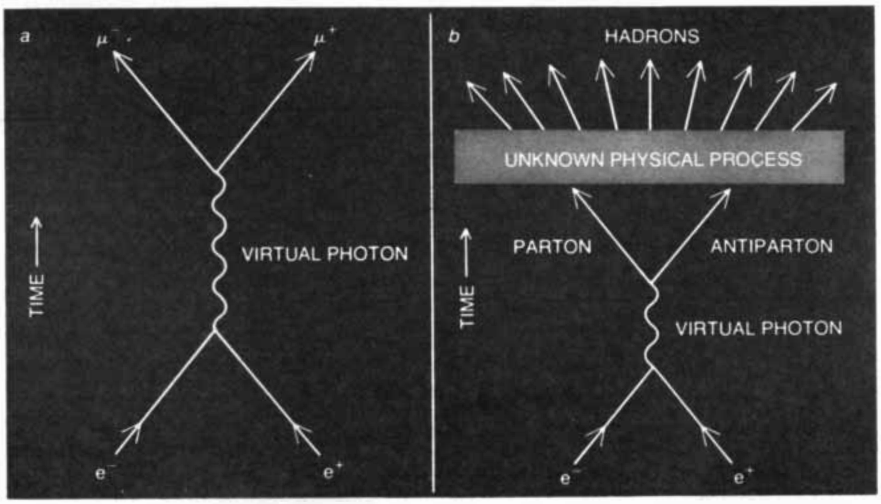
The beams collide at two regions on the perimeter of the ring, where detectors have been installed to record and analyze the products of interactions. The probability of even a single e^-e^+ annihilation in any one pass-through of the bunches is quite low, but because the particles' velocity is nearly that of light they pass through one another several hundred thousand times per second, so that an acceptable reaction rate is achieved. The beams of particles can be stored in the ring for minutes at a time.

Virtual Particles

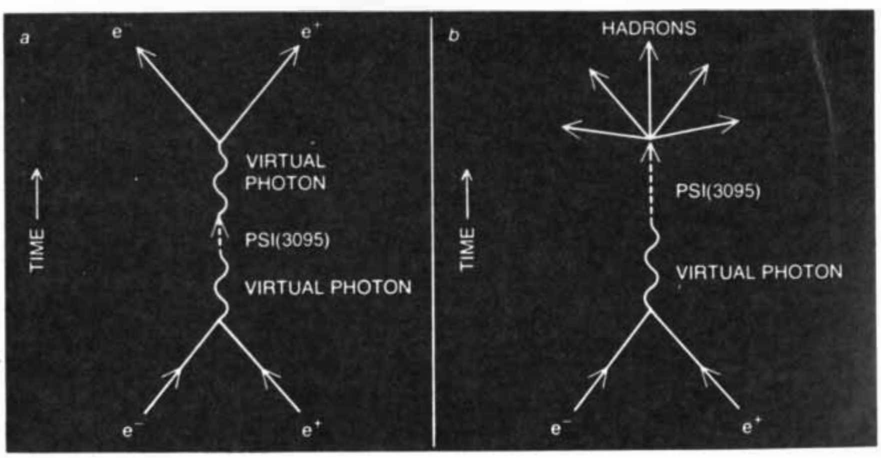
What happens when an electron and a positron meet and annihilate each other with a combined energy of a few billion electron volts? Because the particles



TWO INTERPRETATIONS are possible when the product of an encounter between an electron and a positron is another electron-positron pair. The initial pair may have annihilated (*a*), producing a virtual photon that materialized into an electron and a positron. Or the particles may have passed nearby and been scattered by the exchange of a photon (*b*).



MUONS AND HADRONS can be produced in electron-positron encounters only by the annihilation of the particles, not by scattering. The creation of muons (*a*) requires that the quantum number mu-ness be changed for each of the particles, although the total mu-ness of the pair remains zero. Hadrons are thought to be formed (*b*) through the creation of a parton and an antiparton (or a quark and an antiquark). The partons interact to form hadrons.



NEW PARTICLES, psi(3095) and psi(3684), materialize from the virtual photon when it has exactly the right energy (3.095 GeV and 3.684 GeV respectively). They can decay through a virtual photon into leptons (*a*), such as an electron and a positron, or into hadrons (*b*). The hadrons most commonly produced by psi particles are pions, the least massive of the hadrons.

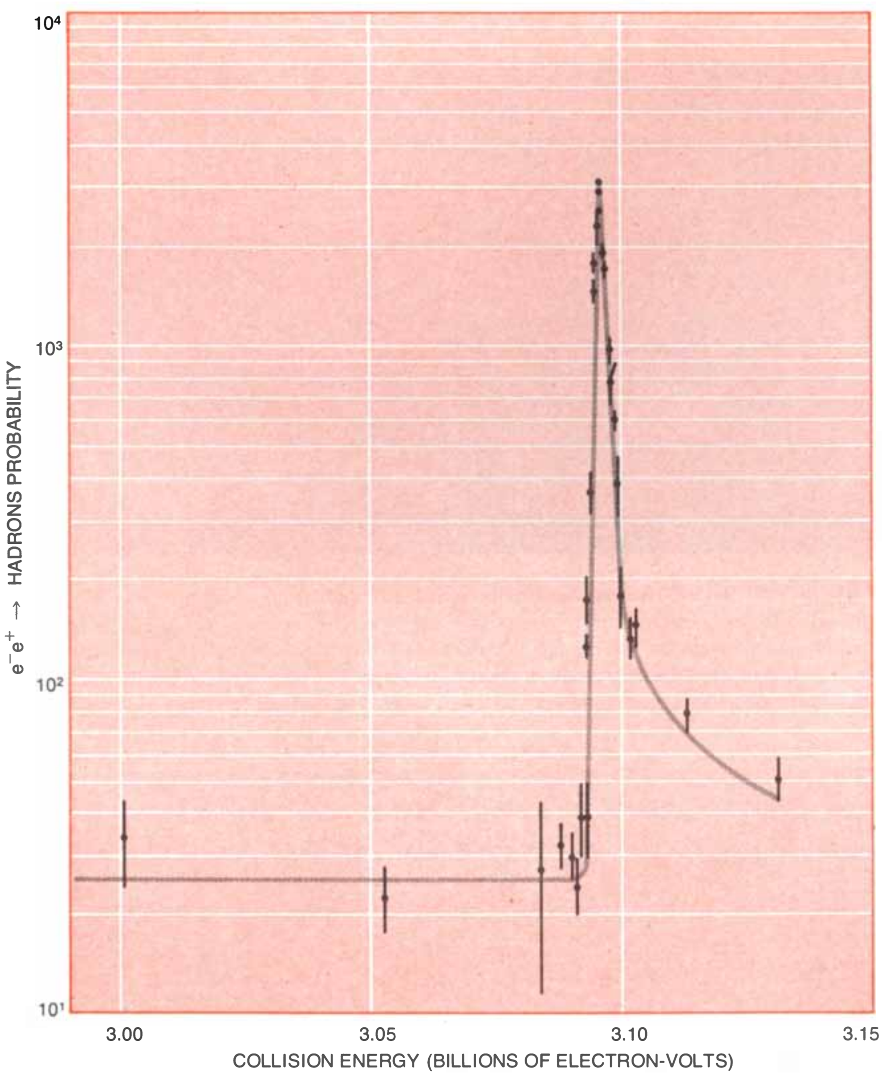
are leptons they do not feel the strong force, and at the energies studied so far the weak interactions are feeble enough to be neglected. The particles are electrically charged, however, so that they do feel the electromagnetic force, and the energy produced by their annihilation is (to a very good approximation) entirely electromagnetic. In other words, the electron and the positron annihilate each other, canceling their electric charges and lepton numbers, to produce a very energetic photon (a gamma ray).

The photon emitted is not, however, a "real" photon such as those that are observed in nature as the quanta of electromagnetic energy. It cannot be real because it has the wrong proportions of energy and momentum, quantities that must be conserved in all interactions. For the photon, which has no mass and which travels at the speed of light, the relation of momentum to energy is constant: the momentum is a fixed fraction of the energy, equal to the energy divided by c. This energy-momentum relation cannot be reconciled with the energy and momentum of the colliding particles. In the storage ring the electron and positron move with equal energy but opposite momentum, and the state formed by their annihilation must therefore have large energy but zero momentum. A photon cannot have that combination of properties.

One possible resolution of this dilemma is for the annihilation to produce two photons that have equal but opposite momenta, thereby satisfying the condition that the sum of the momenta of the products be zero. This reaction does in fact take place, and measurement of it is of major interest. Generally, however, the annihilation process generates as few photons as it possibly can. The probability that an electron or a positron will interact with or produce a single photon is measured by a dimensionless number called the fine-structure constant, equal to about 1/137. For each additional photon the probability is reduced by a higher power of the same factor.

The most likely outcome of the annihilation is therefore the creation of a single photon. As we have seen, however, it cannot be a real particle; it is called a virtual photon, and its most important characteristic is that it can never be observed; it can never emerge from the reaction as radiation. The virtual photon serves merely as a coupling between the initial electron-positron pair having zero total momentum and some subsequent ensemble of particles that must also have zero total momentum.

The virtual photon cannot be observed because it decays before it can be detected. According to the uncertainty principle formulated by Werner Heisenberg, the lifetime of a virtual particle is necessarily too brief for the particle to be observed. In the case of the recent electron-positron annihilation experiments the virtual photon materializes in less than 10^{-25} second into particles with the correct combination of energy and momentum.



RESONANCE detected in electron-positron annihilations at SPEAR signifies the presence of a new particle, psi(3095). It was discovered last November by a group of physicists under the direction of Burton Richter, William Chinowsky, Gerson Goldhaber, Martin L. Perl and George H. Trilling. The resonance represents a greatly increased probability of interaction between the colliding particles at the resonance energy. In this case the electrons and positrons are about 150 times as likely to annihilate each other and yield hadrons when their combined energy is 3.095 GeV as they are at adjacent, "background" energies. The psi(3095) resonance is exceptionally narrow, which indicates that the lifetime of the particle is long.

Particle Production

When the virtual photon decays, several kinds of particles can be created. At the energies investigated so far pairs of electrons and positrons, pairs of muons and antimuons, and hadrons have all been observed.

If the collision yields an electron-positron pair, the annihilation and rebirth of such a pair is phenomenologically indistinguishable from the mere elastic scattering of the incident electron and positron. In the first case the pair disappears in forming a virtual photon and then reappears; in the second the two particles are said to exchange a photon and are deflected, so that their direction is changed but the magnitudes of their energy and momentum are not [*see top illustration on preceding page*].

The creation of a muon-antimuon pair is not complicated by this ambiguity.

Because mu-ness must be conserved the pair can be created only through the annihilation of the electron and the positron. It is also necessary, of course, that the $e^- e^+$ pair have enough energy to account for the mass of the muons. These reactions are of great interest because they provide a method of testing the validity of present theories of electromagnetism at high energy. In addition there is enormous interest in studying the reactions that lead to the production of hadrons.

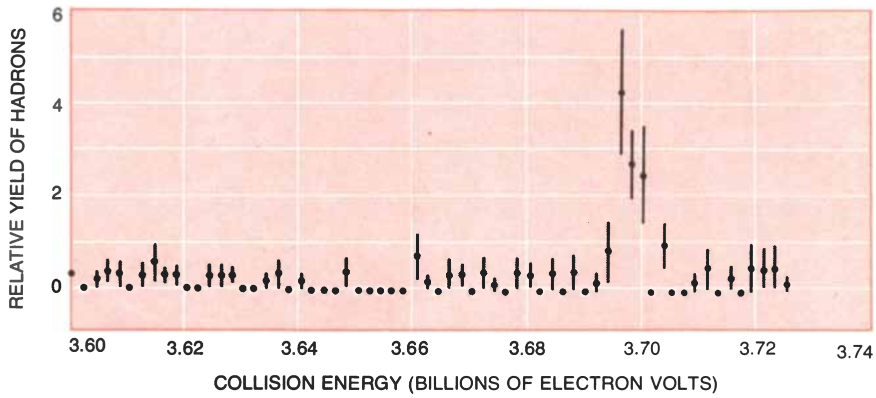
Processes that involve only leptons and photons can be described entirely in terms of the electromagnetic interaction. To predict hadron production, however, one must have a theory of the structure of hadrons; as yet there is no completely satisfactory theory. Most ideas of hadron structure that are under consideration today rely on the concept of partons or quarks as constituents of the hadron. They predict that at high energy a virtual photon can decay into a parton and an antiparton (or a quark and an antiquark). The parton-antiparton pair is then transformed by the strong interaction into hadrons that emerge in the aftermath of the collision.

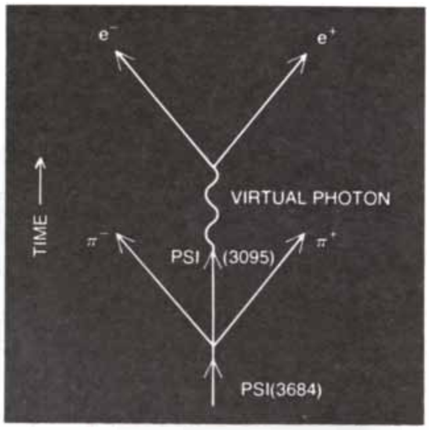
Even though partons and quarks have not been observed in isolation, many of their properties are described in detail by theory; they are *required* to have certain properties in order to explain the properties of the families of hadrons that have been observed. In addition the interpretation of the scattering experiments performed at SLAC (which originally led to the parton hypothesis) indicates that at high energy quarks or partons behave as independent, pointlike entities, just as the leptons do. For this reason it was expected that once the total energy of the colliding particles exceeded a threshold value, suggested by the scattering experiments to be about 2 GeV, the production of leptons and partons would obey similar rules. In particular it was predicted that above 2 GeV the probability of producing hadrons would vary with the collision energy in the same way that the probability of producing a pair of muons varies. This relation was expressed mathematically by stating that the ratio of hadrons to muon-antimuon pairs would be constant and independent of the collision energy. Before the experiments were performed there was argument over what the numercial value of this ratio would be, but there was little doubt that the ratio would in fact be constant at all energies within reach of experiment.

The disagreement arose because several versions of the quark theory predict different values for the hadron/muon ratio. In each case the value is calculated simply by adding the squares of the charges of all the quarks postulated by the model [*see lower illustration on page 62*]. The original formulation of the quark hypothesis, for example, predicts a value of 2/3. Three of the more prominent variations on the theory give values of 2, 10/3 and 4.

Initial measurements of the ratio were made in 1973 at Frascati in Italy and at the Cambridge Electron Accelerator in Cambridge, Mass.; they were soon followed by the first round of measurements with the SPEAR storage ring. To the surprise of most particle physicists, none of the predictions was confirmed; the experiments indicated that the hadron/muon ratio is not constant at all. At 2 GeV the value appeared to be about 2, and it increased gradually to about 5 at 5 GeV. In other words, collisions at 2 GeV were twice as likely to produce hadrons as to produce muon pairs, and at 5 GeV hadrons were about five times as likely. This disturbing development had not yet been accommodated by theory (and it still has not been) when the unexpected massive particles were encountered last November.

At the SPEAR storage ring the first of the new particles was discovered during a second round of measurements of the hadron/muon ratio made at many (and more closely spaced) values of the collision energy. The mass of the particle has now been determined accurately as being 3.095 GeV, and the particle has been named by the SPEAR group psi(3095). (The particle was given another name, *J*, at Brookhaven, but here I am adopting the SPEAR nomenclature.) A second, heavier particle was subsequently found at SPEAR, and it was designated psi(3684), to denote its mass of 3.684 GeV. The heavier particle can decay to form the lighter one, along with two pions. The Brookhaven workers have searched for the 3.684-GeV resonance, but they have found that it cannot be detected by their methods, a fact that may provide a clue to the properties of the particle.

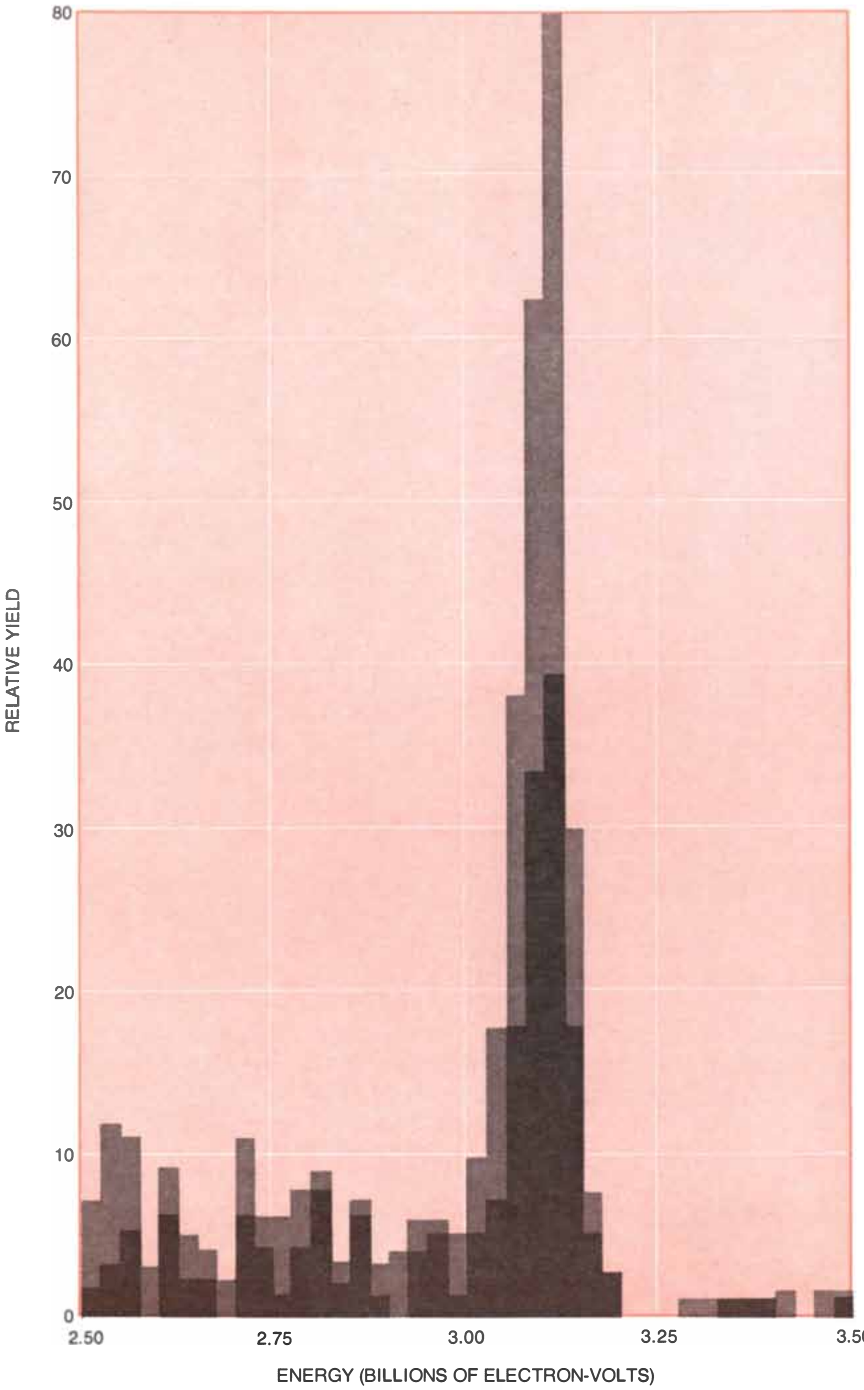




PSI(3684) RESONANCE is neither as tall nor as narrow as that of the psi(3095) particle. The graph (*above*) is at lower resolution than the one on the opposite page, that is, the yield of the electron-positron collisions was measured at fewer energies. The yield is expressed in relative numbers of hadrons produced at each energy. Psi(3684) decays spontaneously into psi(3095) with the emission of a pair of charged pions (*left*). Psi(3095) can then decay in turn, producing other particles, such as an electron-positron pair or hadrons. Because psi(3684) can produce psi(3095) by decay, it has been suggested that the more massive particle is an excited state of the less massive one.

The New Particles

The new particles are unstable, and like all other very massive and short-lived particles they are detected as "resonances," or enhancements of the probability of an interaction [see "Resonance Particles," by R. D. Hill; Scientific American, January, 1963]. As the energy of the colliding beams is increased in small steps a sudden peak in the production of particles is observed at the resonance energy; when the beam energy is increased further, the production drops off again. Such a pattern indicates that a particle exists with a mass equal to the combined masses of the colliding particles; when the colliding particles have exactly the required energy, they are more likely to interact than they are at other energies. In the case of psi(3095) the probability that the electron and the positron will interact was observed to increase by a factor of about 150 at the resonance energy compared with its value at adjacent energies.



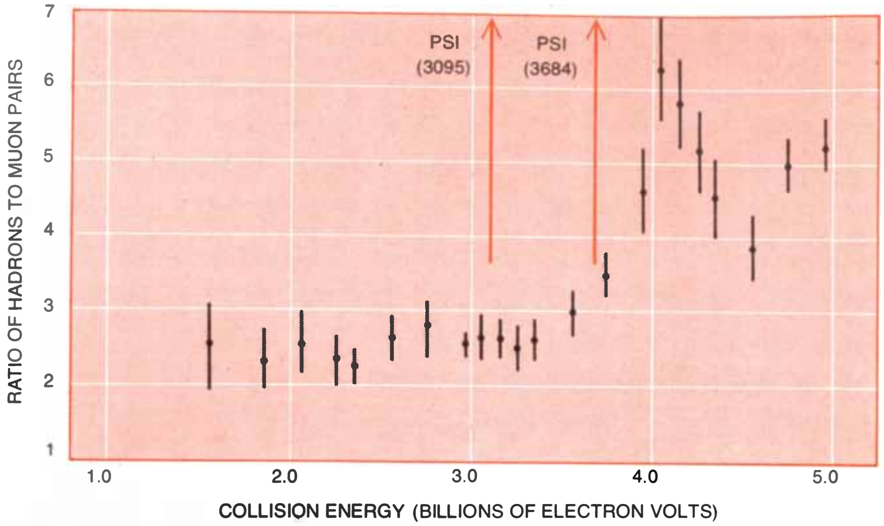
HADRON COLLISIONS also result in the production of one of the new particles. The resonance is detected as an increased probability of the production of electron-positron pairs in the collision of protons with beryllium nuclei; this process is the inverse of an electron-positron annihilation. The production of the new particle in hadron collisions was observed by a group of physicists, directed by Samuel C. C. Ting, working at the Brookhaven National Laboratory. They discovered it simultaneously with the SPEAR group and named it *J*. The psi(3684) particle has proved to be undetectable in the reaction studied at Brookhaven. The 3.095-GeV resonance is measured by the number of electron-positron pairs detected. Measurements at two calibrations of the detector are superposed (*gray and black*).

The height of the resonance peaks is related to their most remarkable feature—their narrowness. The width of a resonance represents the uncertainty in the determination of the energy at which the resonance occurs. This uncertainty in the energy is in turn related to the lifetime of the particle by Heisenberg's uncertainty principle. The equation that gives the lifetime is $\Delta E \Delta t = h/2\pi$, where ΔE is a measure of the uncertainty in the determination of the energy, Δt is a measure of the lifetime and h is Planck's constant (approximately 6.6×10^{-27}). Only for a stable particle—one that has an indefinitely long lifetime, as the proton apparently does—can the energy or mass be precisely defined. In that case Δt is large and ΔE is correspondingly small; the lifetime is unlimited and the resonance width is vanishingly small, so that the energy of the particle can be known with arbitrarily great precision.

The width of the psi(3095) resonance has been determined to be about 77 KeV (thousand electron volts), which represents a very sharp peak. Substituting this value in the equation above yields a lifetime of about 10^{-20} second. The psi(3684) resonance is somewhat broader, and that particle therefore decays faster, with a lifetime of about 10^{-21} second.

Although 10^{-20} second is an almost unimaginably brief interval, far too brief to be measured directly, it is 1,000 times longer than the expected lifetime of such a particle. All other heavy resonances are far broader (their width is typically measured in millions of electron volts rather than thousands), and the particles they represent decay in the characteristic time of the strong interaction: about 10^{-23} second. We are immediately compelled to ask what properties of psi(3095) hold it together so long. Does it exhibit a new kind of structure? Is there a previously unknown quantum number that nature wants to conserve but that must be changed when psi(3095) decays? Questions of this kind are among the most fundamental that can be asked in the physics of elementary particles.

Although the psi particles are themselves mystifying, their discovery has to some extent simplified, if not clarified, the earlier measurements of the hadron/muon ratio. It now appears that for energies between 2 GeV and 3.8 GeV the ratio is roughly constant at about 2.5; the only significant deviations are those associated with the two psi resonances. At higher energy the ratio increases, reaches a broad peak at about 4.1 GeV, then stabilizes at a value of about 5 [*see illustration at right*]. Recent experiments at SPEAR have confirmed that the hadron/muon ratio remains approximately constant up to 6 GeV. The causes of the observed peculiarities near 4.1 GeV are not yet understood. The broad peak may or may not represent an additional resonance or perhaps several resonances.



RATIO of hadrons to muon-antimuon pairs produced in electron-positron annihilations was predicted to be constant at collision energies above 2 GeV. The SPEAR experiments demonstrated that the ratio is roughly constant up to about 3.8 GeV if the effects introduced by the two psi particles are ignored. It rises to a peak at about 4.1 GeV, then declines again to a nearly constant value of approximately 5. Various versions of the quark model predict different constant values for the ratio, but none of them explains the observed fluctuations.

Quantum Electrodynamics

The interpretation of the puzzling events revealed by electron-positron collisions is made substantially easier by the fact that the initial state emerging from the annihilation—the virtual photon—is simple and well understood. Because it is formed in a purely electromagnetic process it is described by the theory of quantum electrodynamics, one of the most successful theories in physics.

Quantum electrodynamics is the theory constructed by imposing the laws of quantum mechanics on the classical theory of electromagnetism. It is a basic tenet of quantum electrodynamics that electrons, when they interact with electromagnetic radiation, act as point charges. This assumption, however, may be only an idealization that, although useful at large distances, fails to describe events at subnuclear dimensions.

Physicists are eager to test the theory of quantum electrodynamics at ever higher energy and with greater precision in order to probe it in finer detail. It is particularly important to know whether or not quantum electrodynamics is valid for the very-high-energy collisions studied at SPEAR, since the theory is central to the interpretation of hadron structure.

Quantum electrodynamics has been tested at distances ranging from more than 250,000 miles (in measurements of the earth's magnetic field) to subatomic dimensions much smaller than 10^{-8} centimeter (in describing the detailed spectrum of the hydrogen atom). The atomic measurements have been made so precisely that theory and experiment agree to roughly four parts in 10^9.

In interactions of electrons and positrons whose final products are either electron-positron pairs, muon-antimuon pairs or two or more real photons only the electromagnetic interaction makes a significant contribution. The entire process, therefore, should be comprehended by quantum electrodynamics and can be employed to test the theory. Another application of the uncertainty principle demonstrates that at the energies available with existing storage rings such events probe distances as small as 10^{-15} centimeter, or roughly 1 percent of the diameter of a hadron. Even at that small scale experiment has so far revealed no flaws in the theory.

The quantitative verification of quantum electrodynamics is the 20th-century parallel to the exciting voyage of Isaac Newton 300 years ago. The elegant simplicity of nature was first comprehended when Newton realized that the same law of gravity that governs the apple's fall to the ground also describes the motion of the planets, hundreds of millions of miles apart. We have now also learned that nature applies the same laws of electrodynamics both on a large scale, extending to about 80 earth radii, and in a realm less than a millionth the size of an atom. Quantum electrodynamics has been verified for distances encompassing more than 25 orders of magnitude.

If a colliding electron-positron pair can create a psi(3095) resonance, then of necessity the resonance particle, by the reverse process, must be able to decay into an electron-positron pair. (Indeed, that is how the particle was discovered by Ting and his colleagues at Brookhaven.) Psi(3095) decays in this way about 7 percent of the time, and muon pairs are produced in another 7 percent of the decay events. The remaining 86 percent of the time the resonance decays to yield hadrons. Since both muon pairs and electron pairs are created through purely electromagnetic forces, quantum electrodynamics should describe these processes, and it is possible to determine to what extent the two mechanisms interfere with each other.

In practice the interference is detected by measuring the ratio of muon pairs to electron pairs produced at energies near the psi(3095) resonance. Of particular importance is the observed decline in the ratio just below the peak energy of the psi(3095) resonance. On fundamental theoretical grounds such an "interference dip" is expected only when the electron-positron annihilation has two possible final states and the two states are, with regard to the electromagnetic interactions, interchangeable. In this instance the two final states can only be the photon and psi(3095), and the experiment has the important implication that psi(3095) has the same quantum numbers as the photon.

This discovery enables us to specify some of the characteristics of the new particles. For example, psi(3095), like the photon, must have a spin angular momentum of 1. Thus our detailed understanding of quantum electrodynamics provides us with information on the new particles.

Color and Charm

There is more to the psi particles, however, than their quantum numbers alone. In particular we want to know what kind of quarks they are made up of. For now all attempts to deduce their quark constitution must be considered speculations only, because experiment has not provided the evidence to discriminate conclusively among a welter of competing theories; nevertheless, two proposals that have been given considerable attention deserve discussion. Both were proposed to explain unrelated phenomena long before the psi particles were discovered.

Quarks, like the particles they compose, are assigned quantum numbers. All of them have spin angular momentum of 1/2, for example, and baryon number of 1/3. Of the original triplet of quarks proposed by Gell-Mann and Zweig, the *u* quark has a charge of 2/3 and the *d* and *s* quarks have a charge of −1/3. Antiquarks, of course, have the opposite quantum numbers. According to these assignments, the baryons, being made up of three quarks, must have a half-integral spin, a baryon number of +1 and a charge of +2, +1, 0 or −1. The mesons, as aggregates of a quark and an antiquark, must have an integral spin, a baryon number of 0 and a charge of +1, 0 or −1.

This ingenious scheme neatly accounted for all the particles and resonances that had been observed when it was proposed, and it soon proved its predictive power by postulating an unknown resonance that was promptly discovered. It contains a deeply disturbing peculiarity, however: the quarks are required to be particles with a half-integral spin but they do not behave as such particles are expected to.

All observed particles with a spin of 1/2 obey Wolfgang Pauli's exclusion principle, which demands that no two be in an identical state. The electrons of an atom, for example, always differ in at least one quantum number; if they have the same orbital configuration, they have opposite spin. Our understanding of atomic structure and of the periodic table of the elements is based on this concept. Particles with integral spin (such as the mesons and the photon) are not affected by the exclusion principle; arbitrarily large numbers of them can be assembled in the same state. (The half-integral-spin particles are said to obey Fermi-Dirac statistics, and they are called fermions; the integral-spin particles obey Bose-Einstein statistics and are called bosons.)

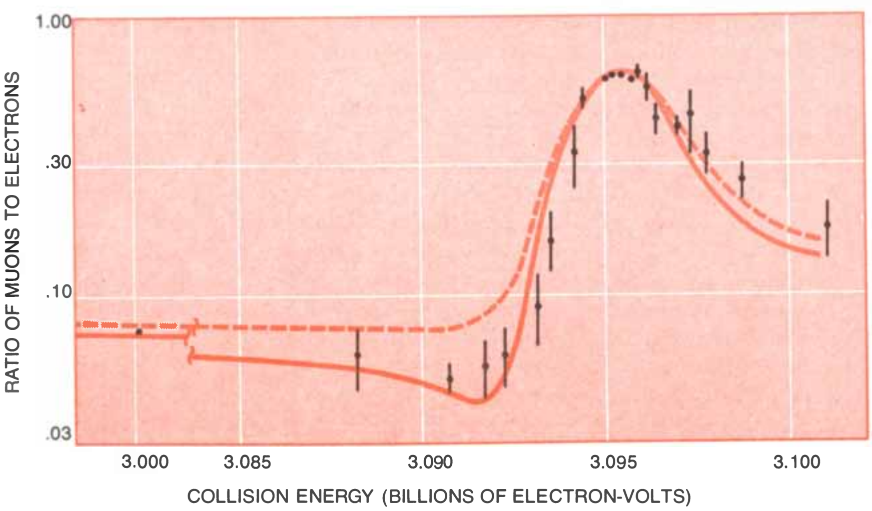
Quarks seem to violate these rules. They must have spin of 1/2, but on the other hand in constructing the baryons it is necessary to assume that two or more are bound together in the same state. The paradox threatens to do violence to some basic and cherished theoretical principles. It can be resolved, however, simply by insisting that quarks do obey the exclusion principle. All that is necessary to make them conform to the rule is to endow them with a new quantum number having three possible values, so that the quarks bound together in a baryon, although identical in all other properties, can differ in this new one. The new property, first suggested in 1964 by Oscar W. Greenberg of the University of Maryland, is called color, although it has nothing to do with vision or the color of objects in the macroscopic world; in this context color is merely a label for a property that expands the original ensemble of three quarks to nine. Each quark of the original triplet can appear in any of three colors—say red, yellow or blue.

An incidental feature of the color hypothesis is that the addition of six more quarks to the original three enables us to reformulate the quark theory with integral charges rather than fractional ones. A model of this kind was constructed by Moo-Young Han of Duke University and Yoichiro Nambu of the University of Chicago in 1965.

All versions of the color theory assume that in the known baryons the three colors of quarks are equally represented; as a result the particle exhibits no net color. Similarly, the mesons are made up of equal proportions of red, yellow and blue quark-antiquark pairs and are also colorless. Indeed, some physicists have speculated that in nature all particles may be colorless. One of a class of proposed interpretations of the psi particles suggests that they may be the first observed states of colored matter.

The second important theory that has been invoked to explain the psi particles is called the charm hypothesis. It was proposed by a number of theorists in 1964 simply for the sake of symmetry. (In particle physics that is not a trivial motive, since every symmetry has an associated conservation law.) In order to construct a parallel to the four known leptons, the theory adds to the three original quarks a fourth, designated *c* for charm. The *c* quark has a charge of 2/3, and it has +1 unit of the new quantum number charm; all the other quarks have zero charm.

In 1970 the new quark and its quantum number were given an important role in physics through the work of Sheldon L. Glashow, John Iliopoulos and Luciano Maiani of Harvard University. They invoked the charm quark in order to explain the suppression of certain particle decays that in the three-quark model should have proceeded normally.



INTERFERENCE between entirely electromagnetic events and events that involve the strong interaction is detected by measuring the ratio of muon pairs to electron pairs produced in annihilation experiments. The presence of interference is an indication that psi-(3095) has the same quantum numbers as the photon. If the quantum numbers are the same, theory predicts that the muon/electron ratio will be depressed at energies just below the resonance energy (*solid colored line*); if the numbers are different, there will be no interference (*broken colored line*). Experimental results (*black dots and bars*) show interference.

Significance of the Particles

If the *c* quark exists, of course, one would expect to observe charmed states of matter, such as a baryon made up of a *c* quark and two other quarks. No charmed particles have been observed, and hence it is believed that charmed quarks are much more massive than uncharmed ones. The psi particles are not believed to be charmed either, since they are produced electromagnetically and the electromagnetic interactions conserve charm, but it has been suggested that they could be mesons consisting of a *c* quark and an anti-*c*-quark. That combination would not exhibit charm, because the charm quantum numbers of the two quarks would cancel.

The first requirement of any theory of the new particles is that it explain their anomalously long lifetime. At the moment neither color nor charm seems to offer an entirely satisfactory explanation.

If psi(3095) and psi(3684) are considered to be colored particles, then it is assumed that they live long because there are no other particles of lower mass to which they can transfer their color when they decay. In the strong interactions color is conserved, so that a colored particle could not decay by the strong force. A particle might well be able to change color, however, in the electromagnetic interactions; the unit of color would be carried off by a photon, leaving decay products of colorless hadrons. As we have seen, the emission of a photon is suppressed by a factor of 1/137, and so the requirement that psi(3095) decay electromagnetically offers a partial explanation of its stability. The actual suppression of the decay involves a factor of about 1,000 instead of 137, but the discrepancy might be accounted for.

The hypothesis that the psi particles are colored is subject to experimental test by at least two methods. First, the gamma rays needed to carry off color in the electromagnetic decay should be detected among the decay products of the particles. Second, if the psi particles do represent colored particles, they cannot be the only ones; all versions of the color theory predict families of associated colored mesons, some of which would be electrically charged.

The charmed-quark theory is simpler and therefore more appealing, but it offers no compelling explanation for the psi lifetime. There is no known mechanism that would prevent a meson made of a charmed quark and antiquark from simply transforming itself into an ordinary quark and antiquark. In this process there are no quantum numbers to be conserved, since in the particle-antiparticle pair charm and all other properties cancel. The only apparent solution is to postulate a new law of nature, to declare arbitrarily that the decay of mesons made of a charmed quark and antiquark is inhibited by a factor of 1,000. There is precedent for such a rule in the theoretical treatment of other meson decays, but psi(3095) involves a larger inhibition factor. Moreover, it is a mystery why such rules should be required at all.

In spite of this weakness the charm hypothesis has attractive elements. The existence of resonances corresponding to a charmed quark and antiquark meson was predicted before the psi particles were discovered. The existence of the particle was discussed by Thomas W. Appelquist and H. David Politzer of Harvard, who named the hypothetical entity charmonium. They also suggested that it could be formed in electron-positron annihilations.

The charm hypothesis is also susceptible to experimental test. Theorists have already suggested mass regions where it might be profitable to look for excited states of charmonium. Psi(3684) could be one of those states, but there would have to be more. The transition between psi(3684) and psi(3095) with the emission of two pions conforms to theoretical expectations. The energy transitions between other states have also been calculated, and if they exist, it should be possible to detect them. Finally, particles incorporating just one charmed quark and therefore exhibiting charm should exist; the discovery of a charmed meson would provide strong evidence for the theory.

In experiments completed so far none of the phenomena that would confirm the color or charm hypotheses have been detected. Indeed, the failure to find this supporting evidence has become an embarrassment to both theories.

The new particles are not the only challenges to theory issuing from the recent discoveries. Some explanation is also required for the broad enhancement of hadron production in the vicinity of 4.1 GeV, and the questions raised by the odd behavior of the hadron/muon ratio throughout the energy range have not been resolved. The search for bumps and lines in the mass spectrum goes on not only at SPEAR but also at the DORIS storage ring at the German Electron Synchrotron in Hamburg and the ADONE storage ring at Frascati. Other investigators are studying the production of the psi particles through the interaction of photons and hadrons and through hadron-hadron collisions. (The last is the method employed by Ting and his colleagues at Brookhaven.)

In the experimental and theoretical investigations now under way many current concepts are being challenged; one, however, is not in question: that of quarks themselves. The discovery of the psi particles has confirmed again the central importance of quarks as the constituent particles of hadrons. Whether or not we shall ever see free quarks in the laboratory is another question; it is possible that they will always remain unobserved, exhibiting their physical reality only through their success in explaining the structure of hadrons and the forces that act on them.

Furthermore, we have no assurance that the quarks, whether there are three or nine or 12 or more of them, are the fundamental particles of matter. In the 20th century physics has probed the atom to discover the nucleus within, and has broken up the nucleus into its constituent particles. Those particles are now interpreted as being composites of more basic entities, the quarks. It is not unreasonable to imagine that we shall someday penetrate the quark and find an internal structure there as well. Only the experiments of the future can reveal whether quarks are the indivisible building blocks of all matter, the "atoms" of Democritus, or whether they too have a structure, as part of the endless series of seeds within seeds envisioned by Anaxagoras.

QUARK		CHARGE	BARYON NUMBER	CHARM	COLOR
	u_r				RED
u	u_y	$+2/3$	$+1/3$	0	YELLOW
	u_b				BLUE
	d_r				RED
d	d_y	$-1/3$	$+1/3$	0	YELLOW
	d_b				BLUE
	s_r				RED
s	s_y	$-1/3$	$+1/3$	0	YELLOW
	s_b				BLUE
	c_r				RED
c	c_y	$+2/3$	$+1/3$	$+1$	YELLOW
	c_b				BLUE

QUARK HYPOTHESIS states that hadrons are not elementary particles but composites of more fundamental entities called quarks. The original formulation of the theory, proposed independently by Murray Gell-Mann and George Zweig, postulated three quarks, *u, d* and *s*. Charge and baryon number (and other quantities not shown) are assigned to them according to the principle that baryons are made up of three quarks and mesons of a quark and an antiquark. Modifications of the theory add a fourth quark, *c*, which exhibits a property arbitrarily called charm, and propose that each quark exists in three states, distinguished by another property, called color. Thus there could be three, four, nine, 12 or more quarks.

MODEL OF HADRON STRUCTURE	PREDICTED VALUE OF HADRON-MUON PAIR RATIO
GELL-MANN-ZWEIG MODEL THREE QUARKS FRACTIONAL CHARGE	$(\frac{2}{3})^2 + (-\frac{1}{3})^2 + (-\frac{1}{3})^2 = \frac{2}{3}$
COLOR HYPOTHESIS THREE TRIPLETS OF QUARKS FRACTIONAL CHARGE	$3\left[(\frac{2}{3})^2 + (-\frac{1}{3})^2 + (-\frac{1}{3})^2\right] = 2$
COLOR HYPOTHESIS AND CHARM HYPOTHESIS THREE QUARTETS OF QUARKS FRACTIONAL CHARGE	$3\left[(\frac{2}{3})^2 + (-\frac{1}{3})^2 + (-\frac{1}{3})^2 + (\frac{2}{3})^2\right] = \frac{10}{3}$
HAN-NAMBU MODEL THREE QUARTETS OF QUARKS INTEGRAL CHARGE	$(1)^2 + (1)^2 + (-1)^2 + (-1)^2 = 4$

VARIANTS OF THE QUARK THEORY predict different values for the ratio of hadrons to muon pairs produced in electron-positron annihilations. In each case the predicted value is equal to the sum of the squares of the charges of the quarks included in the theory. The Gell-Mann–Zweig model gives a value of 2/3; the color hypothesis would increase the ratio to 2; assuming that both color and charm exist would yield a value of 10/3. The scheme devised by Moo-Young Han and Yoichiro Nambu eliminates the fractional charges common to other quark theories and predicts a value of 4 for the ratio. So far experimental findings at SPEAR and other laboratories cannot be reconciled with any of the predictions.

People all over the world talk the same language about the DC-10: "I like it."

Chicago...
"I like it."

Rome...
"Veramente fantastico."

Copenhagen...
"Den er skøn."

Frankfurt...
"Ein phantastisches Flugzeug."

Dakar...
"J'adore."

New York...
"I like it."

Tokyo...
好きですね

Paris...
"J'aime beaucoup."

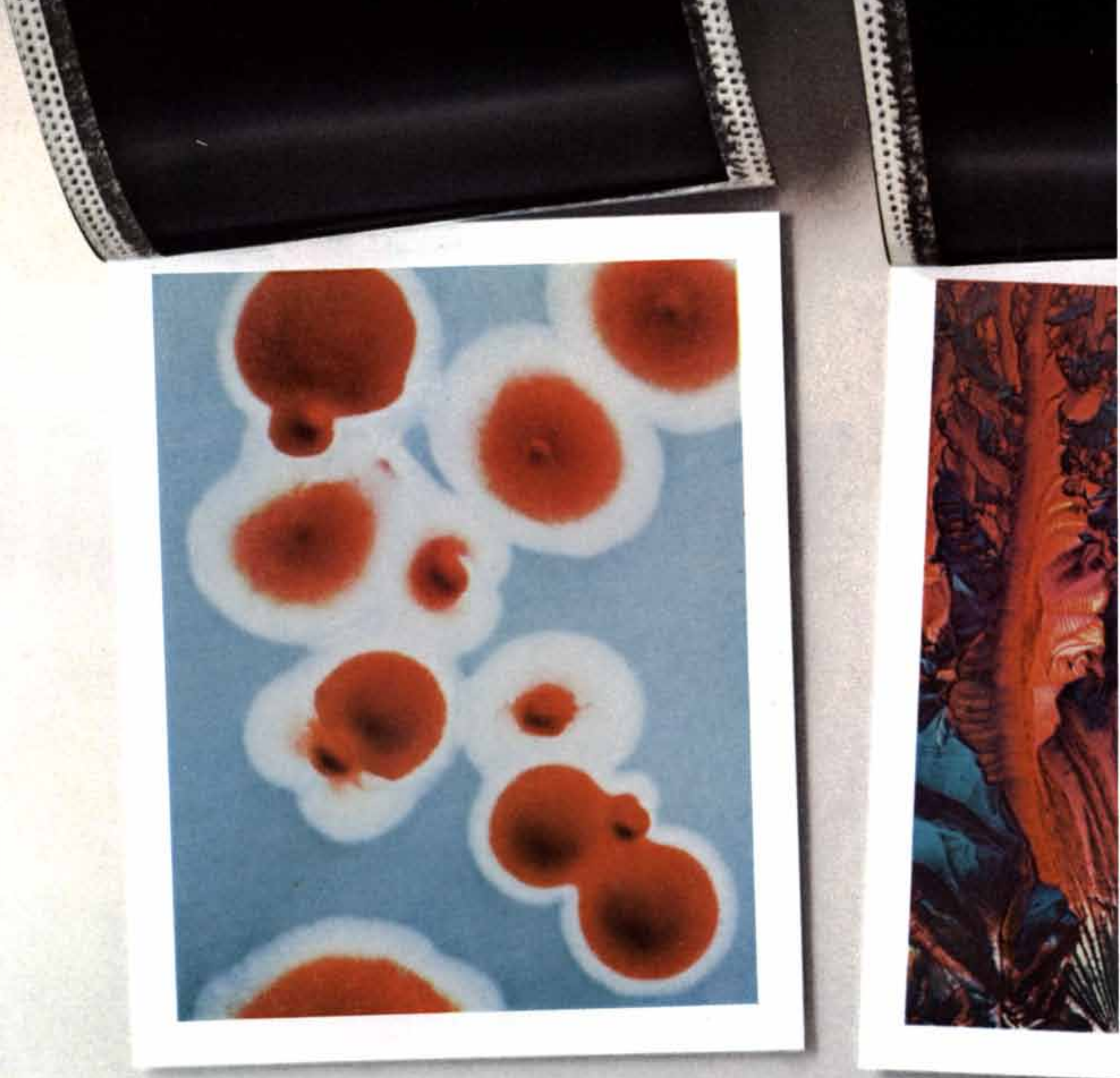
Hong Kong...
我中意

Caracas...
"Me gusta."

If you've already enjoyed flying on a DC-10, you've shared a pleasant experience with millions of other people. Each day, more than 75,000 travelers fly a DC-10 to more than 120 cities in 67 countries. In many languages, they say they like the spacious, quiet comfort of the DC-10. So, ask your favorite airline or travel agent to book you aboard the DC-10.

DC-10 the choice of 34 airlines

MCDONNELL DOUGLAS



Now: Beautiful Bacteria, Stunni

Polaroid presents Polacolor 2. A new instant film that can literally saturate your pictures with color.

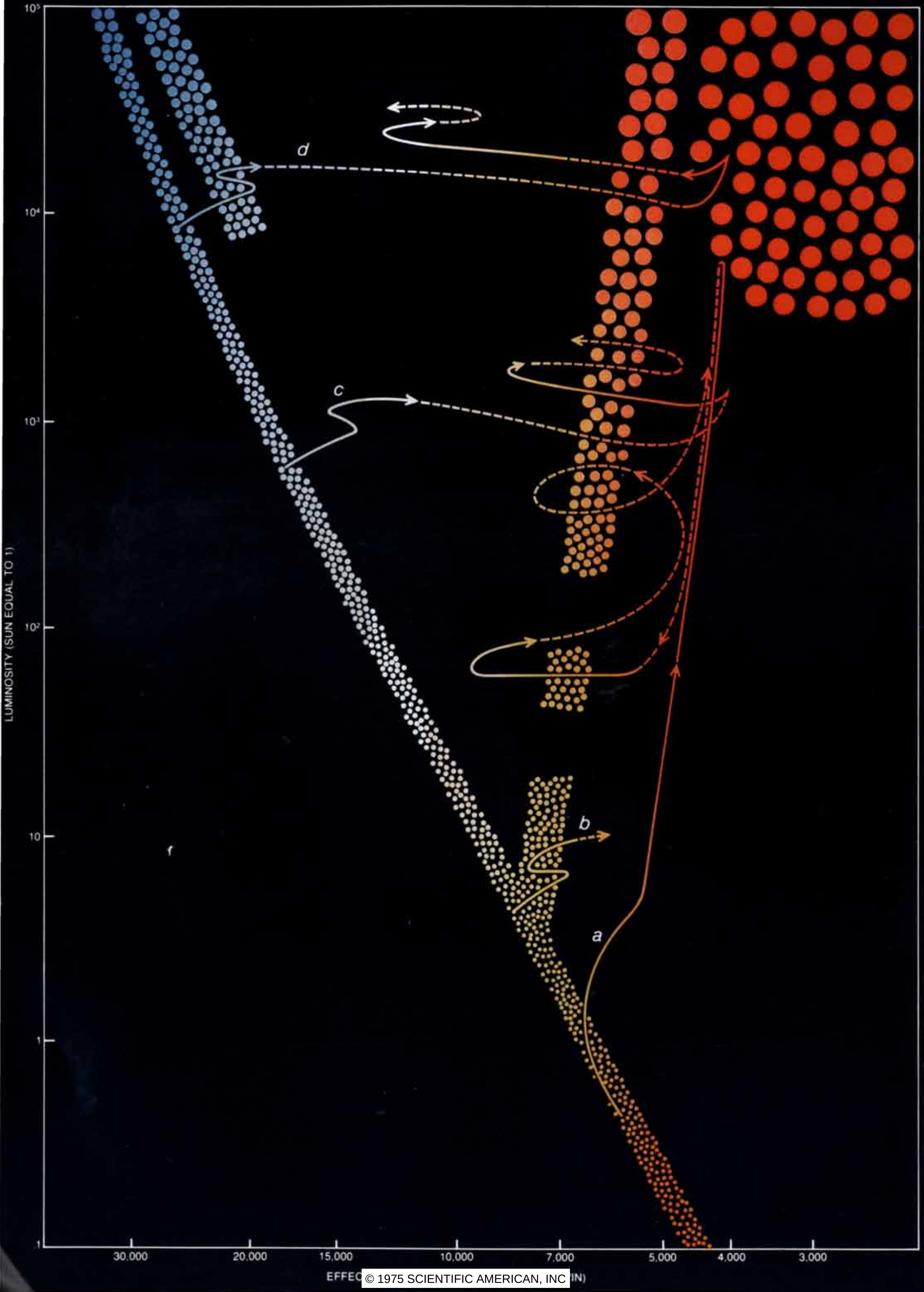
The next time you call a germ a beauty, the world will understand.

Polacolor 2 will be there to back you up.

It's a totally new 60-second color print film with brilliant dyes, high acutance, and precise separation of color.

Polacolor 2 gives you fiery reds, electric blues, and chlorine greens.

Maybe you like beauty, but you still need accuracy.



LUMINOSITY (SUN EQUAL TO 1)
10^5
10^4
10^3
10^2
10
1
.1
30,000
20,000
15,000
10,000
7,000
5,000
4,000
3,000
a
b
c
d

PULSATING STARS

A star that varies regularly in brightness is vibrating like the air in an organ pipe. The fundamental vibration or its harmonics are strong clues to the star's composition and internal architecture

by John R. Percy

To the casual observer the stars in the heavens appear to be changeless. There are, however, dozens of stars that fade and brighten and whose behavior is clearly visible to the unaided eye. In fact, it now appears that most stars—even the sun—vary in brightness to some extent. Each star in its unique way may be rotating on its axis, or revolving around a companion, or pulsating or erupting, often doing all four at the same time and often in a quite complex manner. These stellar goings-on are monitored with great interest by astronomers, who can use them to gain a better understanding of the nature and the evolution of stars.

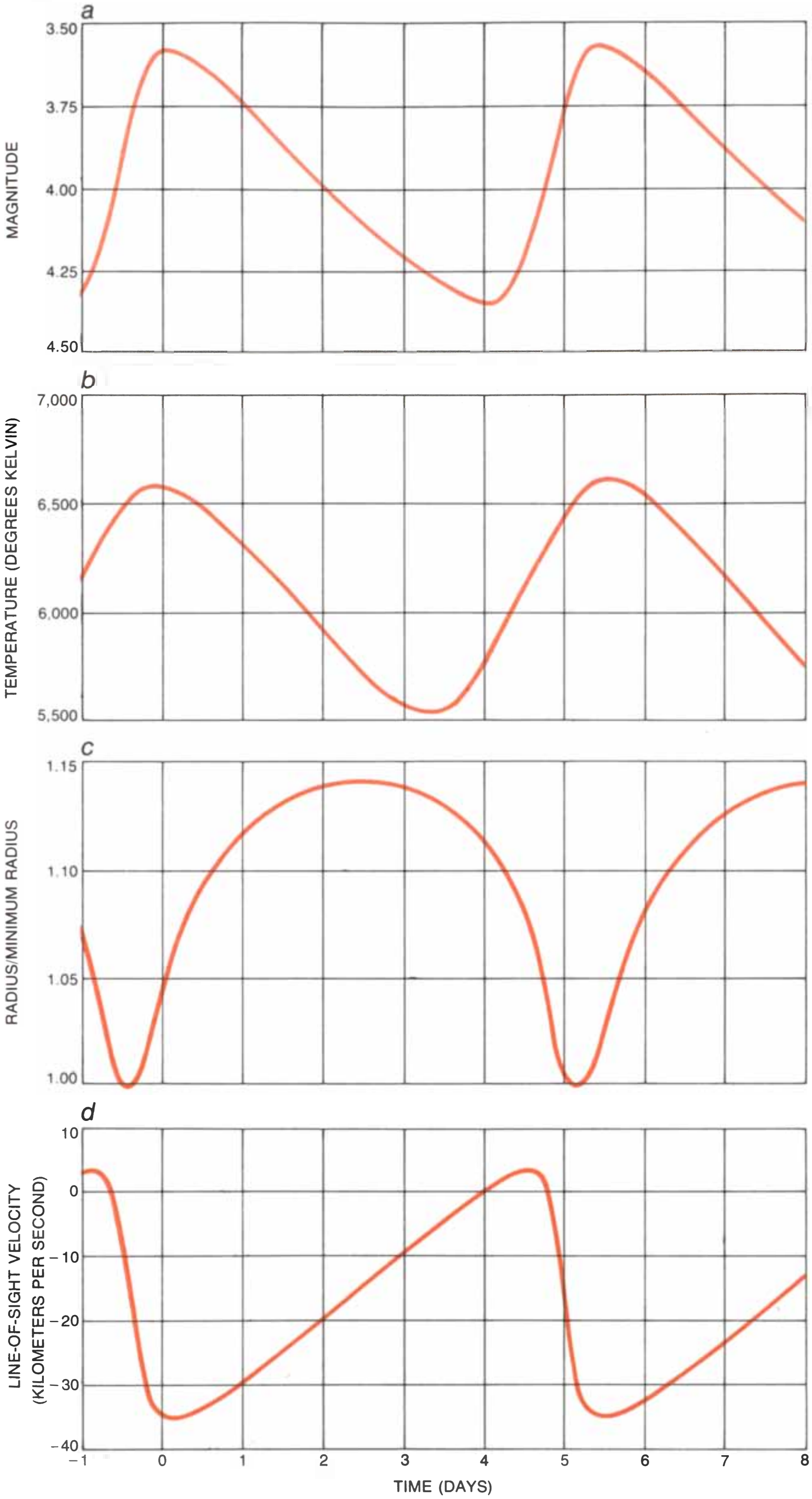
A veritable menagerie of variable stars are known. With the exception of a few particularly baffling objects, the menagerie can be divided into four broad species. The geometric variables are systems of two or more stars whose light varies because one star is periodically eclipsed or tidally distorted by a companion star. The rotating variables are stars that have a hotter or a cooler region on their surface that periodically rotates through our line of sight; the pulsars are related to this species. Eruptive variables are stars that are characterized by sudden and unpredictable outbursts; these performances range in violence from the explosions of supernovas that nearly obliterate themselves through the lesser outbursts of ordinary novas to the flares of more commonplace stars such as the sun. Pulsating variables are stars that periodically brighten and fade because they go through a cycle of expansion and contraction.

Although many observable properties of a variable star may be changing at the same time, its variation in magnitude, or brightness, is usually the most pronounced property and the easiest to measure. The magnitude scale is a logarithmic one on which each magnitude represents a factor of 2.512 times in brightness. The more negative the magnitude, the brighter the object. If the magnitude of a variable is measured repeatedly and plotted with respect to time, the resulting graph is a light curve. Most pulsating stars vary in a regular way, that is, the period required for one complete cycle is approximately constant, the range in magnitude from maximum to minimum brightness remains the same and the shape of the light curve does not change from cycle to cycle.

The constellation Cepheus in the northern sky includes several interesting pulsating variables. The star Delta Cephei is the prototype of a class of pulsating variables known as classical Cepheid variables. It has a period of 5.366341 days and a range in brightness of .9 magnitude. The star T Cephei, another pulsating variable, is dimly visible to the unaided eye and is clearly visible with binoculars. It is distinctly red in color and its variations are much more leisurely and pronounced than those of Delta Cephei. It has a period of more than a year and a range of six magnitudes. Mu Cephei varies irregularly on a time scale of several years. Beta Cephei and Epsilon Cephei are also pulsating variables, with periods of only a few hours. Their ranges in brightness, however, are too small to be detected with the unaided eye. At least three other bright stars in Cepheus are suspected to be variables. Our knowledge of the variability of even the brightest stars is clearly far from complete.

Studying the variability of a pulsating star is just one approach to understanding the object. A second approach is to measure the variable's equilibrium or average properties, that is, the properties it presumably would have if it were

HERTZSPRUNG-RUSSELL DIAGRAM, which plots the luminosity of stars against their temperature (and color), shows that the distribution of pulsating variable stars differs from the distribution of normal stars. The colors and sizes of the dots represent the apparent colors and the approximate relative sizes of the stars. The majority of normal stars lie on or near the main sequence, the swath of stars running diagonally from upper left to lower right. The more massive a normal star is, the higher it lies on the main sequence. Pulsating stars are distributed quite differently. In the region of large red stars at upper right are the irregular variables and the long-period variables: stars with pulsation periods of more than a year and a wide range in brightness from maximum to minimum. The nearly vertical region extending from the top of the diagram to the main sequence (and possibly extending below it) is the Cepheid instability strip. The classical Cepheid variables and the Population II Cepheid variables are the large orange stars in the top portion, the RR Lyrae variables are in the middle region and the Delta Scuti stars and the dwarf Cepheids are in the section of the instability strip that crosses the main sequence. The isolated region of hot blue stars above the main sequence at upper left is the region of the Beta Cephei stars. Stars reach the instability regions by proceeding along evolutionary tracks (*arrows*) in the course of their lifetime. Some phases of their evolution are more rapid than others; the rapid phases are indicated by the broken lines. The color of the track indicates the approximate temperature or color of the star at that point on the track. Sample evolutionary tracks are shown for four stars of different mass: .65 times the mass of the sun (*a*), 1.5 times the mass of the sun (*b*), five times the mass of the sun (*c*) and 10 times the mass of the sun (*d*).



VARIATION IN BRIGHTNESS COINCIDES with other changes in a pulsating star, as is shown here in the case of Delta Cephei. The light curve (*a*) of the star shows that its brightness varies regularly, with a period of 5.366341 days and a range of .9 magnitude. The changes in brightness are due mainly to variations in the star's temperature (*b*) and to a lesser extent to variations in its radius (*c*). As the star expands and contracts, the radiating layers alternately approach the observer and then recede from him. The line-of-sight velocity of the layers can be measured by the Doppler shift of the lines in the star's spectrum (*d*).

not pulsating. The equilibrium properties of particular interest are the mass, the effective surface temperature, the luminosity and the radius. We should ultimately like to know whether or not the nature of the variations is related to the equilibrium properties and, if it is, how pulsating stars then differ from nonpulsating ones.

The mass of a star can be determined directly only if it has a companion star in orbit around it. The effective surface temperature can be determined from the star's observed color or from other characteristics of its spectrum. The luminosity can be obtained by measuring the distance to the star and relating that distance to the star's brightness as it is perceived from the earth. The radius can be calculated from the effective temperature and the luminosity, or from a close analysis of the variations in the star's brightness, in its color and in the velocity of its surface along the line of sight as it pulsates [*see illustration at left*]. Because of the star's cycle of expansion and contraction its surface alternately approaches and recedes from the earth. As a result the lines in its spectrum are Doppler-shifted toward shorter wavelengths as the surface approaches us and toward longer wavelengths as the surface recedes. The variations in the velocity of the surface along the line of sight are determined from these shifts.

When a star pulsates, it expands past its equilibrium size until the expansion is slowed and reversed by gravity. Like a pendulum, the star overshoots its equilibrium size again and continues to contract. As the gas pressure within the star builds up, the contraction is slowed and the surface is pushed outward again. If the pulsation were not somehow driven, dissipative forces such as friction would eventually cause it to stop. The fact that it continues tells us that some process is constantly feeding it mechanical energy. Moreover, the velocity attained by the pulsating gas is quite high, typically 50 kilometers per second. In some stars the motion may be vigorous enough to eject material from the star completely.

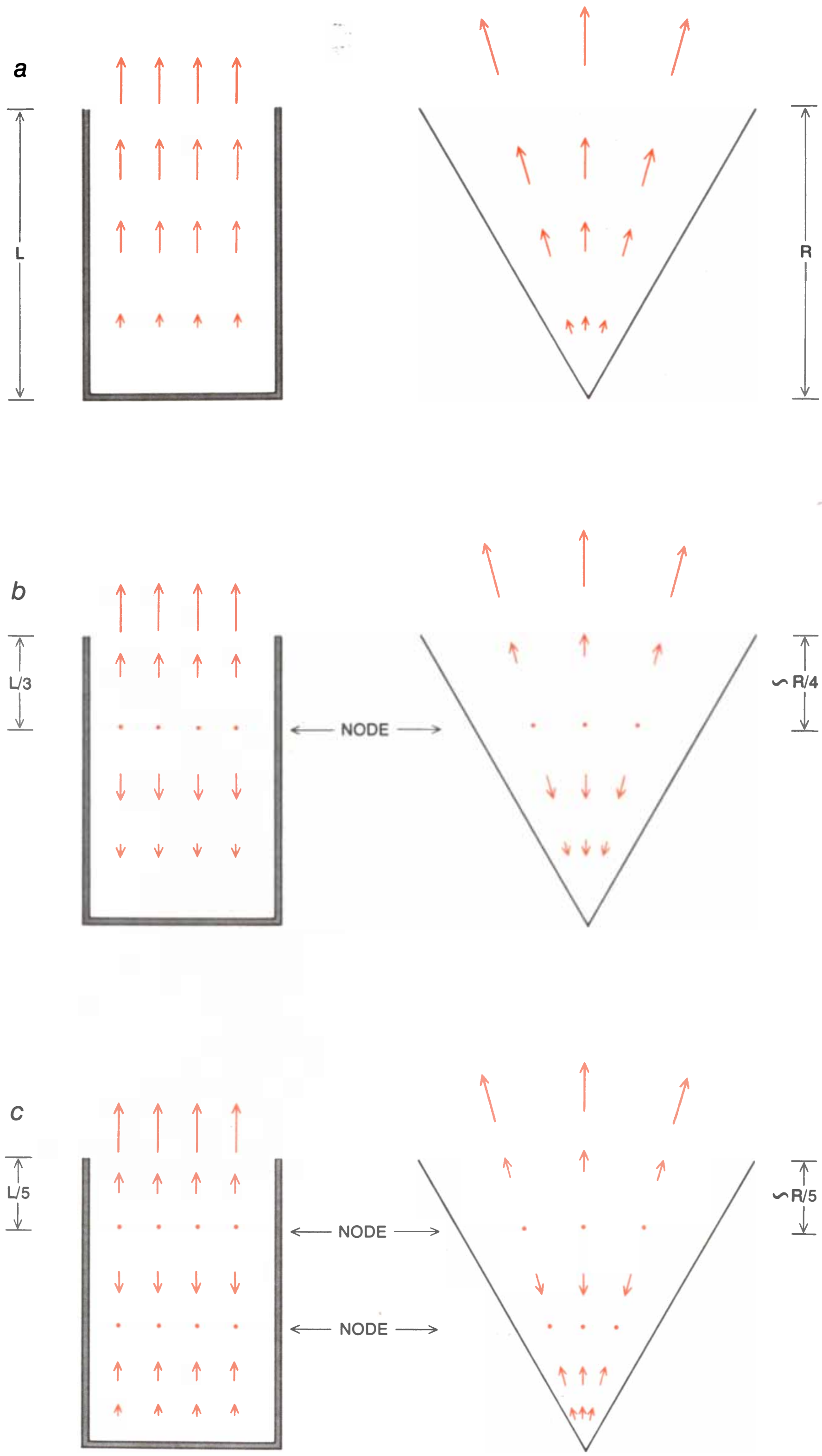
The modes of pulsation of a star are not unlike the modes of vibration of air in an organ pipe. An organ pipe is a column of gas closed at one end and open at the other; a star is a sphere of gas confined at the center and open at the surface. The simplest mode of the vibration or the pulsation is the fundamental mode, in which the amplitude of the gas's movement decreases smoothly toward the confined end. There are also harmonic modes that differ from the fundamental mode in that the periods of the vibration or pulsation are shorter and the distribution of the gas's amplitude of movement is different. There are nodes, or stationary points, within the gas, and the amplitude is appreciable only at the free end. Which mode is excited in either an organ pipe or a star depends on how and where the driving force is applied.

Once a star's pulsation properties and equilibrium properties have been determined it is useful to compare them and to look for correlations. The comparison can best be shown in the Hertzsprung-Russell diagram, which displays the luminosity of stars with respect to their effective temperature [*see illustration on page 66*]. In the diagram large stars lie at the upper right and small ones at the lower left. Most normal stars lie in or near the main sequence, a band that crosses the diagram from the top left to the bottom right. Along the main sequence luminosity and effective temperature increase uniformly with the mass of the star. The distribution of pulsating variables is completely different from that of normal stars. Most of the pulsating variables lie in a nearly vertical band known as the Cepheid instability strip, which merges into a broader instability region at the top right. Within these regions many if not most stars pulsate.

Pulsating stars are classified according to their pulsation properties (particularly their light curves) and their equilibrium properties. One large class is the class of Cepheid variables: giant yellowish stars. This group is subdivided into the classical Cepheids such as Delta Cephei, and the Population II Cepheids, which are very old stars. The two types of Cepheid have similar light curves, but they have quite different ranges of mass and age. They arrived at their common location in the Hertzsprung-Russell diagram by entirely different evolutionary routes.

The RR Lyrae stars, along with the Population II Cepheids, are among the very oldest stars in our galaxy and are abundant in globular star clusters. There are two subclasses of the RR Lyrae stars, those that pulsate in the fundamental mode and those that pulsate in the first harmonic mode.

The Delta Scuti stars and the dwarf Cepheid variables also share a common location in the Hertzsprung-Russell diagram, but they may also differ in mass and age. Their light curves can be com-



MODES OF PULSATION OF A STAR (*right*) are not unlike the modes of vibration of the column of air in an open organ pipe (*left*). The arrows show the relative amplitude and direction of the gas's motion in both cases at the moment when the gas at the free end is moving outward at its highest velocity. In the fundamental mode the amplitude of the vibration is greatest at the free end and decreases inward to the fixed end (*a*). In the first harmonic mode there is a node, or stationary point, in the gas (*b*). In the second harmonic mode there are two nodes (*c*). The period of the modes of vibration in the organ pipe is proportional to the length (L) of the pipe. The period of the modes of pulsation in the star is roughly proportional to the radius (R) to the power of 1.5. In each case the periods of the harmonic modes are shorter than the period of the fundamental mode. Mode in which gas vibrates in pipe or star is determined by how and where driving mechanism is applied.

plex and apparently irregular, but close analysis shows that they are actually pulsating in two or more modes simultaneously; the result is that their light curves are reminiscent of the forms of sound waves from a musical instrument. The Delta Scuti stars lie at the intersection of the Cepheid instability strip and the densely populated main sequence, and they are the most common kind of pulsating star. The dwarf Cepheids are not actually a subclass of the Cepheid variables but are the smallest of the stars lying in the Cepheid instability strip.

At the top left in the Hertzsprung-Russell diagram, off by themselves, lie the Beta Cephei stars. At least a dozen of the brightest stars in the sky belong to this class, yet the cause of their pulsation remains unknown. At the top right in the diagram lie the irregular variables and the long-period variables: cool red-giant stars with periods of more than 100 days. The best-known of the long-period variables is Mira in the constellation Cetus. Both types of variables are numerous and conspicuous; some of them differ by as much as 10 magnitudes from their dimmest to their brightest.

The distribution of pulsating variables in the Hertzsprung-Russell diagram is not random. The existence of definite regions of instability suggests that certain combinations of a star's effective temperature and luminosity favor a state of pulsation as opposed to a state of rest. Furthermore, if we compare the pulsation properties and the equilibrium properties of pulsating stars of any particular class, we discover another important correlation: The period of the pulsation is closely related to the luminosity. To be precise, the greater the luminosity, the longer the period. The period-luminosity relation for the Cepheid variables was first discovered by Henrietta S. Leavitt in 1908 in the course of a study of variables in the Small Cloud of Magellan, a companion galaxy of our own. The exact form of the relation was worked out over the past two decades by John D. Fernie, Robert P. Kraft, Allan R. Sandage and Gustav A. Tammann [see "Pulsating Stars and Cosmic Distances," by Robert P. Kraft; SCIENTIFIC AMERICAN, July, 1959]. Knowledge of the period-luminosity relation makes it possible to determine the luminosity, and hence the distance, of any Cepheid variable whose period can be measured. The Cepheids are extremely luminous stars and can be seen at great distances from the sun. They have been utilized to determine distances within our own galaxy and distances to nearby galaxies, and they are one of the fundamental tools of observational cosmology.

Why should there be a relation between the period and the luminosity of a pulsating star? In an organ pipe the period of vibration of the column of air is proportional to the length of the column. In a star the period of pulsation is proportional to the radius of the star, raised to the power of 1.5. The star's luminosity is proportional to its surface area; because surface area is proportional to the square of the radius of the star the luminosity too depends on the radius. Therefore since both the period of the pulsation and the luminosity are related to the star's radius, they are related to each other. The fact that the luminosity of the star depends on its effective temperature means that the period also depends on the effective temperature, although in a minor way.

We have assumed so far that a star's pulsation properties and its equilibrium properties are themselves constant. Actually they are not, since all stars slowly evolve. Our present view of the stars is only a snapshot in the history of their lives. Stars evolve because their supply of energy is finite. That energy is generated by a series of nuclear reactions that first fuse hydrogen into helium, then fuse helium into carbon and eventually may fuse carbon into heavier elements. Both during and between each of these phases of energy production the internal structure and the external properties of a star change profoundly. The rate of change depends on three factors: the total mass of the nuclear fuel, the energy content of the fuel per unit of mass and the rate at which the star is radiating away its energy. The most luminous stars exhaust their fuel in a few million years, whereas a star such as the sun should have a lifetime of at least 10,000 million years.

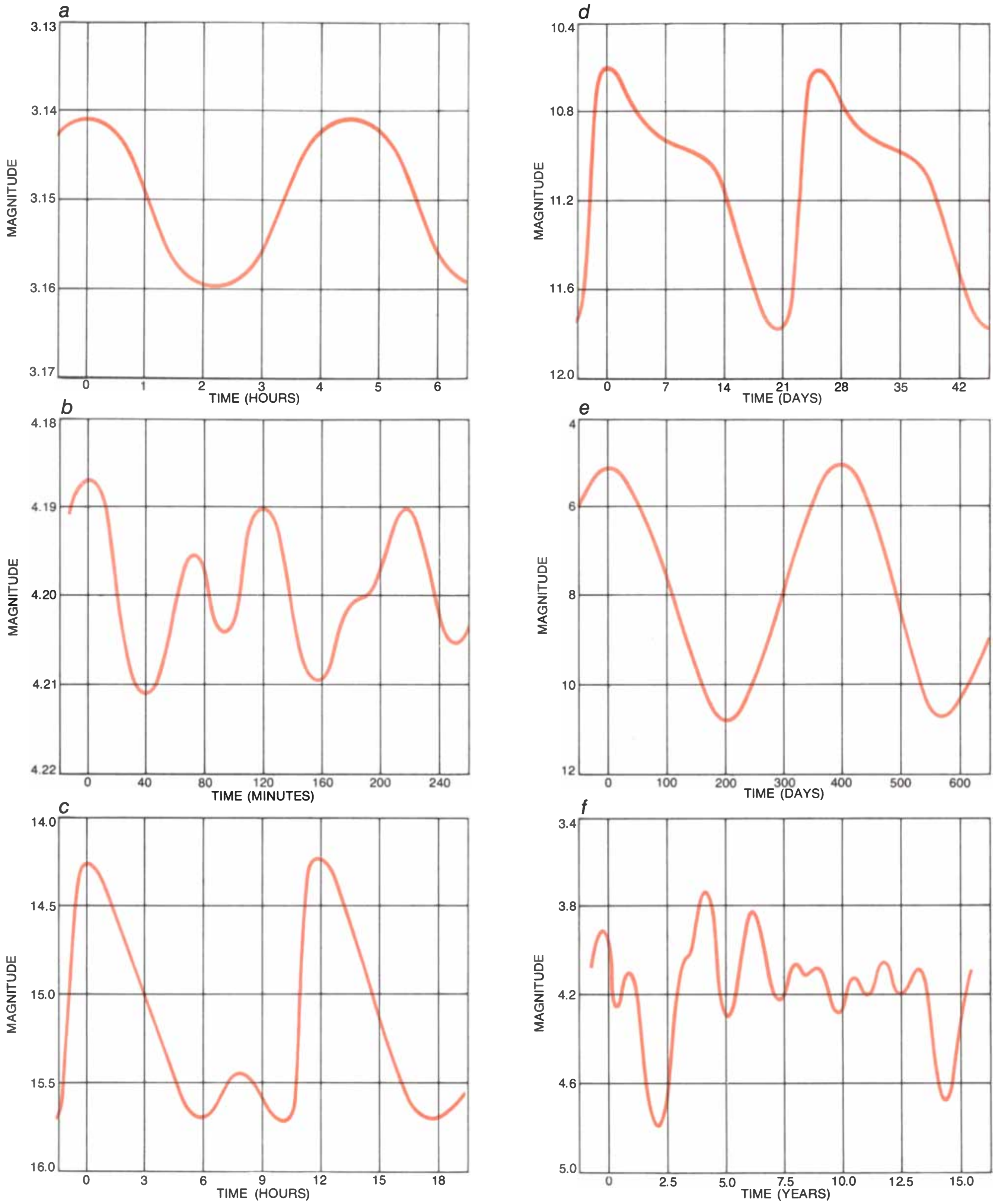
The evolutionary changes in a star's luminosity and effective temperature can be calculated theoretically and shown as an evolutionary track in the Hertzsprung-Russell diagram. Such calculations have been made by Icko Iben, Jr., who is now at the University of Illinois [see "Globular-Cluster Stars," by Icko Iben, Jr.; SCIENTIFIC AMERICAN, July, 1970]. The tracks begin on the main sequence at a position that depends on the mass of the star. Throughout most of their lifetime stars remain near the main sequence. As the hydrogen in their core is depleted they expand and move to the right in the Hertzsprung-Russell diagram. Their evolutionary tracks become quite complicated and may make one excursion or more into the regions of instability. Finally the tracks move rapidly to the lower left, where the stars become tiny white dwarfs. Evolution and the wide variety of masses with which stars are formed together account for the diversity of stellar objects.

Evolution affects pulsation in quite a number of ways. First, a star pulsates only if its evolutionary track carries it into an instability region. Billions of years from now the sun's evolutionary track may carry it into one of the instability regions to become a long-period variable or perhaps a dwarf Cepheid.

Second, the rate at which the evolutionary track crosses an instability region determines how long the star remains a pulsating star and therefore determines the number of stars in that region at any one time. Occasionally we might even witness a star entering an instability region or leaving it. A nonpulsating star

CLASS	PERIOD	RANGE	MASS (SUN = 1)	AGE
BETA CEPHEI STARS	3 TO 6 HOURS	SMALL	10 TO 20	YOUNG
DELTA SCUTI STARS	.5 TO 6 HOURS	SMALL	1.5 TO 3	YOUNG
DWARF CEPHEIDS	1 TO 6 HOURS	MODERATE	.5 ?	OLD?
RR LYRAE STARS	6 TO 24 HOURS	MODERATE	.5	OLD
CLASSICAL CEPHEIDS	1 TO 50 DAYS	MODERATE	2 TO 20	YOUNG
POPULATION-II CEPHEIDS	1 TO 50 DAYS	MODERATE	.5	OLD
LONG-PERIOD VARIABLES	100 TO 500 DAYS	LARGE	.5 TO 2	MIXED
IRREGULAR VARIABLES	1 TO 10 YEARS	MODERATE	2 TO 10	YOUNG

PULSATING STARS ARE CLASSIFIED according to their period, range and physical properties. As can be seen, the classes include stars with a wide variety of masses and ages.



LIGHT CURVES OF SIX VARIABLES compare the properties of different types of pulsating star. The Beta Cephei stars (*a*), exemplified by the prototype Beta Cephei, are hot, luminous stars with short periods of pulsation and small ranges in brightness. The cause of their pulsation is completely unknown. Delta Scuti stars (*b*) are by far the most numerous of the pulsating stars, but they are inconspicuous because of their small range in brightness. The light curve shown is for the star Epsilon Cephei. The variations in brightness look irregular, but they are actually a superposition of two or more harmonic modes in which the star is pulsating. RR Lyrae stars (*c*) have a short period and a sharp maximum; this star is Variable 19 in the globular cluster Messier 5. Population II Cepheids (*d*) have a light curve that is very similar to the light curve for the classical Cepheids (*see light curve for Delta Cephei at top of page 68*), as is shown in the case of Variable 42 in Messier 5. Long-period variables (*e*), as is shown by the light curve of T Cephei, have a very large range in brightness (up to six magnitudes). Irregular variables (*f*) have a somewhat smaller range and in spite of their designation do have some degree of periodicity. Light curve that is given as an example is for the star Mu Cephei.

may become a pulsating star or vice versa. Although such an event would be rare, there are one or two cases where it may have been observed. In 1966 Fernie and Serge Demers of the University of Toronto discovered that the Population II Cepheid RU Camelopardalis had apparently stopped pulsating. Since then the star has shown some signs of regaining its pulsation, and its future behavior will be watched with interest. The simplest interpretation is that RU Camelopardalis has moved out of the instability strip.

Although evolutionary changes in luminosity and effective temperature are far too slow to be observed directly, the changes can be detected indirectly in a novel way. If the star's pulsation period is related to its luminosity, and if the luminosity changes because of the star's evolution, then the period should also change, albeit rather slowly. The rate of change in the period should reflect the rate at which the star is evolving.

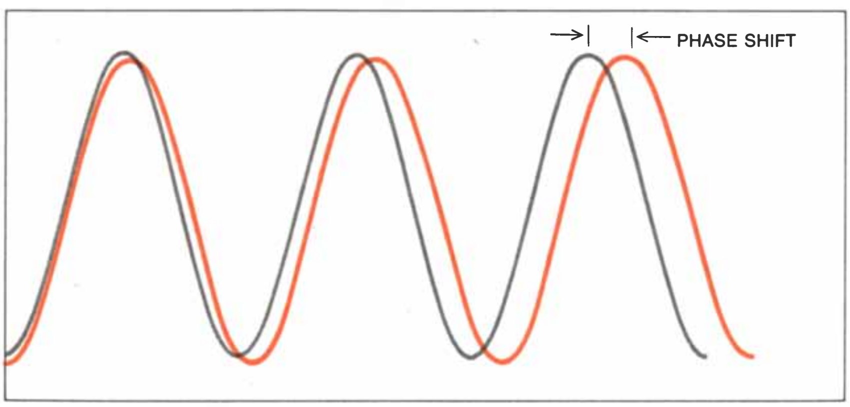
The rate of change in the period can be measured by comparing the observed time at which the star reaches its maximum brightness with a computed time based on the assumption that its period is constant. The difference between the times is the phase shift. The beauty of monitoring the phase shift is that a change in the star's period produces a cumulative shift. Consider the following analogy. A standard clock that always completes a cycle in 86,400 seconds will always register perfect time in a day that is exactly 24 hours long. A clock that always completes its cycle in 86,401 seconds will lose one second per day; with respect to the standard clock it will accumulate an error of one second after one day, two seconds after two days, three seconds after three days and n seconds after n days. The cumulative error is proportional to the elapsed time, and when it is plotted against the elapsed time, the resulting curve is a straight line. After n days (if n is large) the clock's error of one second per day can be measured to great accuracy by dividing the cumulative error by n. The period of a pulsating star can be measured with great precision in the same way.

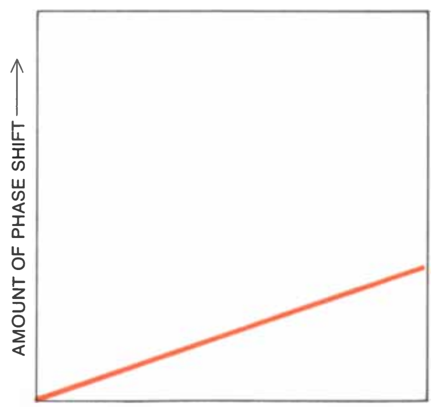
Now consider a clock whose rate is not constant but is slowing down, with the length of its period increasing by two seconds per day. If its period is 86,401 seconds the first day, then it will complete its cycle in 86,403 seconds the second day, 86,405 seconds the third day and so on. With respect to the standard clock it will accumulate an error of one second after one day, four seconds after two days, nine seconds after three days and n^2 seconds after n days. The cumulative error is proportional to the square of the elapsed time, and when it is plotted against the elapsed time, the resulting curve is a parabola. The same kind of curve obtained by monitoring a star's progressive phase shift with respect to time should yield information on the way a star is slowly evolving.

RR Lyrae stars in globular clusters are ideally suited to such studies. The time when they reach their maximum brightness is accurately defined, and many stars can be measured on a single photographic plate. The study of variable stars in globular clusters was initiated by Solon I. Bailey at the Harvard College Observatory in the 1890's, continued by Harlow Shapley and then for nearly 50 years by Helen S. Hogg of the University of Toronto. An intensive study of the globular cluster Messier 5 by Hogg and Christine M. Coutts, utilizing photographs made between 1895 and 1965, shows that the periods of some of the RR Lyrae stars are indeed changing at a constant rate as evolution theory predicts. Other stars behave differently: their periods remain constant for many years, change abruptly and again remain constant for many years. Although the period changes are quite small, typically being equivalent to one second per century, they can be measured with great accuracy by the phase-shift method. Furthermore, the longer the studies are continued, the more accurate and productive they will be.

The noted British astronomer A. S. Eddington made the first serious attempt to explain the actual cause of stellar pulsation. By 1917 he had derived a wave equation that describes the period and the relative amplitude of pulsation throughout the star. He found that the relative amplitude decreases rapidly toward the center of the star because the density there is so high. As a result the center of the star neither affects nor is affected by the pulsation to any great extent. That condition makes the understanding of stellar pulsation much simpler because we can ignore the complicated changes in the chemical composition of the center of the star due to nuclear reactions.

Eddington's wave equation described only the mechanical aspects of the pulsation, not its source or its absolute amplitude. Those properties must be determined from the laws of thermodynamics. Eddington reasoned that if a star pulsates, then it must somehow have acquired mechanical energy both to begin its pulsation and to maintain it against the dissipative or frictional forces. The star must therefore function as a heat en-





PHASE SHIFT is the difference between the observed time of maximum brightness of a star and the calculated time (on the assumption that the period of pulsation is constant). When the amount of the phase shift is plotted against the elapsed time, the

gine, converting a small fraction of its abundant supply of radiant energy into motion. The basic requirement of a heat engine is that it absorb excess heat when it is hot and compressed and release the excess heat when it is cool and expanded.

Eddington suggested two mechanisms for making a star behave as a heat engine. The first mechanism was the processes that generated the energy at the center of the star, the nature of which was not known in his day. If the processes were sensitive to changes in temperature (which they are), they might produce excess heat energy when the star was most compressed. Eddington discarded this mechanism, because he knew that even if it worked, the amplitude of the gas's pulsation at the star's center would be too small to create the observed pulsation of the entire star. Later detailed calculations verified Eddington's conclusion.

The second mechanism Eddington proposed was a hypothetical "valve" near the surface of the star, where the pulsation amplitude was greatest. If the valve closed, trapping heat, when the star was hottest and most compressed, and opened, releasing heat, when the star was coolest and most expanded, then the requirements for a heat engine would be satisfied.

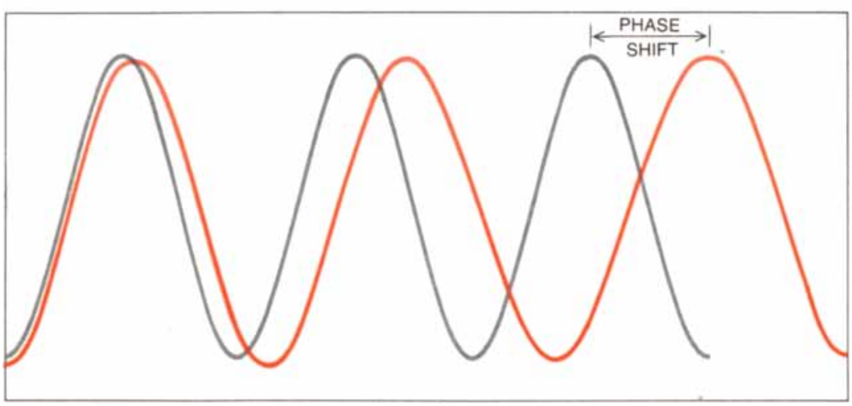
Eddington considered whether or not the natural opacity of the atoms in a star could provide the valve. He determined that under normal circumstances they would not. The atoms in the interior of most normal stars are fully ionized: stripped of all their electrons. Fully ionized atoms become more transparent when they are compressed, so that they let excess heat escape and the valve mechanism does not work. Indeed, most stars do not pulsate and therefore do not behave as heat engines.

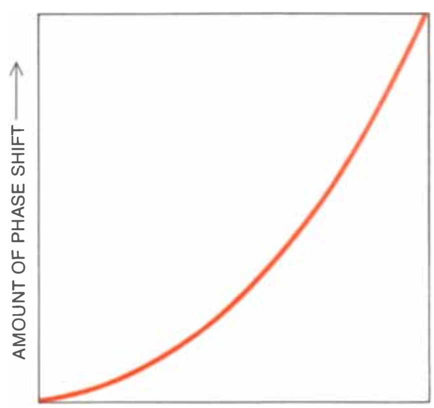
There is one situation where atoms may become more opaque when they are compressed, namely when the atoms are only partially ionized or, to be more precise, when the atoms of a particular element are present in two states, with the atoms in one state being stripped of one more electron than the atoms in the other. A region in a star where atoms of a particular element are only partially ionized is called an ionization zone. Eddington suggested that ionization zones could act as heat engines and thus could cause the star to pulsate. He was quite right. In Eddington's time, however, it was believed that stars were made up of heavy elements such as iron and that the lighter elements such as hydrogen and helium were quite rare. Hence Eddington could not specify the true source of the pulsation, which is the ionization zones of hydrogen and helium. In stars with an effective temperature of about 7,000 degrees Kelvin the ionization zones are close enough to the surface for the gas's pulsation amplitude to be large, and at the same time they are deep enough within the star for their density to be high and for them to have enough mass and momentum to drive the pulsation.

After Eddington's death the problem of stellar pulsation was taken up by the Russian astronomer S. A. Zhevakin, John P. Cox of the University of Colorado and Charles A. Whitney of the Smithsonian Astrophysical Observatory. Working with more realistic estimates for the chemical composition of the stars and more sophisticated models for their structure, they showed theoretically that the second ionization zone of helium, in which helium atoms already stripped of one electron lose the second electron as well, was capable of causing pulsation in stars resembling Cepheid variables. Cox also demonstrated that certain properties of the ionization zones in addition to their opacity contribute to their destabilizing effect. For example, when an ionization zone contracts, much of the energy liberated by the contraction goes into ionizing the remaining un-ionized atoms in the zone and little energy goes into raising the temperature of the zone's gas. In fact, the ionization zone actually becomes cooler than its surroundings and thus absorbs excess heat, as a heat engine in its compression cycle should. When the star subsequently expands, the excess heat increases the pressure within the zone and provides the extra "lift" needed to build up and maintain the star's pulsation.

Pulsation theory advanced rapidly in the 1960's with the advent of large electronic computers. With the aid of such machines it became possible for investigators to amass detailed tables of the physical properties of the matter in a star. Arthur N. Cox and his collaborators at the Los Alamos Scientific Laboratory compiled elaborate tables of the opacity of matter in a star with respect to temperature and density. Since pulsation depends heavily on the changes in the opacity of ionization zones, the opacity tables are essential for understanding the pulsation. The tables are also essential for calculating the evolutionary tracks of stars.

Computers were more directly applied to the study of stellar pulsation and of stellar evolution in general by acting as tools for building models. In a star the temperature, the pressure and the density decrease smoothly from the center





resulting graph will be a straight line if the period of the star is constant (*colored curve at left*), even if it differs from the predicted period (*gray curve*). If the star's period is changing uniformly as a result of stellar evolution, however (*right*), graph will be a parabola.

to the surface. In a model of a nonpulsating star the star is approximated by a series of static concentric shells, each with its own constant temperature, pressure and density. If we hypothesize a certain mass for the star, the temperature, the pressure and the density for each shell can be determined by applying the law of hydrostatic equilibrium and the laws of the conservation of mass and of energy. The resulting equations must be solved numerically for each shell.

In a model of a pulsating star the shells are not static. The equation of hydrostatic equilibrium must be replaced by Newton's second law of motion, which states that the force exerted on the shell by the rest of the star is equal to the mass of the shell times its acceleration. Since acceleration involves the change in velocity with respect to time, the problem becomes more complicated: the equations now have an extra variable, namely the time. If the shells are given some small initial motion, will the motion grow and develop into pulsation or will it decay until the shells come to rest? If the motion grows, will it level off at some constant amplitude? If it does, why does it do so and at what amplitude? This approach to the study of how pulsation increases and decreases with respect to time is called the nonlinear approach; answering the questions properly can take several hours of expensive computer time.

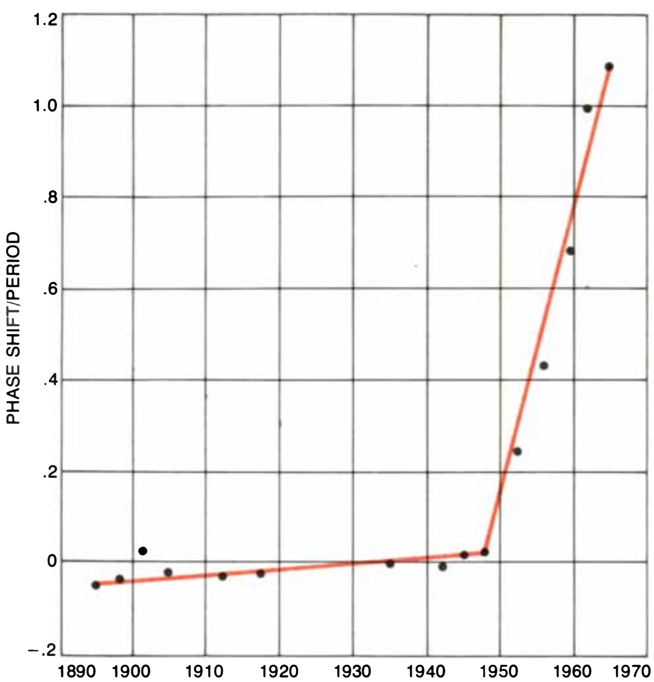
The problem can be made simpler (and cheaper) by assuming that the pulsation is small and sinusoidal, that is, that it repeats each cycle in a specific way. The time variable thus disappears from the equations because we have now assumed how the star will pulsate with respect to time. This approach is called the linear approach, and it enables us to determine whether or not a certain model star will sustain small pulsations. We then hope that if the model can develop small pulsations, it will be able to develop large ones as well. The linear approach is an economical way of mapping the instability regions in the Hertzsprung-Russell diagram by choosing an array of models each of which has a different luminosity and temperature. It does not, however, tell us how large the pulsation will eventually become. That information can come only from the nonlinear approach.

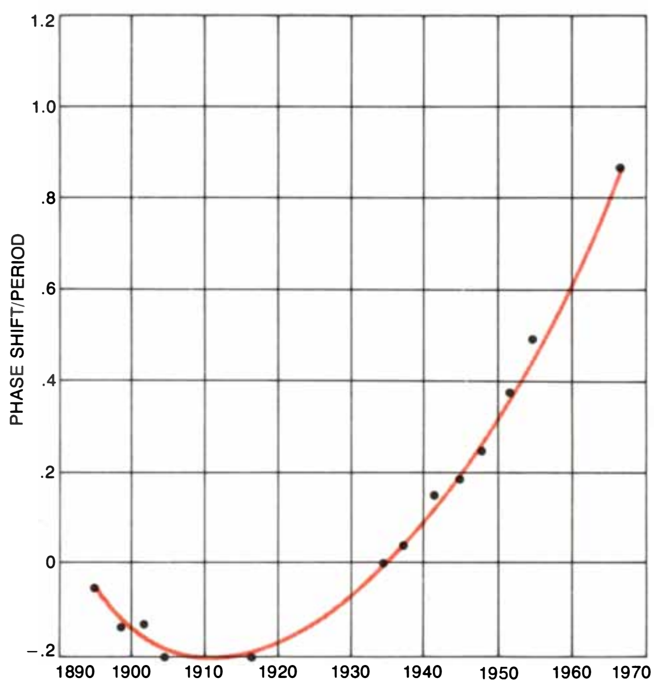
Robert F. Christy of the California Institute of Technology conducted the first major study with the nonlinear approach. He constructed models of RR Lyrae stars with different values of mass, luminosity, effective temperature and chemical composition. He then compared the pulsation properties predicted by the models with the properties observed in real stars. In that way he was able to theoretically determine the mass, the luminosity and the chemical composition of the real stars.

The RR Lyrae stars are well suited to studies such as Christy's: they are intrinsically interesting because they are such old stars, they are well studied because so many of them are conveniently located together in globular clusters and they are all confined to a small region in the Hertzsprung-Russell diagram (eliminating wide ranges of mass, luminosity and composition). Even with all these advantages Christy's study required more than 100 separate models.

Of the many results of Christy's work perhaps the most important concerns the abundance of helium in the RR Lyrae stars. An ionization zone of hydrogen is created at a temperature of about 10,000 degrees K. The first ionization zone of helium, in which helium atoms are stripped of one electron, is established at 12,000 degrees; the second is established at 40,000 degrees. If the star is too hot, the ionization zones of hydrogen and helium lie too close to the surface of the star, where their density is too low for them to have enough mass to drive the pulsation. The star is then stable. In a cooler star they are located at a somewhat greater depth, where they do have enough mass to drive the pulsation. The star will pulsate in the first harmonic mode if the ionization zones are relatively close to the surface and in the fundamental mode if the zones are somewhat deeper.

Imagine that the star contained no helium. There would be no helium ion-



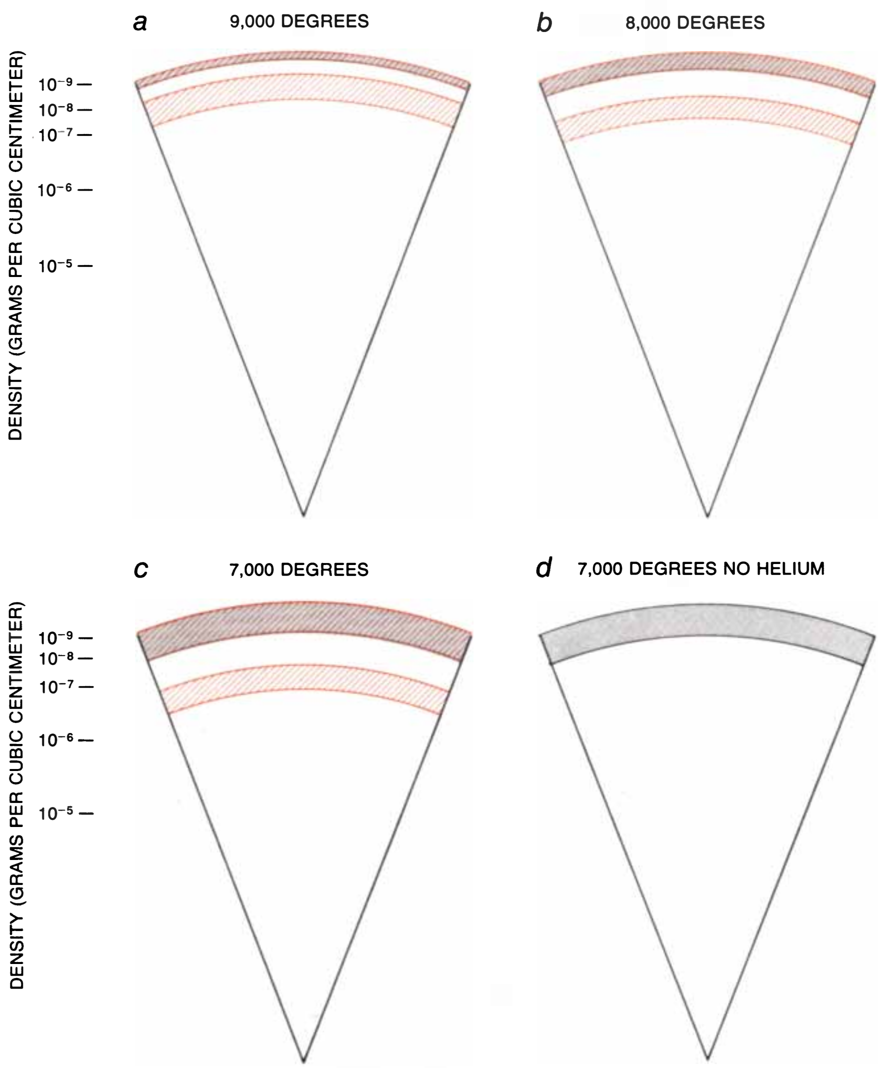


PERIODS OF PULSATING STARS CHANGE with respect to time as the stars evolve. Period of Variable 70 in Messier 5 lengthened uniformly (*right*). Period of Variable 19 remained constant until 1947, changed and has remained constant ever since (*left*).

ization zones. In a star that would normally pulsate because of the valve effect of all three ionization zones acting together there would now be only the hydrogen ionization zone, which is too close to the surface to single-handedly overcome the dissipation or friction in the rest of the star. The existence and the nature of the pulsation in the RR Lyrae stars shows that they must comprise appreciable amounts of helium; indeed, they must be as much as 30 percent helium by weight. Pulsating stars much younger than the RR Lyrae variables, however, are also composed of about 30 percent helium. The presence of helium in such great abundance in the outer layers of all types of pulsating stars implies that the helium must have been a part of the gas from which they were formed. In the case of the ancient RR Lyrae stars this abundance indicates that helium was a primordial element and that it was already present in large amounts when the oldest stars in our galaxy were created. The most likely explanation is that the helium was formed in the "big bang" that marked the beginning of the universe as we know it.

Although the first results from the nonlinear calculations were interesting and useful, they had their limitations. The equations involved were so complex that shortcuts invariably had to be taken, and some of them had undesirable side effects. The calculated light curves show variations that sometimes are only artifacts of the computations, and there is no clear indication of how to tell those variations from the variations that are supposed to be real. The masses obtained by fitting the pulsation properties of real Cepheids with the pulsation properties of theoretical models do not agree with the masses obtained by fitting those same real Cepheids on their theoretical evolutionary tracks in the Hertzsprung-Russell diagram. There is also a problem in trying to determine what happens if a star is susceptible to pulsation in both the fundamental mode and the harmonic modes. The star may take a much longer time to make up its mind which to choose than the astronomer can afford to follow on his computer. Several elegant methods are being developed to handle such problems.

Some larger questions still remain. The valve mechanism explains why stars hotter than those in the instability strip are stable because the ionization zones are too close to the star's surface; it does not explain why some stars, such as the



IONIZATION ZONES within a star, zones where atoms are stripped of some or all of their electrons, are responsible for causing a star to pulsate. If the ionization zones of hydrogen (*gray*) and helium (*hatching*) lie at a critical depth below the surface, they will act as a heat engine and drive the pulsation. If the zones are too close to the star's surface, as they are when the star has a high effective temperature (*a*), their density is too low and there is not enough matter in the zones for them to drive the pulsation. At a somewhat lower temperature the ionization zones will drive the pulsation in the first harmonic mode (*b*). At an even lower temperature they will drive the pulsation in the fundamental mode (*c*). If star has little or no helium (*d*), however, hydrogen ionization zone would be incapable of single-handedly driving pulsation, and the star would not pulsate. The fact that helium is required provides a method for determining the abundance of helium in pulsating stars.

sun, that are cooler than those in the instability strip are stable. A qualitative explanation does exist: In stars with an effective temperature of less than 6,000 degrees the gases are so opaque that radiant energy can no longer flow efficiently and the energy is carried by convection instead. Since the valve mechanism depends on the gas's absorbing the radiant energy, it will not operate efficiently when the energy is not carried primarily in that form. Although it would be heartening to be able to verify this explanation, convection is the most poorly understood process in astrophysics. The same problem hampers our understanding of the cool long-period variables.

To confuse matters further, a few nonpulsating stars have been found in the instability regions. What mechanism stabilizes them? Rotation perhaps? And what mechanism causes the pulsation of the Beta Cephei stars, whose effective temperature is so high that the ionization zones of the common elements are right at the surface of the star? It is clear that although we have a basic understanding of why stars pulsate, many interesting problems remain to be solved.

VISUAL MOTION PERCEPTION

The eye has no shutter, and yet a moving world does not appear as a blur. The visual system works not like a camera but more like a computer with a program of specific mathematical rules

by Gunnar Johansson

The eye is often compared to the camera, but there is one enormous difference between the two. In all ordinary cameras a shutter "freezes" the image; even in a television camera, which has no shutter, the scanning raster of an electron beam serves the same purpose. In all animals, however, the eye operates without a shutter. Why, then, is the world we see through our eyes not a complete blur? As we walk down a street the buildings we pass seem quite stationary. We do not perceive them as the bundles of streaks they optically create on our retina. Other pedestrians and moving vehicles all seem to be traveling through the same static visual space with sharp outlines, even though they are moving in various directions and with quite different velocities. Whether we are standing still or moving through space the eye effortlessly sorts moving objects from stationary ones and transforms the optical flow into a perfectly structured world of objects, all without the benefit of a shutter. How is this remarkable feat accomplished?

From the evolutionary point of view the feat was clearly necessary for survival. The eye has evolved to function essentially as a motion-detecting system. The concept of a motionless animal in a totally static environment has hardly any biological significance; the perception of physical motion is of decisive importance. In many lower animals the efficient perception of moving objects seems to be the most essential visual function. A frog or a chameleon, for example, can perceive and catch its prey only if the prey is moving. A motionless fly, even within easy reach, goes quite unnoticed.

Evidence for a similar dependence on changes in the visual stimulus pattern can also be demonstrated in man. In experiments where a special device holds an image motionless on the retina the corresponding percept rapidly fades and disappears. Tennis and many other sports testify to man's remarkable ability to visually determine the precise spatiotemporal position of a small fast-moving object.

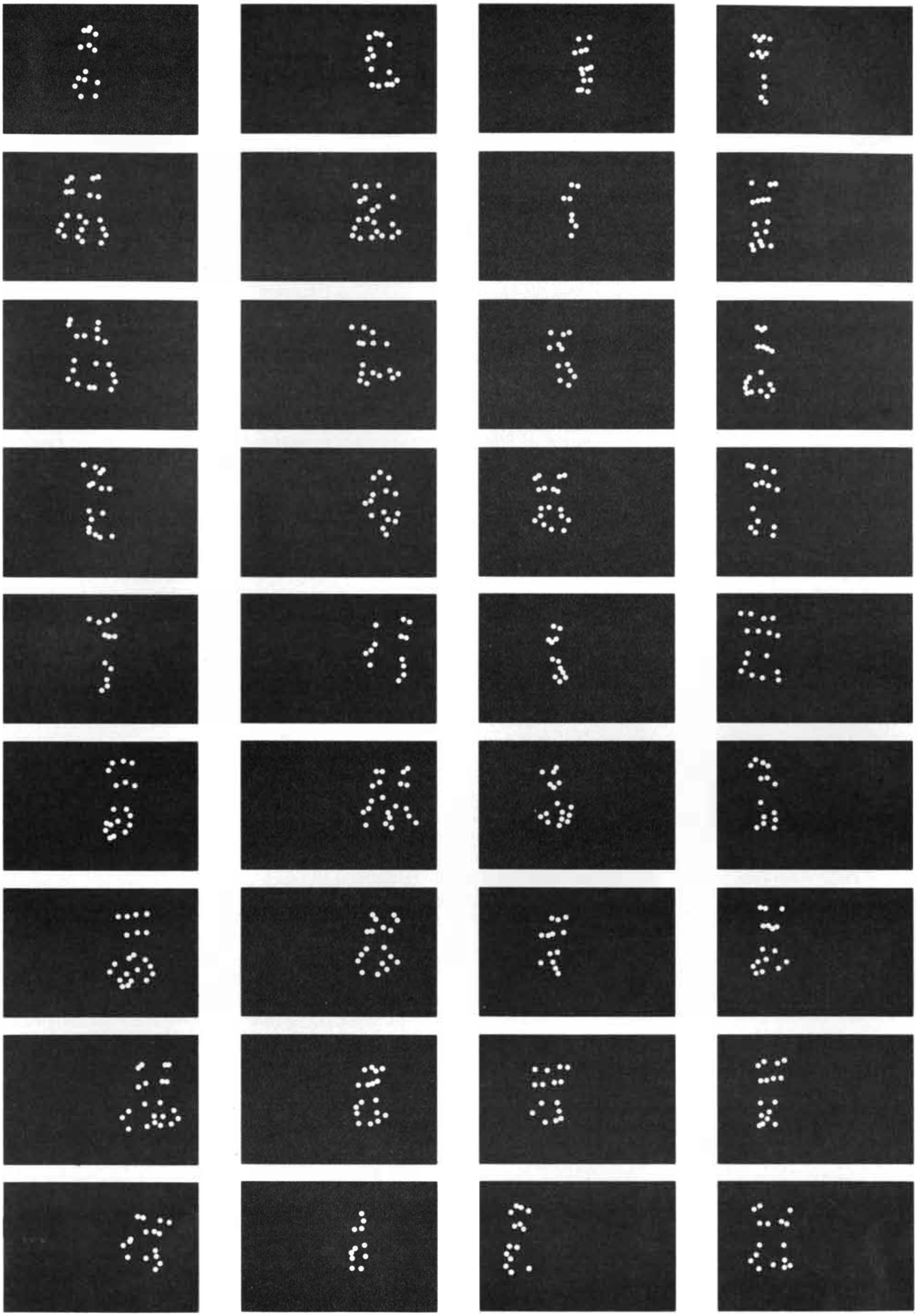
The traditional comparison of the eye and the camera serves the useful didactic purpose of explaining how light rays are focused to produce a two-dimensional image on the surface of the retina. Difficulties arise, however, when the photoreceptors embedded in the retina are likened to a photographic film. Unless one deliberately wants to get a blurred image on the film it must be exposed to the incident light rays for only a brief period, just enough for the photosensitive chemicals in the film to "capture" the image. Although it is true that the retinal receptors have a similar ability to capture photons, their real function is not to capture images but to mediate changes in light flux. The light impinging on the receptors (the rods and the cones) gives rise to a continuous change in the structure of photosensitive molecules. The change in structure releases a flow of ions in the receptor, culminating in a bioelectric signal that travels from the receptor into adjacent nerve cells.

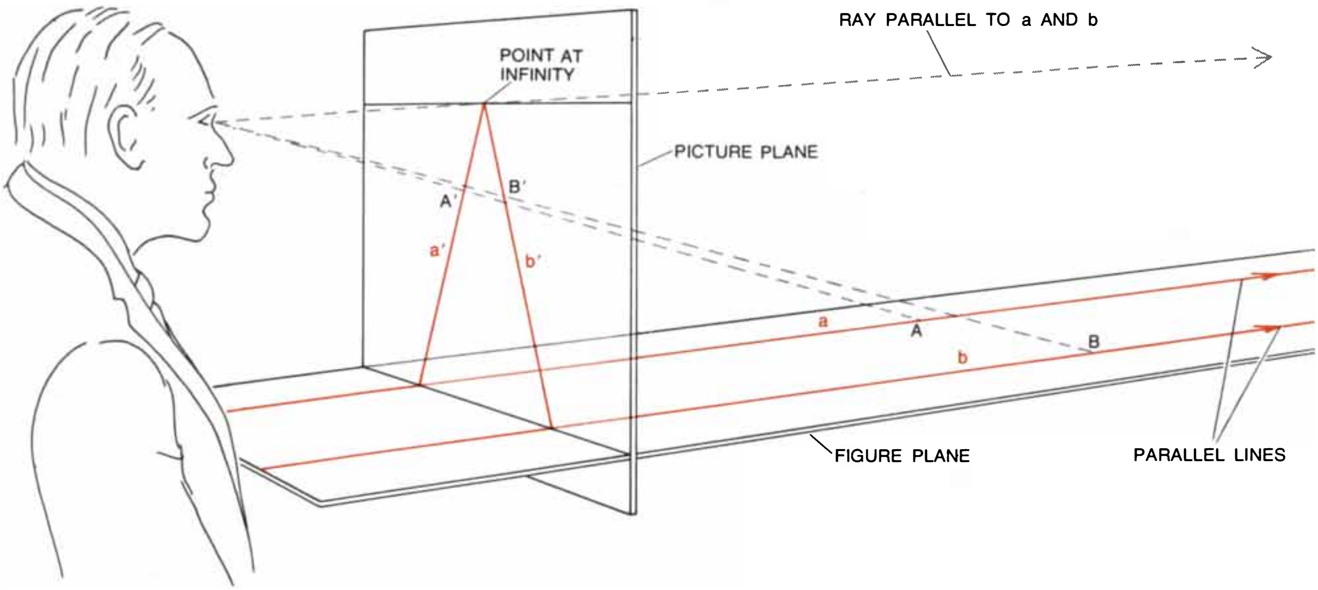
The strength of the signal varies with the light flux. Within a few milliseconds the myriad changes in signal pattern over the entire retina are combined and transformed by an intricate neural network within the retina itself, by other networks at relay stations in the midbrain and finally by the neural networks within a number of receiving terminals in the cerebral cortex. The result at the conscious level is the perception of motion in visual space. Thus the eye is basically an instrument for analyzing changes in light flux over time rather than an instrument for recording static patterns. Roughly speaking, without a change in the light striking the receptor there would be no change in ion flow and no neural response.

In studies of visual perception it is often important to distinguish between monocular and binocular vision. For the range of phenomena I shall take up here, however, the contribution made by binocular perception can be ignored. In our laboratory at the University of Uppsala my colleagues and I have contrived a variety of experiments to examine how the eye deals with moving visual stimuli. We include under this heading the motion of stationary objects perceived by a moving observer as well as the motion of moving objects perceived by a stationary observer.

As an introduction to our experiments, consider what happens when you use a camera to make a picture of a friend. You look through the viewfinder and customarily take a few steps forward or backward until you have the subject properly framed and the image expand-

TWO FIGURES DANCING IN THE DARK appear on the opposite page in a sequence of 36 motion-picture frames from a film made in the author's laboratory at the University of Uppsala. Each dancer is "outlined" by 12 lights: two each at the shoulders, elbows, wrists, hips, knees and ankles. This sequence, which proceeds in vertical columns starting at upper left, consists of every sixth frame from a portion of the film. Naïve subjects shown the film can tell in a fraction of a second that they are seeing the movements of two people.





IN PERSPECTIVE DRAWING parallel lines converge at a fictitious "vanishing point" located on the viewer's horizon. The diagram shows how points *A* and *B* on parallel lines *a* and *b* are converted into their perspective equivalent on a transparent screen.

ed or contracted to the desired size. As you move toward the subject every optical element in the viewfinder streams radially outward from a central point. Conversely, when you step backward, the image contracts radially toward the center. If you are a careful photographer, you probably also check the effects of moving the camera up and down and from side to side. Such movements generate optical flows considerably more complex than the radial flow produced by moving directly toward the subject. All such changes in the viewfinder, however, follow the laws of central perspective. They are continuous perspective transformations.

The optical flow of images into the viewfinder of a camera (or into the camera itself when the lens is open) corresponds to the optical flow impinging on the retina during locomotion. From the geometrical point of view it does not matter whether it is the camera that is moving or the subject in front of the camera. It would be trivial to say that asking your friend to take a step toward you has the same effect on the size of his image as your moving a step toward him. It is significant, however, that in the first case the image of the surrounding environment remains fixed and in the second the image of the environment expands outward slightly from the optical center. To generalize, when objects move in our field of vision, they give rise to local flow patterns; when we move around in the environment, there is an optical flow across the entire retinal surface.

In everyday perception the optical flow across the retina usually represents a complex combination of patterns generated by the observer's own motion and patterns generated by the motion of moving objects. Even when the observer is simply standing still or sitting, the sway of his body or small movements of his head add a small "locomotion component" to the flow of the retinal images. Movements of the eye itself introduce a further component into the total flow; the movement can be smooth, as when an observer follows the flight of a ball, or jerky, as when your eye follows these words by a number of saccadic eye movements. The summation of all such optical flows over the retina determines the character of the incessant flow of nerve impulses from the retinal receptors. In order to study the visual information supplied by a light-reflecting space we must consider the geometry of the optical flow reaching the retina.

A theory of the perception of visual space was outlined as long ago as 1709 by George Berkeley (later Bishop Berkeley). The theory was further developed by Hermann von Helmholtz in the 19th century and is still familiar today in a modified version known as cue theory. According to this theory, the two-dimensional image on the retina is visually interpreted as being three-dimensional by a number of cues, or signs. The cues are available not only in the image itself but also in the activity of the oculomotor apparatus. The cues include binocular disparity in the images seen by the two eyes, convergence and accommodation of the lens, size of image, interposition of figures, binocular perspective and so on. The theory also invokes visual-motor experience and learning as important supplementary factors.

Berkeley knew only Euclidean geometry (the discovery of other geometries was still in the future), and as a result he began his study of the relation between a stimulus and a percept by analyzing the retinal image as if it could be adequately measured with a ruler and a protractor. Even today many excellent theorists stay within the tradition of measuring optical projections in millimeters and degrees of arc. This approach has given rise to many artificial problems, such as trying to explain how retinal images of different sizes and forms can evoke perception of the same object.

New geometries that have come into existence since Berkeley's day are free of the Euclidean parallel axiom, which leads to the postulate that parallel lines do not meet. One of the geometries that is not fettered by the parallel axiom is projective geometry. That geometry is of special interest for the study of vision because it is the geometry of optical paths through pinholes and lenses and provides the theoretical basis for perspective drawing. It is characterized as being a nonmetric geometry because it deals exclusively with relations rather than particular measurements.

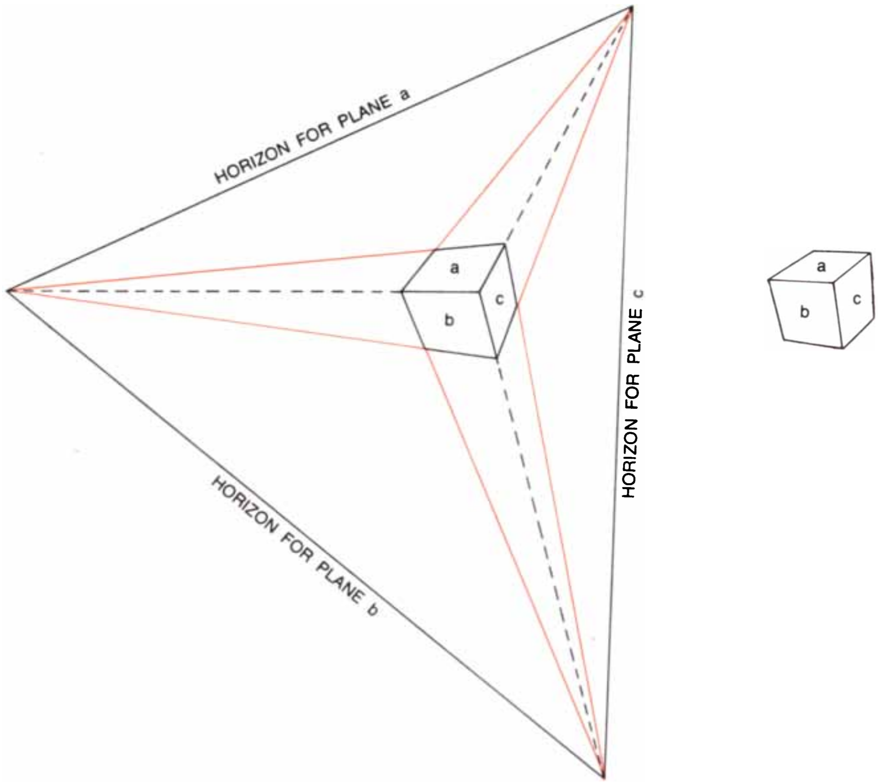
The first comprehensive use of the principles of central perspective in the theoretical analysis of visual space perception was made by J. J. Gibson of Cornell University in his book *The Perception of the Visual World,* published in 1950. Gibson's main thesis is that traditional cue theory is an unnecessary and even misleading construct. According to Gibson, the image itself contains all the information needed for three-dimensional perception, a fact overlooked in cue theory because of its unsophisticated description of the visual stimulus. Mathematical lawfulness in the structural change from point to point in the optical image, involving what Gibson termed "gradients" and "higher-order variables," is the effective stimulus. The gradients and variables are essentially consequences of central projection. Gibson also applied these principles to moving patterns, speaking of stimulus flow rather than stimulus images.

My own thinking closely follows Gibson's. Experimental work over the past two decades has led me to break completely with the Euclidean model and to adopt projective relations as the theoretical foundation for investigations of visual space and motion.

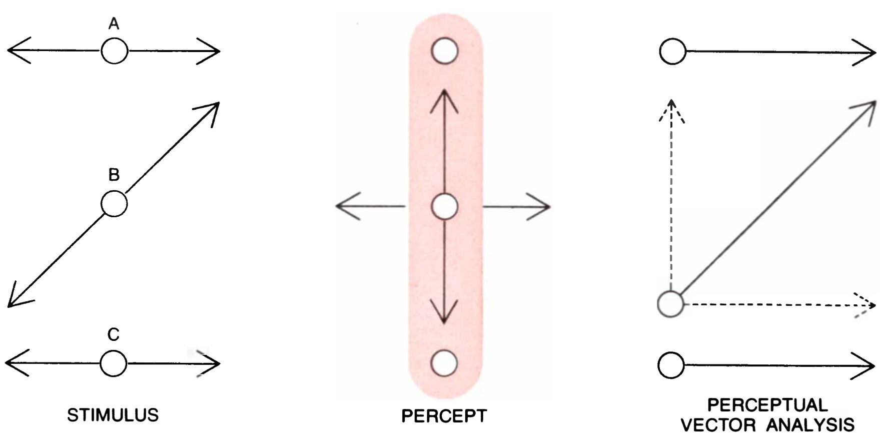
In retrospect it seems strange that it should have been hypothesized, as it was in the classical theory, that organisms searching for spatial information from reflected light developed an eye with a lens and then failed to take advantage of the mathematical laws determining spatial information, available in the trajectories of light through a lens. So strong are the Euclidean and Berkeleyan traditions, however, that a direct experimental approach is needed in order to gain acceptance for a model based on central projection.

The reader may ask how a geometry lacking a fixed metric can be of any use for transferring information about the rigid three-dimensional space that surrounds us, in which the Euclidean metric certainly holds true. The answer is that projective geometry is a geometry dealing with certain relations that remain invariant under perspective transformation. These invariances serve as a counterpart in terms of figural equivalence for the Euclidean figural congruence under the conditions of rigid motion. Mathematicians have also developed a special system of coordinates (homogeneous coordinates) that are determined by distance relations rather than by absolute distances and that make it possible to deal analytically with projective transformations.

For the purposes of the rest of this discussion it is sufficient to say that projective geometry underlies the rules of central perspective [*see illustration on opposite page*]. It is well known that in perspective drawing parallel lines must be pictured as converging at a fictitious "vanishing point." Thus in the perspective system the parallel axiom is abandoned. The angle between the "parallel" lines (actually converging lines) depends on the angle between the figure plane (the surface being pictured) and the picture plane. Hence we know that a rectangular tabletop in a drawing or in a photograph will be trapezoidal, a circular table will be elliptical and so on. No matter how the viewing angle or the distance to an object is changed, the object



CUBE REMAINS A CUBE even when it is seen from different angles. Strictly speaking, each face of the two cubes drawn here is a trapezium. The visual system, however, automatically corrects for the distortions and delivers percept of regular solid with square faces.

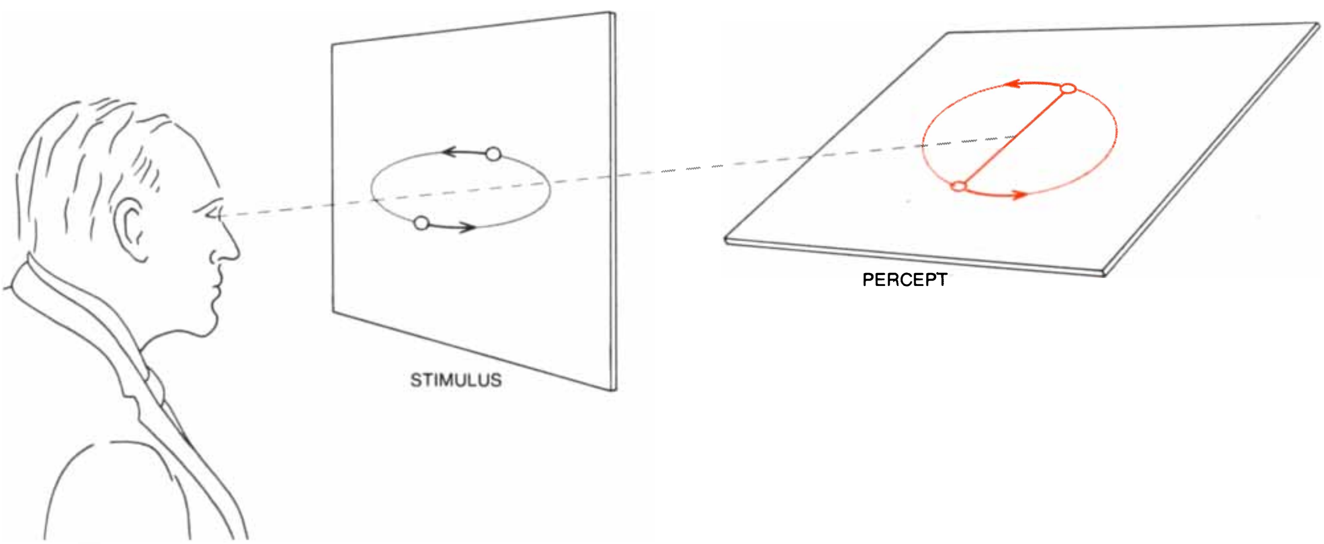


MOVING REFERENCE SYSTEM is formed by three spots of light, *A, B* and *C,* traveling along the paths indicated by the arrows at the left. If seen by itself, *B* simply moves back and forth on a slanting path. When the motions of *A* and *C* are added, however, the three spots form a perceptual unit (*middle*), in which the trajectory of *B* no longer seems to slant. Instead *B* seems to oscillate vertically as if bouncing back and forth between *A* and *C.* In this case the motion of *B* divides into two component vectors: one horizontal and equal to the motion of *A* and *C,* and one vertical, representing the motion of *B* relative to *A* and *C.*

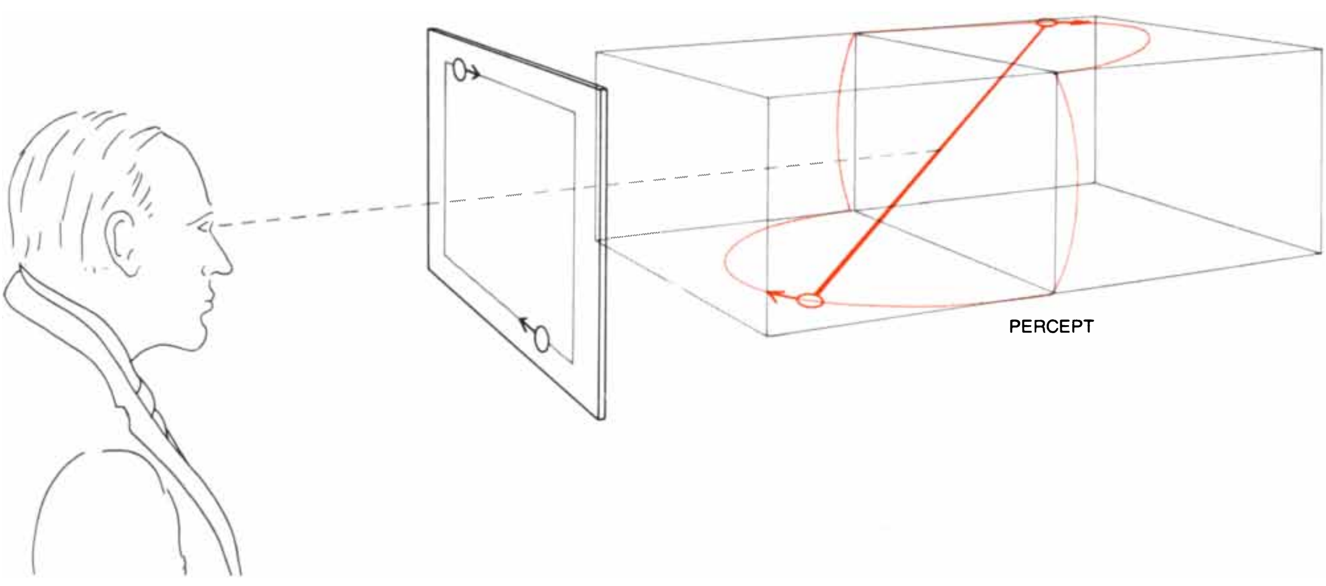
is recognizable as the same object seen at different angles [*see upper illustration on preceding page*]. The forms in the pictures are equivalent because of certain invariant relations, although from a Euclidean point of view they are all different.

From recent studies of motion perception in which continuous figural changes of this type are presented without three-dimensional depth cues we have overwhelming evidence that the visual system spontaneously abstracts relational invariances in the optical flow and constructs percepts of rigid objects moving in three-dimensional space. Indeed, it has been found that continuous perspective transformations always evoke the perception of moving objects with a constant size and shape. This means that the particular projection chosen perceptually by the visual system is one that represents Euclidean invariance under the conditions of motion in rigid three-dimensional space.

A basic and well-established conclusion from a large body of experimental research dating back to the 1920's is that the visual system, in its decoding of a total optical flow, tends to extract components of projective invariances in accordance with specific rules. An example from daily life is perhaps the best way to make this rather abstract statement easier to grasp. My little granddaughter runs across the floor of my study, eager to show me a ladybug walking on her finger. The optical flow produced in my eyes by this scene includes the following components: the light reflected from (1) the floor, the walls and the furniture in my study, (2) the child's body, (3) the child's hand and finger reaching toward me and finally (4) the ladybug moving on the child's moving finger. All these components moving relative to my eyes contribute to the complex optical flow, but



IMAGINARY ROTATING ROD is formed when the stimulus consists of two spots of light moving in an elliptical path. Because the visual system "prefers" to perceive the rod as maintaining a constant length, the viewer has the impression that the rod is rotating in a plane that is slanted either toward him or away from him, as is depicted here. The slants approximate those of projected circles.



BIZARRE THREE-DIMENSIONAL FIGURE seems to be traced by an imaginary rod that is created when two spots of light move at constant speed on the opposite sides of a rectangular path. The built-in tendency for the visual system to perceive the moving spots as being connected to each other and forming a rigid structure leads to the perception of a rod that is rotating around a stationary central point in a jerky manner, executing a strange three-dimensional motion the observer quite probably has never seen before.

Buy it for what it does. And what you don't.

The Victor Programable Calculator is not only a statistical genius, it's also an expert in the management of time.

Stored programs perform complex statistical calculations—such as mean and standard deviation, linear or multiple regression, probability, and the like—instantly and automatically. The drudgery, errors, and cost of repetitious manual sequences, such as those crossed out on the tape above, are eliminated. You and your people get more time to do what you do best: think.

Our extensive library of statistical programs can "educate" the machine to do all this, and more, at the touch of a button. The keyboard itself has eight unique keys which can themselves be custom-programed to perform any operation you select or design, from entry of a numerical constant to a complete automatic routine.

This desk-top, computer-brained, programable calculator is the most advanced of the 5-million office machines we've built over 56 years. And it's backed by the industry's largest service network, more than 800 locations nationwide.

To see how one fits into your program, just send the coupon.

Victor Comptometer Corporation
3900 No. Rockwell
Chicago, Illinois 60618 **SA675**

Yes, I would like to know more about statistical applications for the Victor Programable.

☐ Please send literature
☐ Please arrange demonstration

Name ______________________

Company ______________________

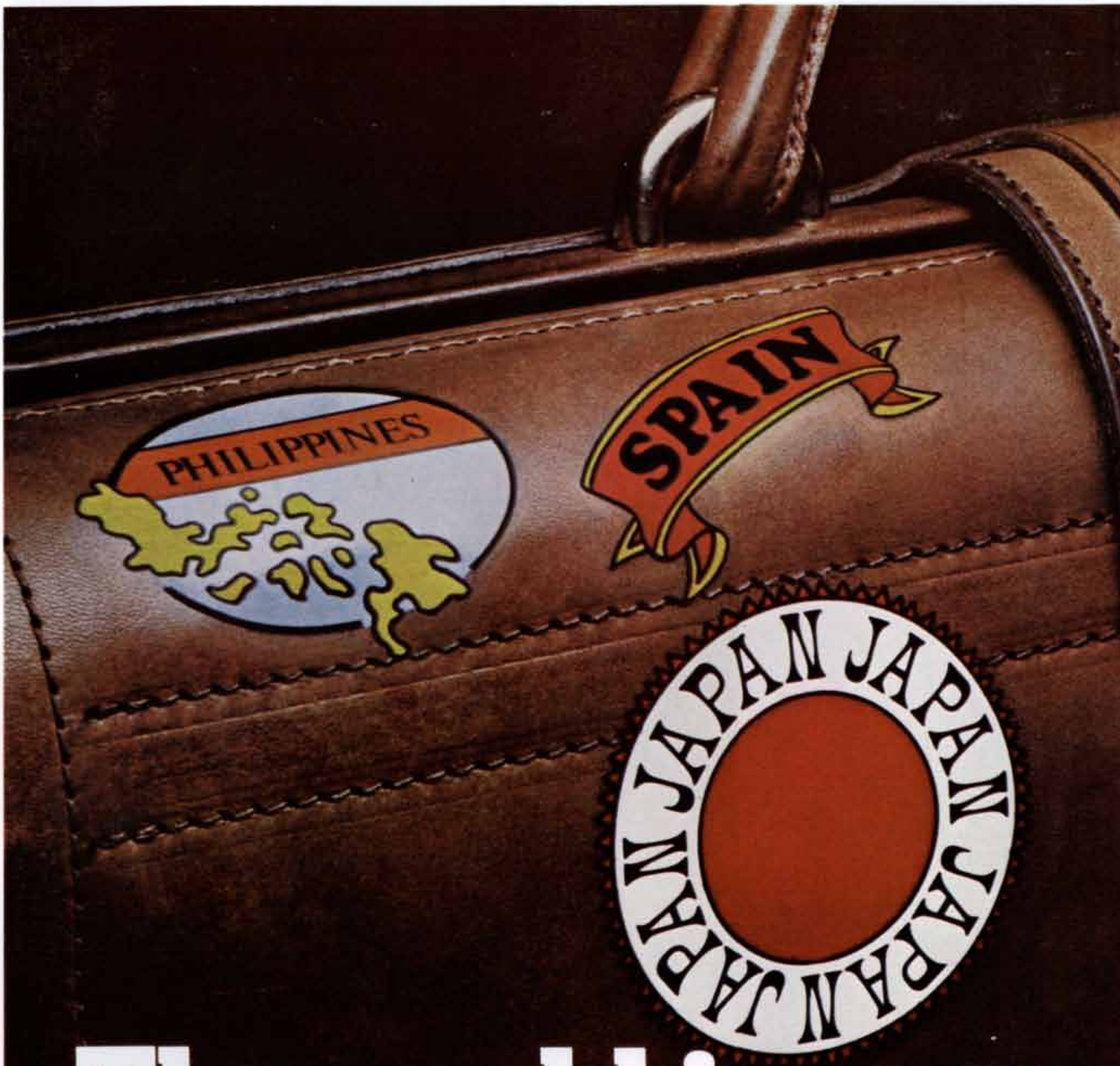
Address ______________________

City ______________ State ______________

Zip ______________ Phone ______________

THE VICTORS

Victor Comptometer Corporation, Chicago, Ill: 60618


PHILIPPINES
SPAIN
JAPAN JAPAN JAPAN JAPAN

to our physicians.

bonus payments. Your dollars go much further, because as an Air Force physician you may take advantage of money-saving purchases through the base exchange and commissary.

And as an Air Force officer, you enjoy an extra measure of prestige, privilege and responsibility.

If you qualify, come be the physician you want to be —gain the professional satisfaction you've trained for— in the Air Force. For the location of the Air Force Medical Recruiting organization serving your area, send in the coupon. Or, call toll free 800-447-4700 (in Illinois call 800-322-4400).

AIR FORCE

C-SA-65

Health Care Opportunities
P.O. Box AF
Peoria, IL 61614

I am interested in Air Force Health Care opportunities. I understand there is no obligation.

Name ____________ (Please Print) ____________ Sex ☐ M ☐ F
Address ____________ Phone ____________
City ____________ State ____________ Zip ____________
Soc. Sec. # ____________ Date of Birth ____________
Specialty ____________
School ____________ Year of Degree ____________

Delivering Computer Power.

Harris SLASH 4VMS Computer System

Access
to big
computers
can be a problem,
and minicomputers have limited power.
Harris medium-scale computer systems
put number crunching power
at the fingertips of the people who need it.
Economically.
Write: Harris Corporation,
55 Public Square, Cleveland, Ohio 44113.

HARRIS

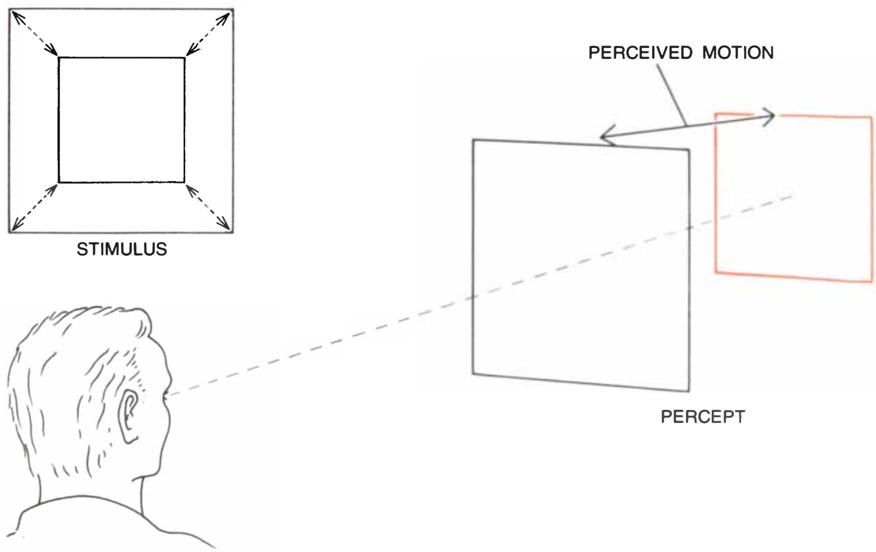
COMMUNICATIONS AND INFORMATION HANDLING

I quite clearly do not perceive them in this way as having a common frame of reference. Perceptually I experience the room as being static, the child as running across the floor, the child's hand and arm as moving relative to her body, the child's finger as moving relative to her hand and the ladybug as moving relative to the child's finger. Thus my visual system abstracts a hierarchical series of moving frames of reference and motions relative to each of them. The perceptual analysis of the optical flow as a hierarchical series of component motions follows closely the principles of ordinary mathematical vector analysis; hence it has been termed perceptual vector analysis. In our laboratory at the University of Uppsala we have devoted much experimental effort to a search for the basic principles underlying this perceptual function.

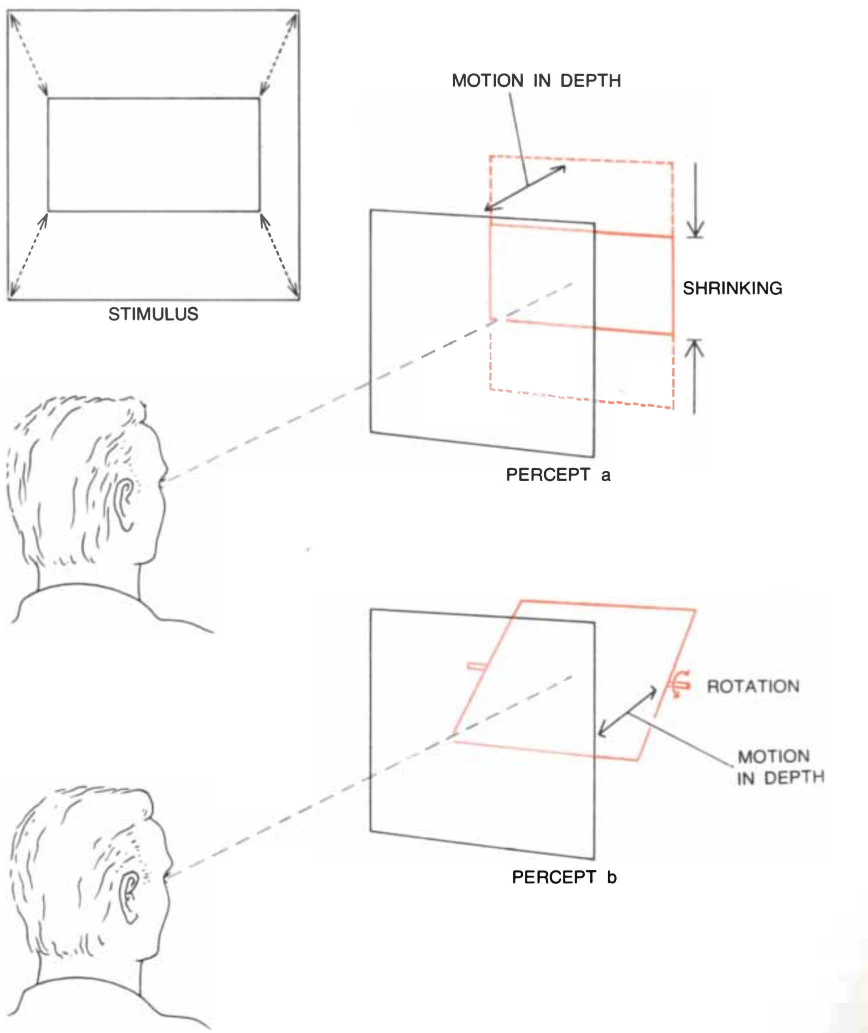
I shall now briefly describe some typical experiments in my laboratory involving perceptual vector analysis and its geometric basis. In most of the experiments the visual stimuli consist of computer-controlled patterns displayed on a televisionlike screen and projected into the eyes of our subjects by means of a collimating device that removes parallax as well as the possibility of seeing the screen.

Some of the fundamentals in the general principle of perceptual relativity are demonstrated in one of my earliest experiments. The stimulus pattern consists simply of three bright spots, *A*, *B*, *C*, one above the other, moving back and forth along straight paths [*see lower illustration on page 79*]. When the top and bottom spots, *A* and *C*, are displayed alone, moving horizontally to the left and to the right, they seem to be rigidly connected. When the middle spot, *B*, is presented alone, it is "correctly" seen as moving in a sloping path. When the three elements are presented simultaneously, however, we get an example of perceptual vector analysis. The entire unit *ABC* seems to be moving horizontally as a unit, but the path of *B* does not appear to be sloping; instead *B* seems to be moving vertically up and down in a straight line. This result can be generalized: Equal vectors or vector components form a perceptual unit that acts as a moving frame of reference in relation to which secondary components seem to move.

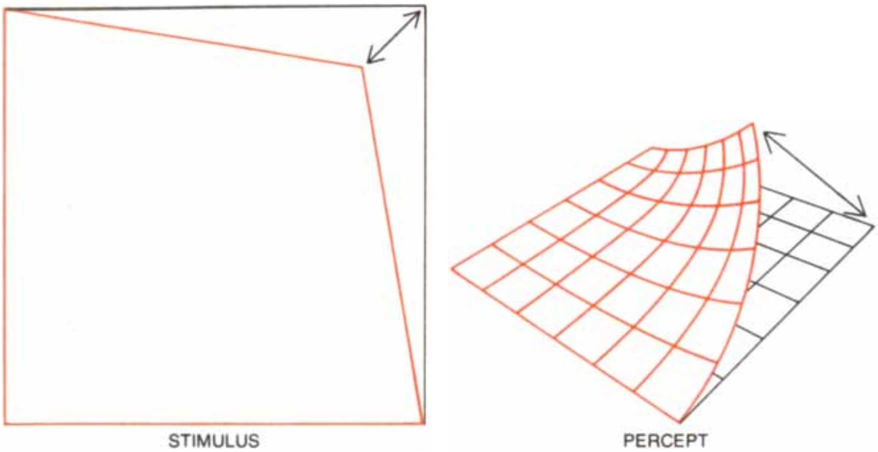
A more recent series of experiments in which a few points trace an ellipse or some other conic section provides other striking insights into the geometry of perception. If we present on our display screen two spots opposite an imaginary



ADVANCING AND RETREATING SQUARE is perceived when the stimulus consists of a square that simply contracts and expands. The visual system interprets change in size as a perspective change produced by a figure of constant size moving back and forth in depth.



ALTERNATIVE VISUAL PERCEPTIONS are evoked if the stimulus square not only contracts and expands but also is simultaneously transformed into a rectangle as it is shrinking. One group of observers sees a figure receding and advancing while it is changing simultaneously from a square into a rectangle and back again (*percept a*). To other observers the square remains a square but one that is oscillating on its horizontal axis (*percept b*).



BENDING MOTION is perceived if one corner of a square figure is moved along a diagonal path. The perception of bending may continue until the moving corner actually touches the opposite corner. A given observer will initially perceive bending as being either toward or away from him, but with some effort he can reverse apparent direction of motion.

center point tracing an ellipse, observers always seem to see a rigid rod of which only the end points are visible [*see upper illustration on page 80*]. Even more surprising, the rod is seen as rotating in a plane that is tilted away from (or toward) the observer. The perceived plane has a slant corresponding roughly to the computed slant of a projected circle. Even though the observer is fully aware that the points on the screen are really tracing an ellipse, he is unable to see the "true" Euclidean pattern; he always sees the ellipse as a circle in perspective. Thus we meet with a convincing indication that the perceptual analysis spontaneously follows the principles of central perspective.

A still more fascinating "illusion" is created by a variation of the experiment in which the two spots of light follow a perfectly rectangular path [*see lower illustration on page 80*]. I must admit I was surprised to find that even in this case the two spots appear to be the lighted ends of a rigid rod rotating around a fixed central point. One might expect that one would simply see two spots (perhaps elastically connected) chasing each other around a rectangular track. Instead an imaginary rod is again seen; its length seems to be constant as the rod describes a curious path in which it rotates for part of the time in a nearly vertical plane and then slants rapidly away from the vertical and back again. So strong is the perceptual tendency toward abstract projective invariance that a highly complex and "unnatural" motion—one that may not have been seen before—is preferred to the simple rectangular track traced by two moving spots. Evidently it is obligatory that the spatial relation between two isolated moving stimuli be perceived as the simplest motion that preserves a rigid connection between the stimuli. The general formula is spatial invariance plus motion.

In a related but slightly different class of experiments the display screen presents the full outline of a simple geometric figure whose shape is systematically altered in a particular way. For example, the observer may be shown a square alternately contracting and expanding [*see top illustration on preceding page*]. What the observer perceives, however, is a square of fixed size alternately receding and approaching. He never perceives the square as a stationary pattern that is changing in size. The result again means that the visual system automatically prefers invariance of figure size, obtained by inferring motion in three-dimensional space.

The next experiment I shall describe is perceived two different ways by different observers. Some observers seem to see it only one way whereas for others the two types of percept alternate. In this presentation the top and bottom of a square alternately shrink and expand as in the preceding experiment while the sides of the square move in and out a smaller distance. Geometrically a large square collapses to form a somewhat smaller rectangle, then expands to its original shape [*see bottom illustration on preceding page*]. All observers have the impression that the figure is alternately advancing and retreating. For one group of observers, however, the figure seems to change during its translatory motion from a square into a rectangle and back again. For a second group of observers the square seems to remain a square at all times, but a square that is rocking back and forth around its horizontal axis as it advances and retreats. Thus we encounter two variants of a vector analysis in the geometric framework of central projection. The first variant is particularly interesting because it represents perception of simultaneous motion and change of shape, such as one might see in a moving cloud or a ring of cigarette smoke.

A final example, taken from a set of experiments that Gunnar Jansson and I have recently published, involves a rather subtle change in the geometry of a square: one corner is made to move slowly toward the center of the square and back again [*see illustration at left*]. The result is interesting because it demonstrates a new type of perceptual invariance. The observer has the illusion that the square is a flexible surface with a corner that is bending toward him. This may seem surprising, but it is just what one would expect if the figural change is interpreted as being a continuous perspective projection.

It is a common characteristic of all the experiments I have described that the observer is evidently not free to choose between a Euclidean interpretation of the changing geometry of the figure in the display and a projective interpretation. For example, he cannot persuade himself that what he sees is simply a square growing larger and smaller in the same visual plane; his visual system insists on telling him that he is seeing a square of constant size approaching and receding. Hence he perceives rigid motion in depth, rotation in a specific slant, bending in depth and so on, paired with the highest possible degree of object constancy.

The theory of visual perception I have outlined here is based on studies with artificial and highly simplified stimulus patterns. Such experiments helped to demonstrate that the visual system uses the geometry of central perspective and enabled us to formulate the principles of perceptual vector analysis. It was natural for my colleagues and me to ask ourselves: Is there any way to show experimentally that the principles of perceptual analysis also hold true for the more complex patterns of motions encountered in everyday life? In an attempt to answer this question we began some years ago to study experimentally the complex patterns of motion generated by men and animals, patterns that might be called biological motions.

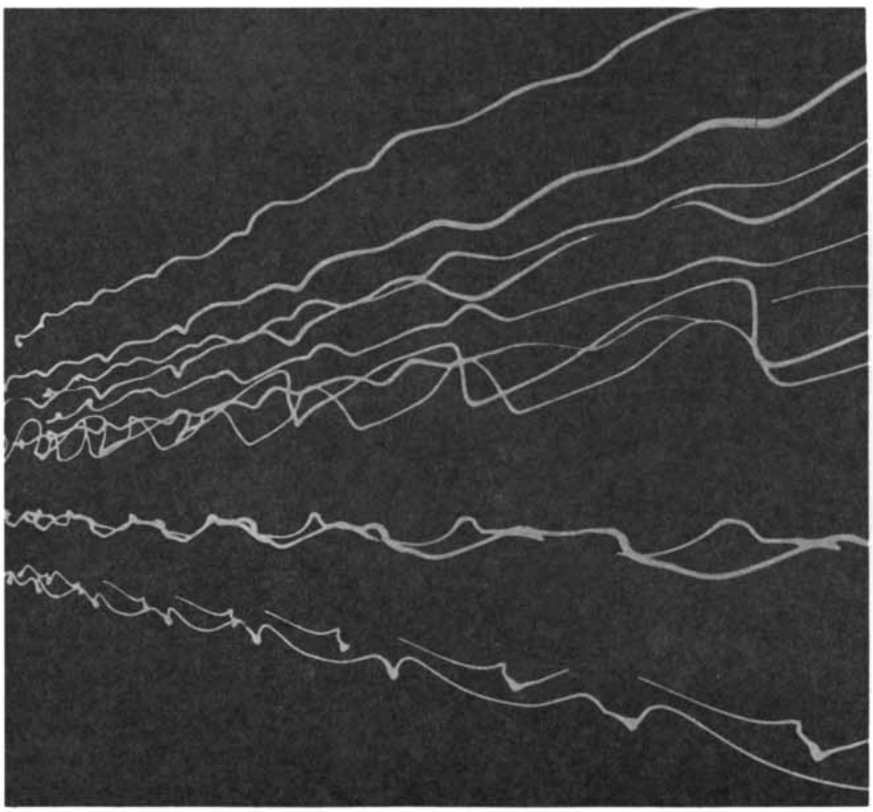
Consider, for example, all the intricate coordinations of frequencies, phase relations, amplitudes and acceleration patterns that are accomplished by one's skeletal structure when one merely walks across the floor. Even in such a simple act scores of articulated bones make precise rotations around dozens of joints.

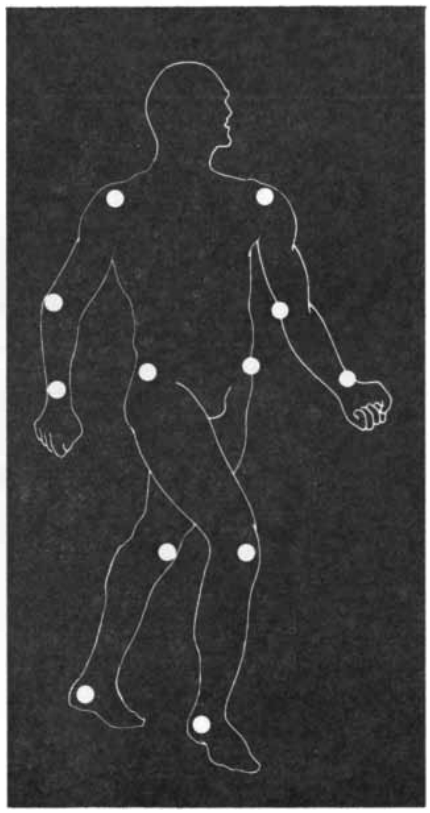
Our simple early experiments had demonstrated that the moving end points of an otherwise invisible straight line carry enough information to convey the impression of a rigid line moving in three-dimensional space. We therefore hypothesized that if we presented the motions of the joints of a walking person in the form of a number of bright spots of light moving against a dark background, an observer might perceive that the spots represented someone walking. We attached small flashlight bulbs to the shoulders, elbows, wrists, hips, knees and ankles of one of our co-workers and made a motion-picture film of him as he moved around in a darkened room [*see illustration below*].

The results, when the motion picture was shown to naïve observers, exceeded our expectations. During the opening scene, when the actor is sitting motionless in a chair, the observers are mystified because they see only a random collection of lights, not unlike a constellation. As soon as the actor rises and starts moving, however, the observers instantly perceive that the lights are attached to an otherwise invisible human being. They are able not only to differentiate without hesitation between walking and jogging movements but also to recognize small anomalies in the actor's behavior, such as the simulation of a slight limp. In another experiment we filmed two people, similarly festooned with lights, performing a lively folk dance. When the film is projected, anyone can see immediately that the 24 swirling spots of light represent a dancing couple [*see illustration on page 77*].

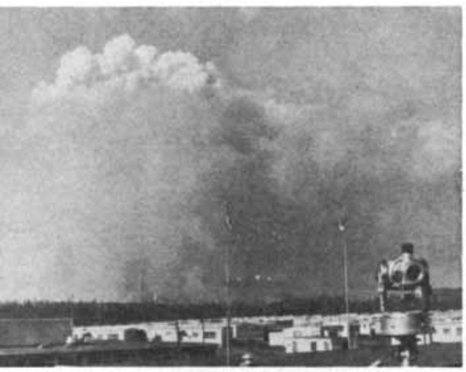
The surprising ability of the human visual system to perceive a dozen or two dozen moving lights as the motions of people led us to study the minimum exposure time required for the sensory organization of such patterns. The result, recently published by our group, is that a tenth of a second (the time needed to project two motion-picture frames) is often enough to enable a naïve observer to identify a familiar biological motion. This finding, together with results not yet published, has led me to believe that the ability of the visual system to abstract invariant relations from the kind of patterns I have been describing is the product of "hard-wired," or fixed, visual pathways originating at the retina and terminating in the cortex. It is as if the hierarchies of relative invariances in the optical flow were filtered out and established before the visual signals reach the level of consciousness. And contrary to expectation the more complex a projectively coherent pattern is from the mathematical point of view, the more effective the sensory decoding is. (Witness the decipherment of the dancing lights.) Evidently as the degrees of freedom are reduced the stimuli become rich in redundant information.

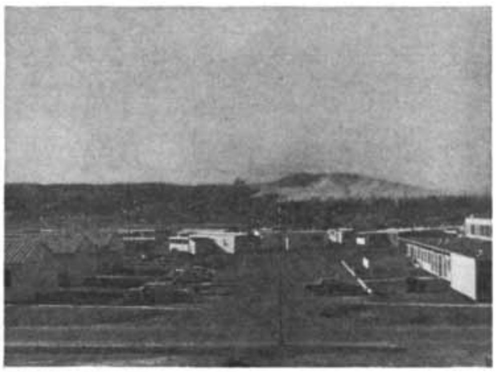
From our investigations we now know that the component in the optical flow that is a consequence of locomotion generally represents a continuous perspective transformation. Generalizing further from our experiments, we conclude that

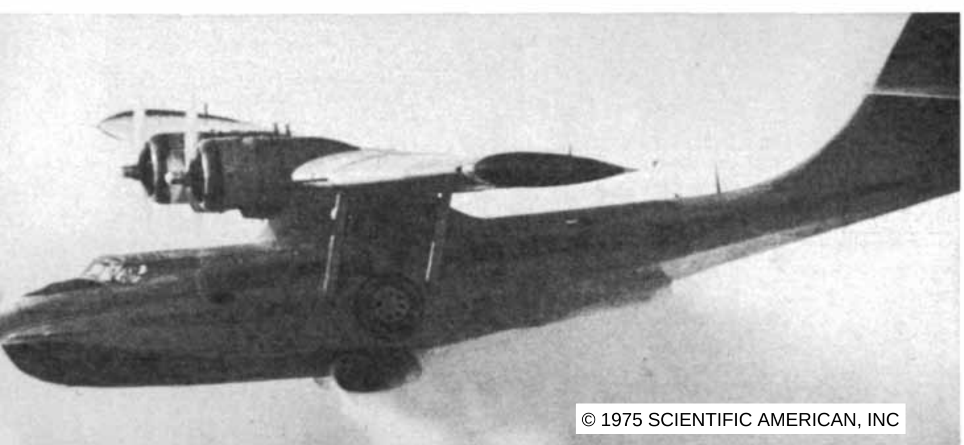




LIGHT TRACKS OF WALKING PERSON (*left*) **are recorded by making a time exposure in a dark room of a subject fitted with 12 small lights at his principal joints, as is shown at the right. The continuous streaks generated in this way have no obvious interpretation. If, however, the moving-light patterns are recorded on motion-picture film, one can see instantly when the film is projected that it portrays a person walking. Motion-picture frames of two similarly lighted subjects dancing in the dark appear on page 77.**







during locomotion the components of the human visual environment are interpreted as rigid structures in relative motion. In this regard the theory and our experiences are in good correspondence; there can be no doubt that we perceive the environment as being rigid.

The term relative motion can imply, however, that either the perceiver or the environment (or both) can be regarded as moving relative to the other. Both experiments and experience indicate that the environment forms the frame of reference for human locomotion. The world is perceived as being stationary and the observer as being in motion. From the point of view of theory we may nonetheless ask: Why is the eye itself not the ultimate reference? Why does one not perceive the ground to be moving instead of oneself? From the point of view of function, the answer is easy: The perceptions supplied by a "stationary" eye would be less informative. But let us ignore function, since we are considering the principles of decoding.

We recognize, of course, that visual information about locomotion does not stand alone; it interacts with signals from other sense organs that report bodily movements: organs in the joints, in the muscles, in the inner ear and so on. It is evident, however, that our consciousness of locomotion requires something more. Experiments have shown that the visual perception of locomotion is able to override conflicting spatial information from those other sensory channels. Thus it seems that the optical flow that covers the retina during locomotion takes precedence over all other sensory information.

The work I have reported, together with comparable studies from many other laboratories, provides the outlines of what one might characterize as a relativistic theory of vision. The central finding is that the geometry of the decoding of visual stimuli is a relational one similar to projective geometry. In accordance with this geometry, series of relative invariances, or perspective transformations, are abstracted from the total optical flow. This results in hierarchical systems of different components that are perceived both in common and in relative perspective transformations. As our experiments make clear, human beings tend to perceive objects as possessing constant Euclidean shapes in rigid motion in a three-dimensional world. In real life these principles of visual analysis taken together give rise to a satisfactorily close correspondence between the physical world and what we perceive that world to be.

PELAGIC TAR

Fine nets towed in the open ocean pick up ubiquitous black lumps, apparently the residue of sludge washed out of the hold of tankers. The stuff could probably be eliminated by good tanker procedures

by James N. Butler

Until a few years ago the term pelagic tar and the material to which it is applied were virtually unknown. The term describes tarry residues of petroleum that are found on the surface of the ocean. (They are also found on ocean beaches, but a precise definition of "pelagic," which comes from the Greek for sea, limits its application to the open sea.) The tar appears to originate mainly from the discharge of waste by oil tankers. Tar residues of this type probably could have been found long ago, but the recent rapid increase in the transportation of oil by tanker and the shift of tanker routes as a result of the closing of the Suez Canal have made the tar considerably more conspicuous.

One naturally asks what becomes of the pelagic tar and what effects it has on marine organisms. My colleagues and I at Harvard University and the Bermuda Biological Station for Research have been investigating pelagic tar in an attempt to answer these questions and others. We do not have complete answers to any of our questions, but a good deal of information about pelagic tar is now in hand.

A factor contributing to the recognition that pelagic tar had become widespread in the world's oceans was the development of the neuston net [*see illustration on page 93*]. It was designed in 1962 by Peter M. David of the National Institute of Oceanography in Britain. The net is towed beside a vessel to sample the population of neuston, which are tiny organisms that float in the film of water at the surface of the ocean. It has been widely employed by oceanographers, many of whom have found that their nets are scooping up significant amounts of tar as well as neuston.

David seems to be the first person to have reported tar lumps at sea. In a communication to me he wrote: "The first tar lumps I caught were taken in hand nets at about 40°N, 20°W [between Portugal and the Azores] in 1954. Many of them were colonized by hydroids and barnacles....

"I began to make neuston nets in 1962 and used them in the International Indian Ocean Expedition in 1963 and 1964. Tar lumps were caught in many, possibly most, of the Indian Ocean samples, although often very tiny fragments in small quantities. The Mediterranean samples taken on the way out to the Indian Ocean were heavily contaminated. I remember thinking that the area off Southeast Arabia where we were going to do our first survey would probably be thick with oil, being right on the main tanker route, but it turned out to be very free of contamination."

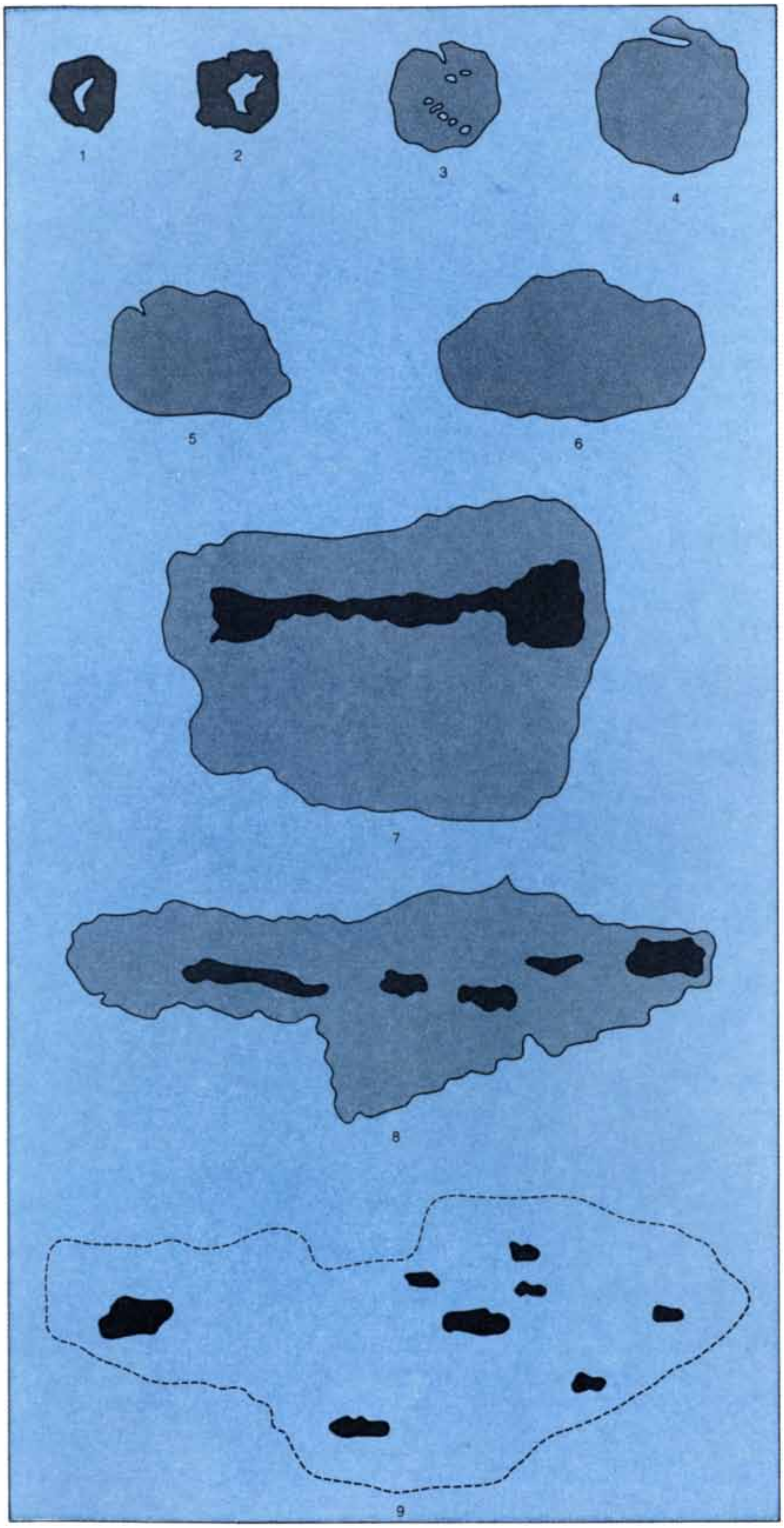
In December, 1968, the research ship *Chain* encountered such a heavy concentration of oil and tar residues on the North American side of the Atlantic that investigators aboard reported the matter to the Smithsonian Institution's Center for Short-Lived Phenomena. (As it has turned out, pelagic tar is a short-lived phenomenon only on a geologic time scale.) The ship was on a cruise from the Woods Hole Oceanographic Institution south through the western Sargasso Sea. Victor E. Noshkin, Jr., and James Craddock, who made many tows with neuston nets, reported on one episode as follows. "In a 30-minute tow the nets skimmed about 1,800 meters of sea surface and contained a cupful of oil-tar lumps.... After from two to four hours of towing the mesh became so encrusted with oil that it was necessary to clean the nets with a strong solvent. On the evening of December 5, at 25°40'N, 67°30'W, the nets were so fouled with oil and tar material that towing had to be discontinued."

In the Mediterranean Sea three other workers from Woods Hole (Michael H. Horn, John M. Teal and Richard H. Backus) found large amounts of tar during a cruise of *Atlantis II* in 1969. Reporting in *Science* that they had collected tarry lumps in 75 percent of more than 700 tows with neuston nets, they wrote that the lumps were irregular in shape, with the greatest dimension varying from one or two millimeters to about 10 centimeters. "Hardness varied," they wrote, "although all lumps were easily deformed by a touch of the finger.... Some of the lumps were very sticky, had a rough, uneven surface and were relatively soft and black. Other lumps were firmer, with a smoother, more even surface, and were usually lighter (brownish black); this type frequently had barnacles attached and appeared to be older than the first."

Several organisms were associated with the tar lumps collected by the Woods Hole group. An isopod, *Idotea metallica,* was collected in large numbers. When these organisms were left with a tar lump in a tank on the ship, they tended to remain attached to the lump. The goose barnacle (*Lepas pectinata*) was frequently found attached to the lumps, particularly to the firmer, older-looking ones. At one station 150 barnacles ranging from two to eight millimeters in length were attached to four lumps. The growth rate of the barnacles was measured and found to be approximately one millimeter per week, thus establishing that the tar lumps to which they were attached were at least two months old. Tar was also found in the stomach of the saury (*Scomberesox saurus*), a fish that feeds on surface-dwelling crustaceans, in three of 10

LARGE TAR LUMP, somewhat magnified here, appears in a clump of the alga *Sargassum*, for which the Sargasso Sea of the western North Atlantic Ocean is named. Many organisms are associated with the alga. They are being studied to see if they are affected by tar.

LUMP OF TAR found in the ocean near Bermuda carried a goose barnacle, *Lepas pectinata*. Barnacles of this species are frequently found attached to pelagic tar, particularly the older and firmer lumps. By ascertaining the growth rate of the barnacles one can estimate the minimum time the tar has been in the water. It is not known whether or not the barnacles are affected by pelagic tar.



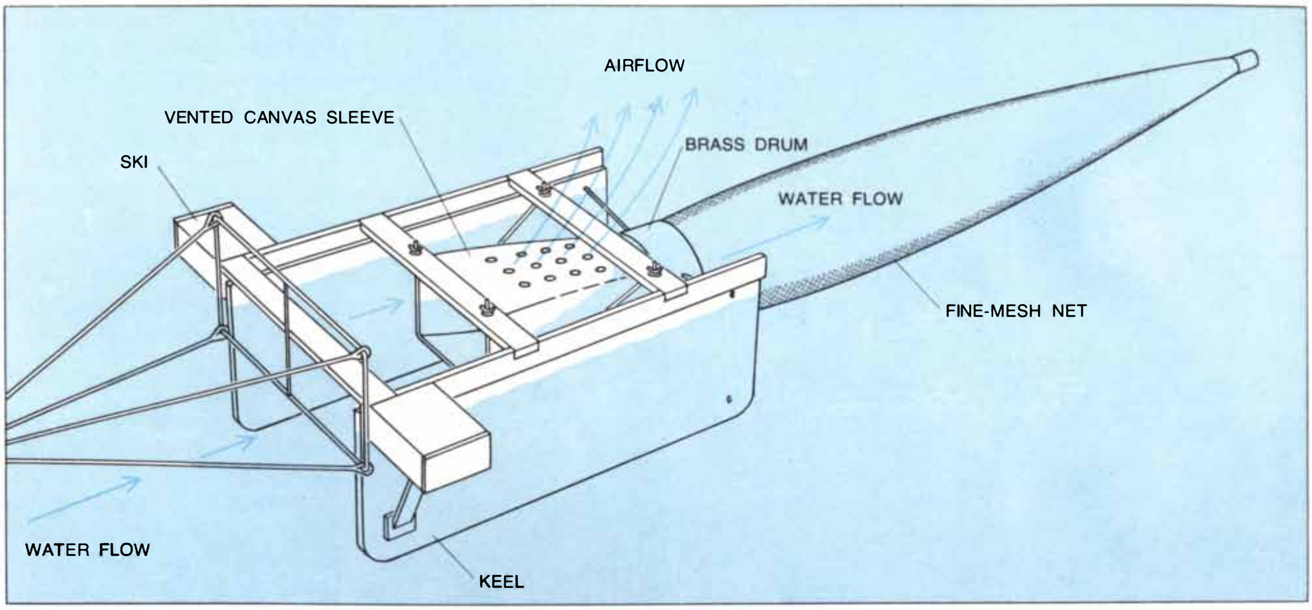
EVOLUTION OF OIL SLICK was traced in an experiment. The oil was dumped in the water by a ship that moved in a circle. Initially (*1*) the slick is about .2 kilometer in diameter, with an open area in the center. Succeeding panels (*2–6*) show it at four minutes, 22 minutes, 42 minutes, two hours and seven hours. After two days (*7*) it consists of a dark patch of thick oil surrounded by a much larger area of thin oil. Half a day later (*8*) the thick patch is breaking up. After four days (*9*) slick is 2.4 kilometers wide but indistinct.

specimens collected by the expedition. Several lumps covered with a grayish film were found to take up oxygen because of the microorganisms on their surface; when the lumps were treated with formalin to kill the microorganisms, the uptake of oxygen dropped to essentially zero.

Observations of tar by Thor Heyerdahl and his associates on their expeditions across the Atlantic on the papyrus rafts *Ra I* and *Ra II* in 1969 and 1970 attracted much more notice than the other reports because the raft voyages were widely publicized. Since the rafts traveled slowly (at a speed of about two knots) and the crew members rode only a foot above the surface of the water, the tar was much more noticeable than it could be to observers on larger ships, who would be from 15 to 40 feet above the water and moving at speeds of from eight to 12 knots. The log of *Ra II* records tar sightings on 40 of the first 43 days of sailing; the rest of the voyage (56 days) was in relatively clear water. The route of both voyages was from North Africa to Barbados, following the natural surface current that forms the southern boundary of the Sargasso Sea.

At approximately the same time Byron Morris, then a graduate student in biological oceanography at Dalhousie University and now my co-worker at the Bermuda Biological Station, was investigating surface-dwelling plankton with a fine-mesh net similar to the one designed by David. "I discovered on my very first tow with the net that plankton was not all I would catch," he told me. "In the bag with the animals were numerous pebble-sized lumps of black tar. I should have expected this, since I had recently heard that workers from Woods Hole had found tar... in the North Atlantic. ... Nevertheless, my first encounter with floating tar came as a sad surprise." Since then Morris has made hundreds of tows in the western North Atlantic and the Sargasso Sea. He has reported that "on only one occasion did I fail to catch any tar."

One of the tar lumps collected by David was analyzed in the gas chromatograph by J. V. Brunnock, D. F. Duckworth and G. G. Stephens of the British Petroleum Company Research Centre in 1968. Their chromatogram showed a large component of paraffinic waxes that have a high boiling point. This finding indicated to the investigators that the lumps were not simply evaporated crude oil but were the sludge of crude oil, which clings to the inner walls of tankers and must be removed with the ballast before the next cargo is loaded. Some of



NEUSTON NET was designed to sample the population of neuston, which are small organisms living at the surface of the ocean. The nets also have collected significant amounts of tar, which had gone largely unnoticed until the nets came into wide use about a decade ago. The net and its supporting framework are designed to be towed alongside a boat from a boom or a crane near the bow.

the tar found in the Mediterranean by the Woods Hole group was analyzed at Woods Hole by Max Blumer, whose gas chromatogram showed it to be a residue of crude oil. His finding was in agreement with the one obtained by the British workers.

Tanker waste, consisting of the heavier and waxier fractions of crude oil, probably forms into tar lumps quickly on entering the water. It is also reasonable to suppose some lumps form more slowly as a result of accretion from oil slicks such as the ones resulting from tanker accidents. A slick would be likely to contain the lighter fractions of petroleum in addition to the heavier ones, but the lighter fractions would evaporate fairly rapidly, leaving the heavier fractions in the water as tar.

Since 1971 Morris and I have been studying the pelagic tar of the Sargasso Sea and the many organisms living within clumps of the alga *Sargassum,* for which the sea is named. We have had the benefit of data accumulated over a long period of time from the hydrographic station that was established 20 miles southeast of Bermuda by Henry M. Stommel of the Massachusetts Institute of Technology in 1954. The station, which is now called Station S, has been monitored continuously by a succession of people from the staff of the Bermuda Biological Station, using the research vessels *Panulirus* and *Panulirus II.*

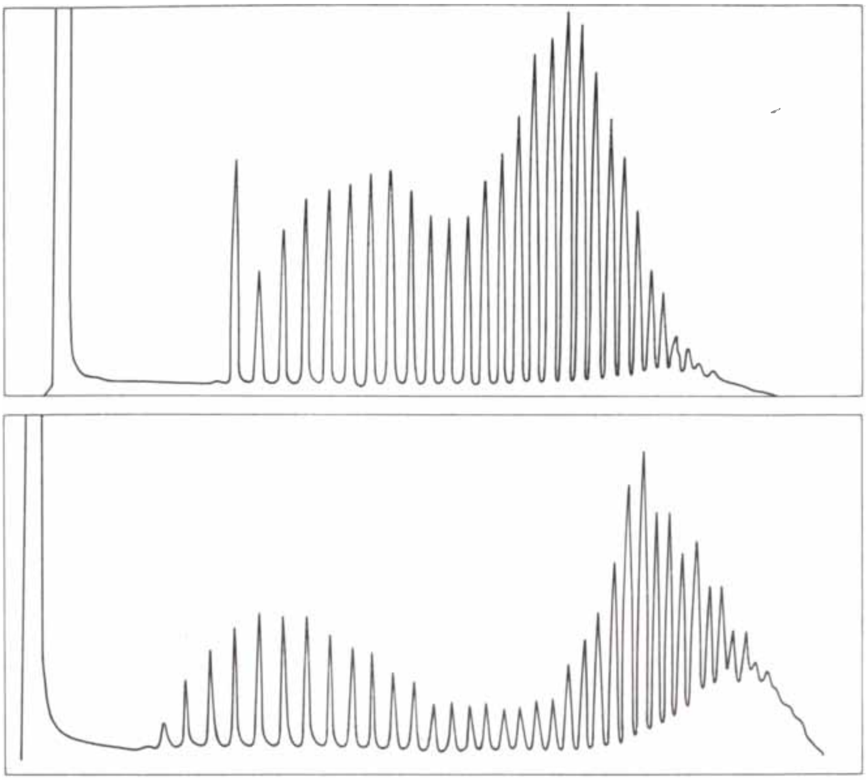
The first samples of pelagic tar to turn up at Station S were collected in 1970 by Roger Pocklington. Morris has carried out a regular series of neuston-net tows in the region since 1971. We have measured the wet weights of tar, *Sargassum* and surface zooplankton. Many of the tar lumps have been analyzed chemically by gas chromatography. Most of the communities found in the alga are being analyzed ecologically, that is, the animals (ranging in size from hydroids to crabs and fishes) are separated, identified and counted. This task can be quite tedious, particularly with the smaller hydroids, worms, copepods, isopods and amphipods.

On two occasions (once in February, 1972, and again six months later) we made several tows through a given area during a period of 24 hours. Surprisingly, the amount of tar collected per tow between 4:00 P.M. on February 15 and 7:00 A.M. the next day varied by a factor of almost 10, and the amount of *Sargassum* varied by a factor of almost 30. There was no correlation between the two. Six months later we found the average amount of tar was 10 times higher than it had been in February, but again the factor of 10 separating the highest and lowest amounts appeared and so did the lack of correlation with the amount of *Sargassum* collected. The ratio of *Sargassum* to tar ranged from a high of nearly 5,000 : 1 to a low of .2 : 1. The lack of correlation between the tar pattern and the *Sargassum* pattern offers us the possibility of studying relatively independent colonies of *Sargassum* with widely different amounts of tar and thus of ascertaining whether or not the presence of tar affects the ecology of the *Sargassum* communities.

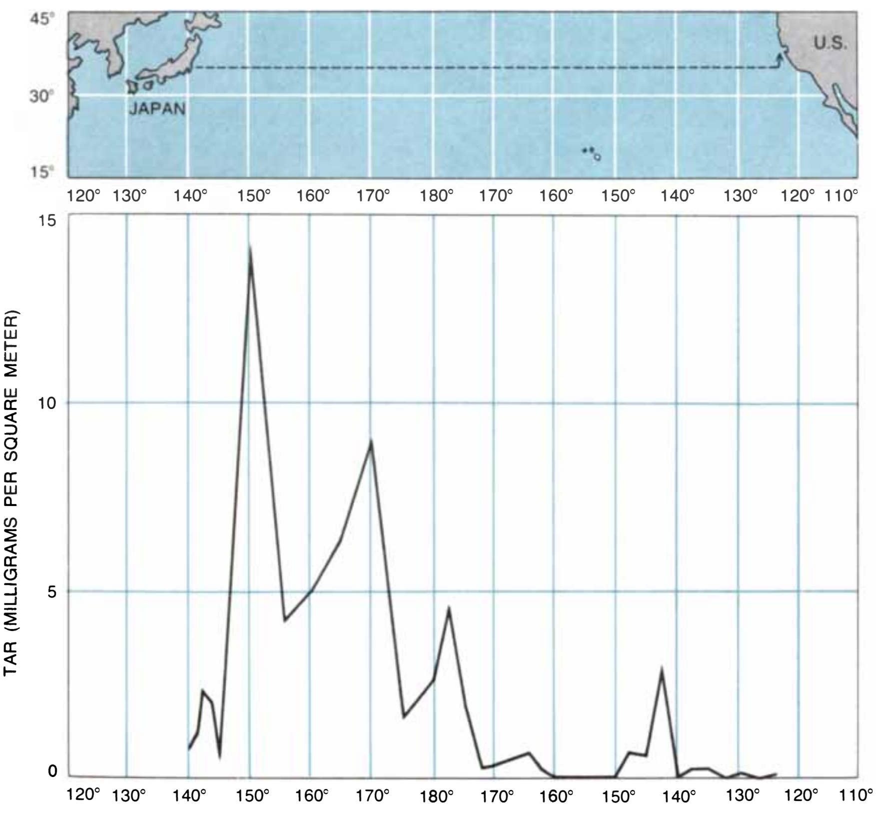
One fact established by these brief repeated tows is that a single measurement in one place at one time tells almost nothing about the amount of either tar or *Sargassum* that would be found in an adjacent area at the same time or in the same place at another time. Even on a time scale as short as one hour the amount of tar or *Sargassum* collected can vary by a factor of 10. One explanation for these variations is provided by visual observations of *Sargassum,* which show that if the wind is steady, the clumps of weed line up in rows parallel to the wind direction. The spacing of these windrows increases with the velocity of the wind from about 20 meters at five knots to about 50 meters at 25 knots.

Nearly 50 years ago the physicist Irving Langmuir demonstrated that the wind is only indirectly the cause of these windrows. He showed that the action of wind over open water produces counterrotating eddies and that between such eddies are bands of sinking water where floating weeds would collect. The Langmuir currents, as they are called, evidently affect accumulations of tar in the same way.

The windrows provide a possible explanation for the large amounts of tar observed by Heyerdahl. Since his rafts moved essentially downwind they could have spent long periods of time in windrows of tar lumps. In sampling at Station S the *Panulirus II* towed the net in a cir-



GAS CHROMATOGRAMS of crude-oil sludge (*top*) and pelagic tar (*bottom*) are compared. A chromatograph separates the molecular components of a sample. Here each peak represents a different paraffin molecule of hydrocarbon. The profiles are similar, indicating that the tar originated in crude-oil sludge, which is a waste discharged by oil tankers.



PATTERN OF TAR found on a crossing of the Pacific Ocean is displayed. The route of the ship is marked by the broken line. The largest amount of tar, comparable to the amount found in the Sargasso Sea, was collected in the Japan Current system, which is east of Japan.

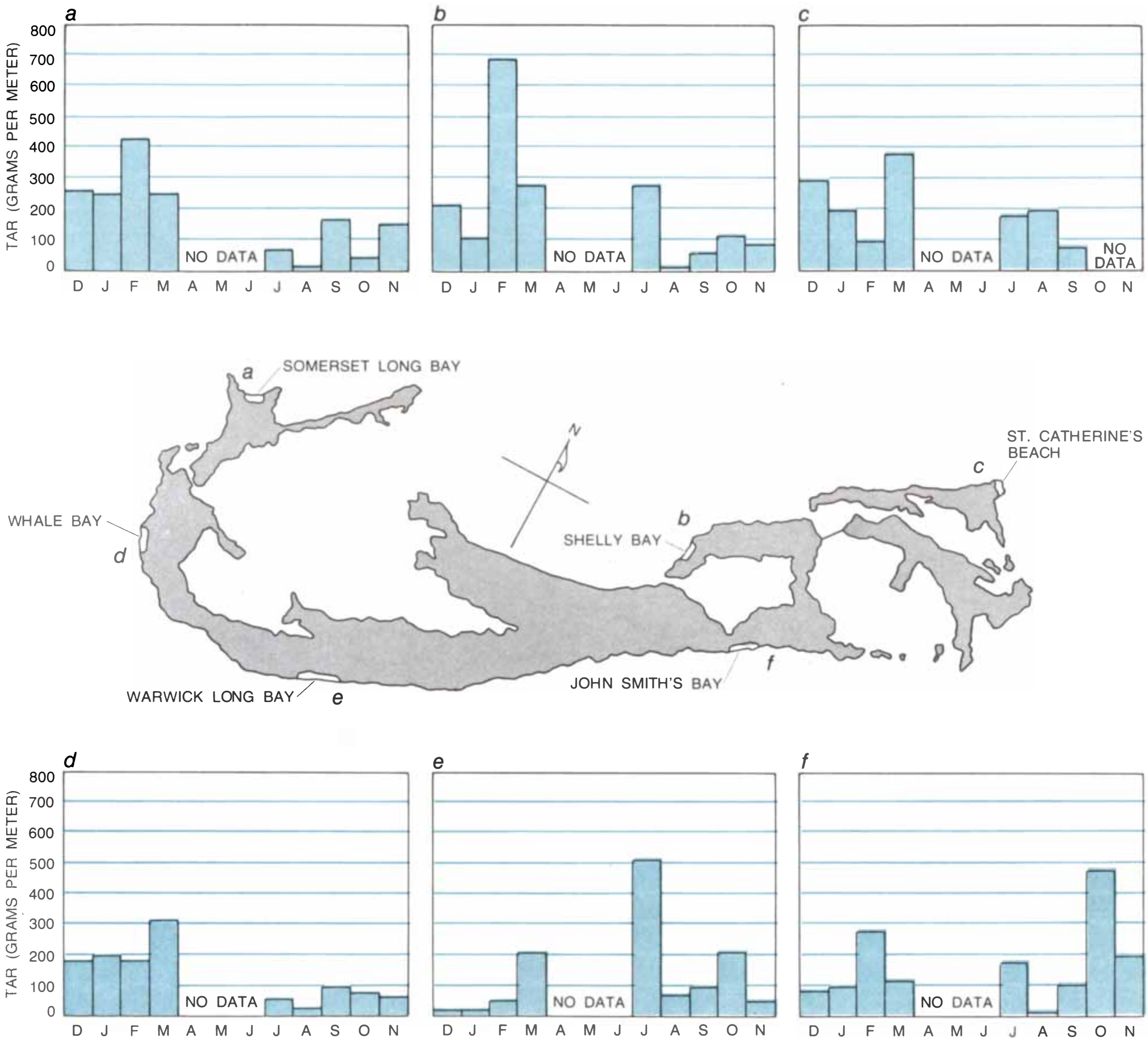
cle so that a single haul would contain material collected not only while the ship was going across the windrows but also while it was going parallel to them; thus this particular source of variation tended to be averaged out. Here one is dealing, however, with a variation that appears from hour to hour. The evidence obtained from two years of sampling at Station S indicates that the concentration of tar also varies over longer periods, possibly as a result of seasonal weather patterns.

The initial reports on pelagic tar were concerned mainly with the physical appearance of the lumps. We wanted also to characterize the lumps chemically. The analysis would not be complete, since crude oil contains hundreds of thousands of different chemical compounds. Instead we chose a simple method of gas chromatography that had been developed by Blumer. It gives a characteristic "fingerprint" of a sample, displaying in the chromatogram primarily the sharp peaks for the normal paraffins. Those peaks are superposed on a background of unresolved materials of greater complexity.

The chromatograph's column of silicone separates the components of a sample approximately in the order of their boiling points. Two important branched-chain compounds, pristane (19-carbon isoprenoid) and phytane (20-carbon isoprenoid), are also resolved. They are found in virtually all crude oils. Pristane is also found in some marine organisms, whereas phytane almost never appears except in organisms that have been contaminated by petroleum. The presence of phytane at about half the concentration of pristane therefore helps in distinguishing hydrocarbons originating in petroleum from hydrocarbons of biological origin. Olefins provide another means of making this distinction, although our analysis did not separate olefins from hydrocarbons. Crude oil contains essentially no olefins, whereas they are a common constituent of marine organisms.

Another distinctive feature of petroleum residues from tankers is the large group of chromatographic peaks that represent compounds with molecules that have from 30 to 40 carbon atoms. They are the paraffin waxes that precipitate from crude oil during shipping. About a third of our samples show this pattern clearly.

The absence of the components of low boiling point, such as are found in gasoline and kerosene, indicates that a sam-



TAR ON BEACHES was sampled by workers in Bermuda. Here the beaches where samples were taken are identified by letters, and the amount of tar collected on each beach is shown in the correspondingly lettered chart. The method of sampling involved collecting all the tar from a path one meter wide, extending from the surf zone of the beach to the high-tide mark. The charts show the weekly average of tar collected in each transect. The workers were volunteers assisting the Bermuda Biological Station for Research.

ple is weathered and therefore has been at sea for at least several days. Unfortunately the loss of volatile components provides a good time scale only in the early stages of weathering. The difference between a residue of tar a month old and one a year old is small, judging from chromatographic analyses of tar stranded on a rock in Bermuda.

We have now analyzed a large number of tar lumps chromatographically. Our experience has been that any two lumps caught in the same net may have distinctly different chemical properties or may be essentially the same. On a number of occasions lumps of almost identical composition were found at intervals of up to two months or at places separated by 100 kilometers. This evidence suggests that the tar comes from many different sources and that the sea tends to mix the lumps from a number of them.

The island of Bermuda itself tends to act as a giant net, collecting whatever is swept onto its beaches of pink sand by the intricate currents of the Sargasso Sea. Since November, 1971, several volunteers assisting our group have sampled six public beaches that face in different directions. The method has involved collecting all the tar lumps from a path one meter wide, extending from the surf zone to the high-tide mark on the beach. The size of the lumps turned out to be relatively large; most of them were between five millimeters and five centimeters in diameter.

On several occasions we marked lumps of tar with orange wax and replaced them on the beach. They subsequently were redistributed on the beach or washed out to sea within one cycle of the tide. In one experiment the lumps were strewn as far as 80 meters along the beach after one tide cycle. On another occasion they had all disappeared by the next day, washed out to sea.

Almost all the Bermuda residents I talked with were quite certain that the lumps of tar were brought in by the prevailing winds. Therefore we measured the wind speed and direction at the beach and also obtained more precise records from the weather station at the airport. Surprisingly, we found little or no correlation between the prevailing wind and the amount of tar discovered on the beach. Apparently the delivery of tar to the beach, like its distribution in the open ocean, is controlled by the small-scale currents of the Sargasso Sea rather than by the wind.

The beach tar is chemically indistinguishable from the lumps found on the open ocean. The occasional large lump obtained from the beach did, however, provide an opportunity for studying the internal structure of the lumps. Blumer and Jeremy Sass made the study at Woods Hole on samples collected in Bermuda by Pocklington.

One analysis involved a 155-gram lump that the investigators froze in dry ice. At the low resulting temperature the tar is quite brittle and can be cracked with a hammer and a spatula, giving clean fracture faces. The appearance of different parts of the lump varied from brown to black, from dull to shiny and from liquid to waxy to crusty in texture. This particular lump contained an inclusion of hard yellow wax in the form of a crystal about seven millimeters long and two millimeters wide.

On chromatographic analysis the crystal was shown to consist solely of paraffins of high carbon number. Analysis of three lumps of tar showed that the dull black crust contained relatively low amounts of paraffins and consisted mostly of unresolved components. The relative amount of paraffins was probably reduced because of bacterial degradation at the surface. The light brown waxy parts contained relatively more paraffins.

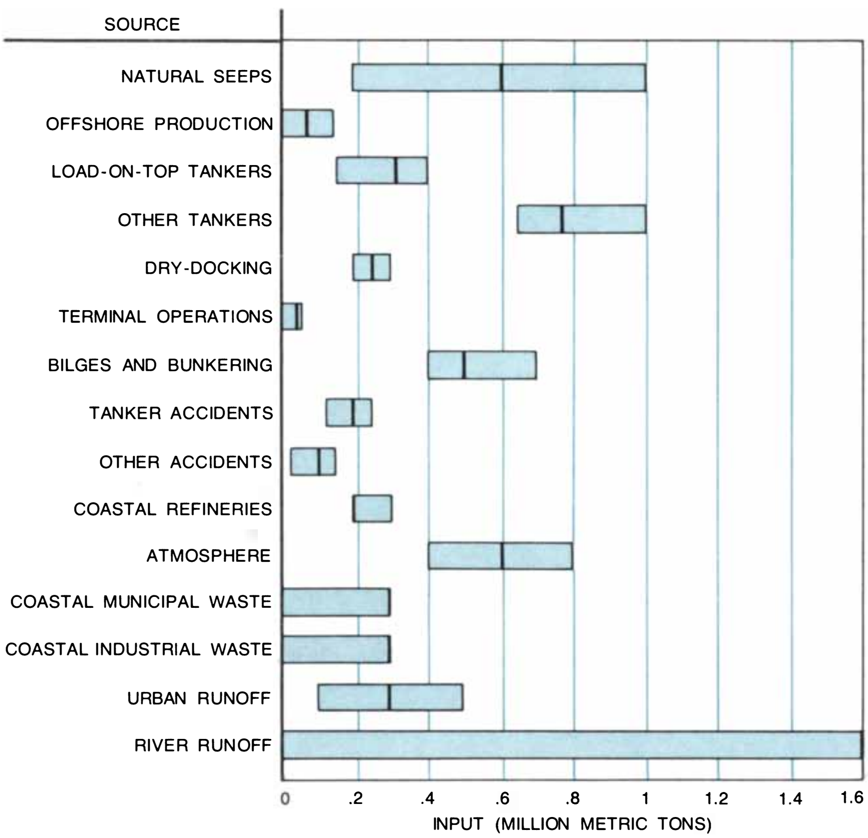
The waxy inclusions are strong evidence that the lumps originate in the material clinging to the inner walls of tankers after their cargo has been unloaded. This sludge and wax is dumped overboard into the sea by about 20 percent of the tanker fleet. The other tankers employ the "load on top" method, in which the residue of the cargo is separated from the ballast water in one compartment and combined with the new cargo instead of being dumped at sea. This procedure is reported to reduce the amount of oil lost with the ballast water to less than 10 percent of what it would be otherwise.

The amount of tar found at sea tells us that the lumps do not accumulate indefinitely. Rough estimates that have been made of the total source of petroleum residues that might form tar, and even rougher estimates of the total amount of tar in the oceans of the world, lead to a guess that the amount of tar now on the surface of the sea is about 700,000 tons. The amount is approximately equal to the estimated annual waste from tankers that do not use the load-on-top method and is about a fourth of all the petroleum wastes deposited in the open ocean.

Since many of the nontanker sources deposit oil containing large fractions of volatile or soluble compounds, their contribution to pelagic tar would not be as large as the contribution from waxy wastes deposited by tankers. A correlation between the estimated standing crop of tar and the estimated influx of petroleum for several oceans gives an average residence time of 2.4 months for petroleum of all kinds reaching the sea. When the figure is corrected for the approximate proportion (80 percent) of volatile and soluble compounds, which disperse in a few weeks, the residence time for tarry residue is estimated to be about a year.

Since the tar lumps do not grow indefinitely and do not seem to weather significantly, where do they go? One hypothesis is that the tar breaks up into fine particles (too fine to be caught by a neuston net) and that the particles are distributed by currents in the top few hundred meters of the ocean. When large amounts of seawater are filtered, black particles are sometimes found on the filter. Preliminary analysis indicates that they are similar to tar lumps in composition. When large amounts of seawater are extracted with an organic solvent, which removes both dissolved and particulate hydrocarbons, one finds concentrations of hydrocarbons near the surface that are small by analytical standards (and hence difficult to verify) but large enough to account for the residue of most of the petroleum ever put into the ocean. As yet no distinction can be made in these analyses between hydrocarbons produced by marine organisms and hydrocarbons resulting from pollution.

Ultimately the fate of petroleum residues at sea is not to remain as oil slicks or tar lumps; they are relatively transient states. The residues either remain in the water as dissolved or particulate hydrocarbons, or they sink to the sediment, or



INPUT OF PETROLEUM to the oceans is charted on an annual basis. For each source the black line represents the best estimate and the colored area indicates the probable range.

they are degraded to carbon dioxide by various organisms. In the laboratory, cultures of bacteria, yeasts and fungi can be developed that will metabolize hydrocarbons rapidly in the presence of enough nutrient. Indeed, the growth of this type of culture in the ballast water of a tanker has been tried experimentally to emulsify and disperse the waxy waste that apparently is the source of tar lumps. In the open sea, however, the population of microorganisms and the concentration of nutrients and hydrocarbons are much lower than they are in typical laboratory experiments, so that the rate of degradation is probably also much lower.

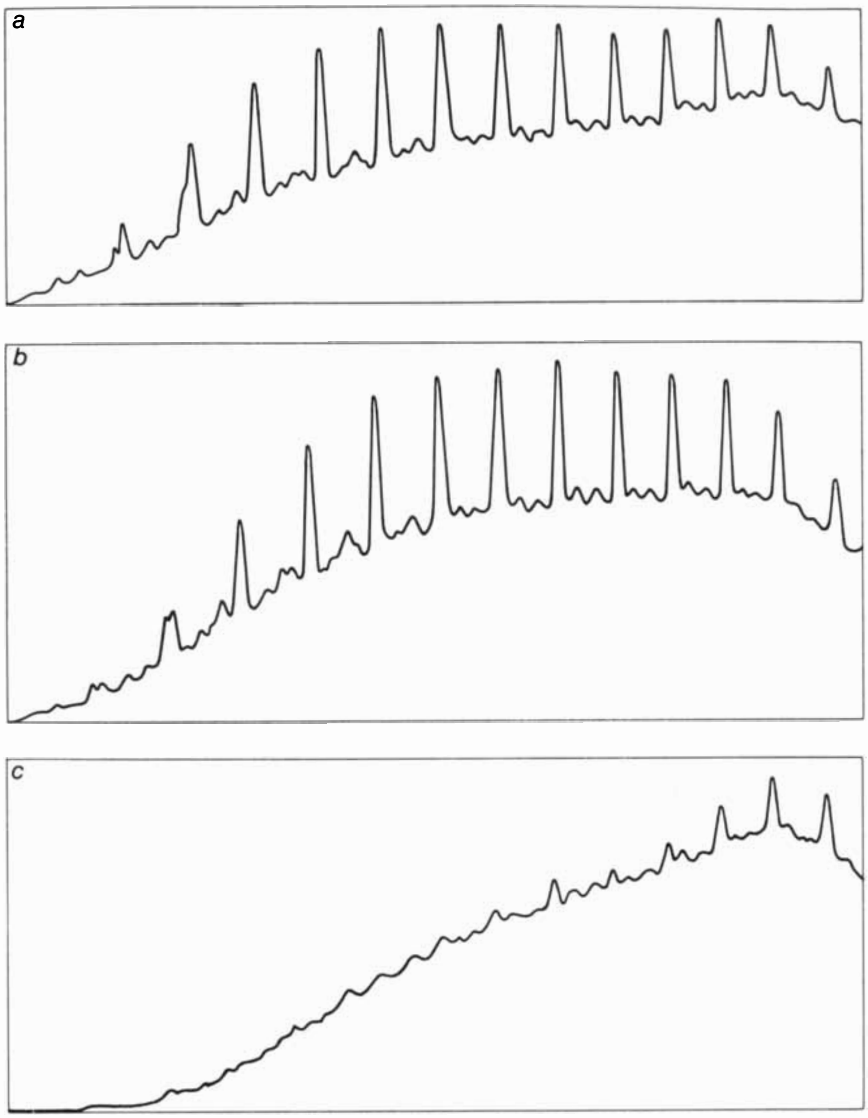
An unresolved matter of major concern is the effect of petroleum wastes on marine organisms. A few large spills of oil near coastlines and a few areas where petroleum wastes are chronically discharged have been studied, but even in these studies there is controversy about how much one may generalize to other situations. The interplay between natural processes of degradation, recolonization by organisms from nearby areas and changes in the kinds of organisms that make up a marine community are all under investigation.

In the open sea, as I have indicated, some organisms find tar lumps suitable as a solid substrate to grow on. Other organisms can take hydrocarbons into their system without changing them or appearing to be affected by them. The average concentration of hydrocarbons in open water is too low for any immediate toxic effects to be observed, although near oil slicks and in association with tar lumps the concentration may be much higher. What is not known is the long-term, low-level effect of compounds from petroleum.

Many marine organisms, ranging in size from bacteria to lobsters, appear to communicate by means of chemical signals. The compounds involved in such communications are in many cases similar to the compounds in petroleum. The communication is effected at concentrations so low that even the relatively small average concentration of hydrocarbons in seawater may be high enough to cause "static" that confuses the chemical senses of some organisms. The possibility of subtle effects of this kind calls for detailed and sophisticated research.

Other investigators are studying the nature of chemical signals between marine organisms in an effort to ascertain what compounds are critical for these communications. We hope to learn from our studies of *Sargassum* communities exposed to varying amounts of tar just how much effect such a pollutant has on the ecological structure of the community. Still other workers are trying to make a clearer distinction between hydrocarbons produced by marine organisms and hydrocarbons introduced by oil pollution. Eventually all these factors may be brought together to form a realistic and quantitative picture of the effect of oil pollution on life in the oceans.

Concern about irreversible damage to the life in the oceans as a result of pollution is widespread among marine scientists, but the question of how much effect petroleum discharges at present levels are having is still unanswered. It may be possible for the oceans to accommodate much higher levels of petroleum waste without significant changes. On the other hand, the critical damage to subtle chemical communication signals between organisms may have already begun; if it has, the measurements that can be made now are simply too crude to detect it. In a situation of such importance for the future, where so little is known, it seems wise to err on the side of caution and attempt to control the dumping of petroleum wastes in the ocean with whatever international sanctions are available.



WEATHERING OF TAR is portrayed in chromatograms of a sample stranded on a rocky shore in Bermuda. The tar came ashore in December, 1970. The chromatograms were made in April, 1971 (*a*), October, 1971 (*b*), and February, 1972 (*c*). The tar was not significantly degraded in that time except for the outermost layer (*c*). Inside the softer material was almost as fresh as when it was deposited. Since tar lumps on the open ocean do not endure indefinitely, they must be disappearing by some mechanism other than simple weathering.

The Role of Music in Galileo's Experiments

A hitherto unpublished page of Galileo's working notes, preserved in Florence, implies a remarkably simple scheme for equalizing short time intervals. The secret of his success, it now appears, was a song

by Stillman Drake

Einstein saw Galileo as a kindred spirit who had faced problems not very different from his own. The testing of relativity theory pressed the limits of observational accuracy, and so had Galileo's experiments on free fall three centuries earlier. Perhaps Einstein had that in mind when he wrote, in the foreword to my English translation of Galileo's *Dialogue Concerning the Two Chief World Systems:* "The experimental methods at Galileo's disposal were so imperfect that only the boldest speculation could possibly bridge the gaps between empirical data. For example, there existed no means to measure times shorter than a second."

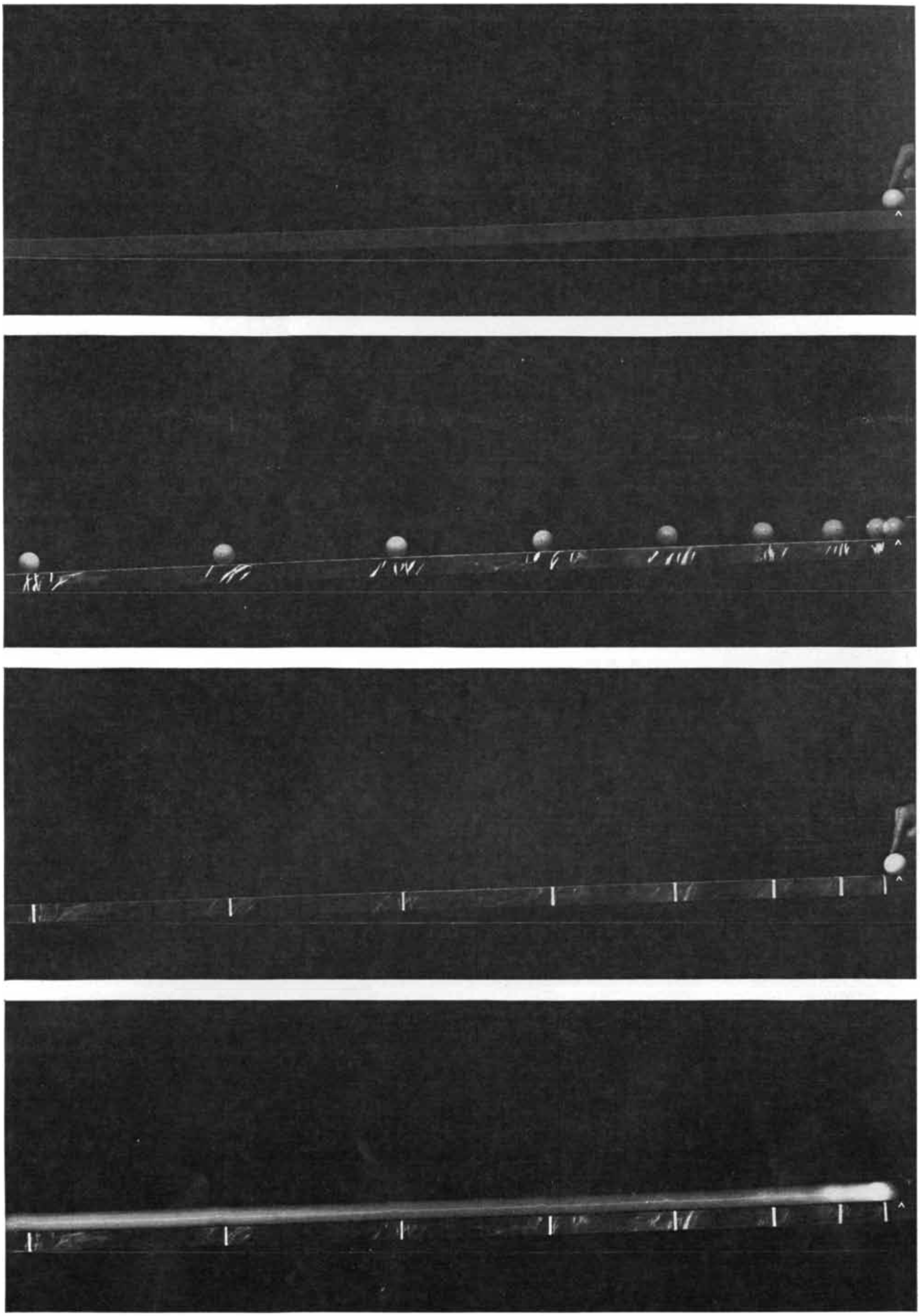
It is true that in Galileo's day there were no reliable clocks or watches, let alone the instruments and techniques that are now available for measuring extremely short times. It does not follow, however, that Galileo lacked any means by which he could divide time into equal intervals shorter than a second. We have grown so accustomed to precise instruments calibrated in standard units that we may think they are necessary for exact experimentation. That is not the case, however; if it were, modern physics could never have got started. Bold speculation has often advanced physical science, but it alone did not suffice to inaugurate modern physics. Neither classical antiquity nor the Middle Ages lacked for brilliant speculations, but for 2,000 years they did nothing useful to bridge the gaps in the empirical data. The modern era began with Galileo's law of falling bodies, and if Galileo had had no way of dividing time into intervals of less than a second, he could never have established that law firmly enough for it to deserve acceptance.

The phrase "measure time" makes us think at once of some standard unit such as the astronomical second. Galileo could not measure time with that kind of accuracy. His mathematical physics was based entirely on ratios, not on standard units as such. In order to compare ratios of times it is necessary only to be able to divide time equally; it is not necessary to name the units, let alone to measure them in seconds. The conductor of an orchestra, moving his baton, divides time evenly with great precision over long periods without thinking of seconds or any other standard unit. He maintains a certain even beat according to an internal rhythm, and he can divide that beat in half again and again with an accuracy rivaling that of any mechanical instrument. If the cymbalist in the orchestra were to miss his entry by a tiny fraction of a second, say by a 64th note in the music, everyone would notice it, not just the conductor. Professional cymbalists have virtually perfect timing, although many of us may feel that theirs is the one place in the orchestra we could learn to fill. In that we are mistaken, but the point is that we would not be tempted to think we could unless we believed it is quite easy for anyone to divide time accurately into very small intervals.

In fact, that is what we do whenever we dance or sing. A dance step may take half a second, but an error of much less than half a second is quite noticeable and even unbearable. Few people can stand the irregularity of the ticks when a metronome or a pendulum clock is slightly tilted. Most of us can sing to an established beat and would wince if a pianist played a grace note where the principal note ought to be (that is, a hundredth of a second too late). And it is not difficult for us to tell, when we hear two notes close together, which one we heard first.

I have been describing equipment that was second nature to Galileo, whose father and brother were professional musicians and who himself composed creditable music for the lute, an instru-

RECONSTRUCTION OF AN EXPERIMENT believed to have been conducted originally by Galileo in 1604 appears in the sequence of photographs on the opposite page. The experiment, part of a series of investigations that led Galileo ultimately to the correct "times squared" rule for the proportionality between the distance an object falls from rest and the elapsed time, was reconstructed for Scientific American by photographer Ben Rose according to specifications supplied by the author. The objective of this modernized test was to measure as precisely as possible the distances traveled from rest by a ball rolling down an inclined plane at the ends of eight equal times (in this case at .55-second intervals). The grooved inclined plane used in the reconstruction was 6½ feet long and was set at an angle of 1.7 degrees; it was fitted with a stop at the higher end, against which the two-inch steel ball could be held (*top*). The time intervals were established by singing "Onward, Christian Soldiers" at a tempo of about two notes per second. At one note the ball was released, and the positions of the ball at subsequent notes were marked with chalk; for comparison the exact .55-second positions were also captured by multiple-flash photography (*second from top*). A rubber band was then put around the plane at each chalk mark (*third from top*). The positions of the rubber bands were adjusted so that the audible bump made by the ball in passing each band would always come exactly at a note of the march; the bumps were visualized by leaving the camera shutter open during an entire run (*bottom*). Finally the distances between pairs of adjacent bands were measured. The ratios of the successive intervals were found to agree closely with a set of figures recorded by Galileo on a page designated *f. 107v* in Volume 72 of the Galilean manuscripts (*see illustration on page 100*).

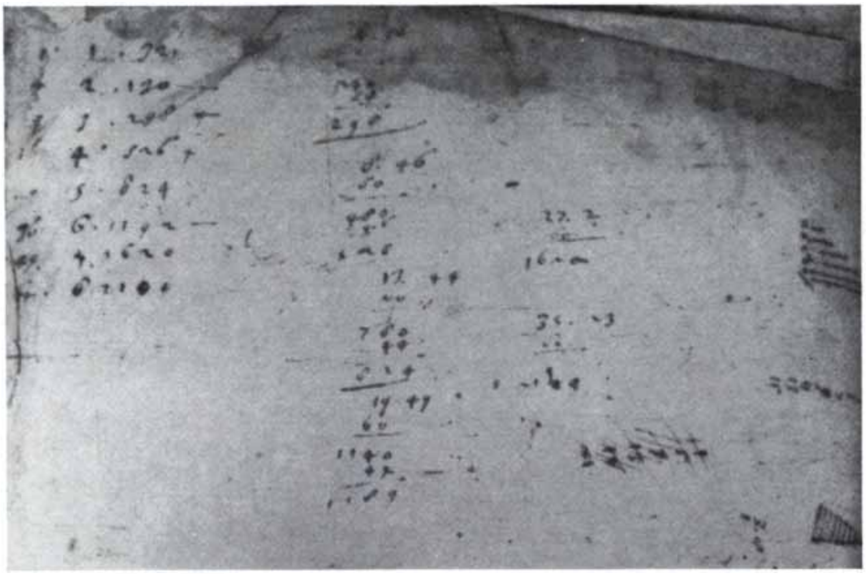


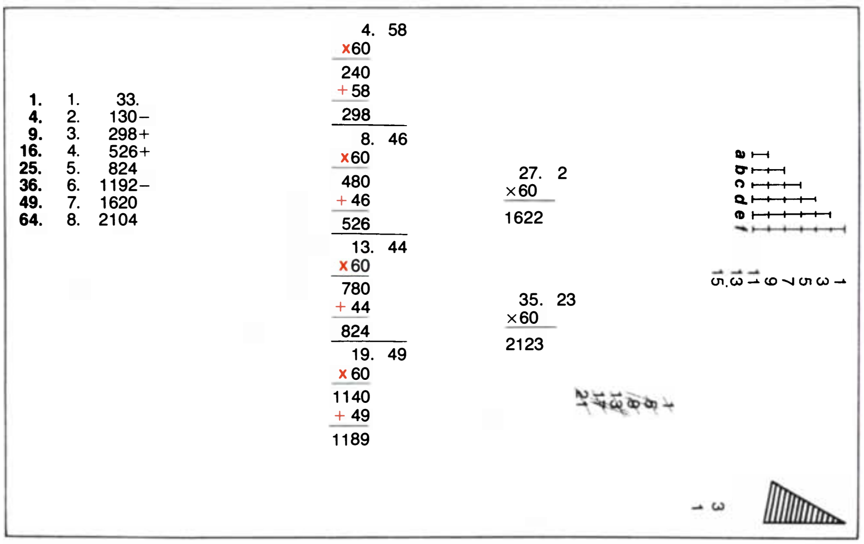
ment he is said to have played quite well. We may think such equipment is not very suitable for scientific experiments, but that is mainly because we no longer need it for them. It was certainly not suitable for a published experiment designed to convince Galileo's readers; even in his day it would have been foolish to write, "I tested this law by singing a song while a ball was rolling down a plane, and it proved quite exact." As a result the experiments later described in print by Galileo were different. Instead of dividing the time equally he fixed the distances in established ratios and then measured the times over the distances by weighing on a delicate balance the water that had flowed from a large container through a narrow orifice. The ratios of distances agreed with the square of the ratio of the times thus measured—within a tenth of a pulse beat, Galileo declared. His published experiments were duplicated in 1960 by the historian of science Thomas B. Settle, who found that Galileo's stated accuracy was more than confirmed.

Galileo did not say, however, that he first obtained his law by performing experiments. All he said about the discovery was that nature had led him almost by the hand to his rule that in natural motion equal speeds are added in equal times. Historians have poked fun at Galileo for that picturesque statement. In October, 1604, Galileo wrote to his friend Paolo Sarpi that the times-squared law followed from a different assumption: that "speeds" in fall are proportional to the distances fallen from rest. No evidence was forthcoming that nature's guiding hand had led Galileo out of his apparent error during at least the next four years, and possibly longer.

Recently, however, things have begun to look different. In an earlier article in this magazine I presented evidence that Galileo's word for "speed" in his letter to Sarpi did not mean our v but rather what we write as v^2 [see "Galileo's Discovery of the Law of Free Fall," by Stillman Drake; SCIENTIFIC AMERICAN, May, 1973]. For the first time a certain page of Galileo's notes on motion (designated *f. 152r* in Volume 72 of the Galilean manuscripts preserved at the National Central Library in Florence) was published in full. The proportionality of speeds in acceleration to times elapsed from rest was already implied in that document, which was almost certainly written before Galileo's letter to Sarpi. In his attempt to derive the law of free fall mathematically from a physical notion, however, he had decided to define "speed" as something that is in fact proportional to distance. His reasons can now be filled in from marginal notations on *f. 107v*, another previously unpublished page of Galileo's notes in the same volume [*see illustration at left*].

First, from *f. 107v* we can reconstruct an experiment that probably took place in 1604, after Galileo had abandoned his first hypothesis on *f. 152r* and before he went on to establish a correct rule for speeds and distances from rest. This new document looks singularly unpromising at first glance, but it records an actual experiment. The figures listed in the third column at the top left represent very nearly the distances from rest of a ball rolling down an inclined plane at the ends of eight equal times. The dis-





GALILEO'S RESULTS of his 1604 experiment appear on the manuscript page shown in the photograph at top. The page, designated *f. 107v*, is reproduced in part here through the courtesy of the National Central Library in Florence. The transcription at bottom gives the relevant figures in a modern typeface; some computational signs have been added in color for clarity. The figures listed in the third column at upper left in the reproduction represent very nearly the distances from rest of a ball rolling down an inclined plane at the ends of eight equal times. The figures differ slightly from the products of the first number, 33, by the squares of the numbers from 1 to 8. The correct square-number multipliers (1, 4, 9, 16, 25, 36, 49, 64) were later added in the first column with a different pen and ink. The calculations at center show how Galileo actually got the numbers he entered in the third column. In each case he multiplied 60 by some integer and then added a number less than 60, probably reflecting the fact that he measured the distances with a short ruler divided into 60 equal parts. The series of alternate odd numbers (1, 5, 9, 13, 17, 21) written down at lower right (and then canceled) appears to represent his first guess at a rule for the growth of the figures in the third column. The significance of the other notations is discussed in the text.

tances were actual, not theoretical, since they differ slightly from the products of the first number, 33, by the square numbers 1, 4, 9, ..., 64. Those numbers in the first column, barely legible in reproduction, were added later with a different pen and ink. Galileo's figures in the third column are not exact multiples of squares, and neither could they have been obtained by using water-timing as in Galileo's published experiment. To get such accurate figures a single time had to be divided into eight equal intervals. I believe this was done as outlined below, which will explain why in spite of this experiment's interest and historical importance it was not published by Galileo.

TIME	TIME^2	DISTANCE (CENTIMETERS)	DISTANCE (POINTS)	FIGURES IN *f.107v*	DIFFERENCE (POINTS)	DISTANCE IN 1/64 SECOND (POINTS)
.55	.30	3.176	32.9	33	+ .1	1.8
1.10	1.21	12.705	131.4	130	− 1.4	3.7
1.65	2.72	28.59	295.7	298	+ 2.3	5.6
2.20	4.84	50.82	525.7	526	+ .3	7.4
2.75	7.56	79.41	821.5	824	+ 2.5	9.3
3.30	10.89	114.3	1,182.4	1,192	+ 9.6	11.2
3.85	14.82	155.6	1,609.8	1,620	+10.2	13.1
4.40	19.36	203.3	2,103.1	2,123 [2,104]	+20.1 [+ .9]	14.9

CLOSE AGREEMENT is evident between Galileo's figures and the theoretical figures for the experiment assumed by the author and his colleagues to underlie the *f. 107v* data. The distances passed by a ball rolling on the inclined plane in eight equal times were calculated both in centimeters and in Galilean "points." (A "point" in Galileo's system of measurement equaled 29/30ths of a millimeter.) The plane was assumed to have an elevation of 60 points in 2,000, giving a theoretical acceleration of 21 centimeters per second squared. Two figures are given in the table for Galileo's final distance, the first being the one he originally measured and the second being the altered figure he entered in the third column of *f. 107v*. Before correction this figure had been the only one showing a difference greater than can be accounted for by assuming a tolerance equal to the distance run by the ball in a 64th of a second at each speed; calculated tolerances are indicated in last column at right.

Place a grooved inclined plane about 6½ feet long at an angle of about 1.7 degrees. Fit a stop at the higher end, against which a steel ball can be held by resting a finger on it lightly. Now sing a simple march such as "Onward, Christian Soldiers" at a tempo of about two notes per second, very crisply. When the tempo is established, release the ball at some note and mark with chalk the position of the ball at other notes (half-second intervals). In three or four runs eight positions can be marked; put a rubber band around the plane at each mark. (Galileo would have tied gut frets at these places, in the way that adjustable frets were tied around the neck of a lute, but of course today rubber bands are easier to place and adjust.) Then, making many repeated runs, adjust the rubber bands so that the audible bump made by the ball in passing each one always coincides exactly with a note of the march. When the inclined plane has thus become a kind of metronome, measure the distance in millimeters from the resting position of the ball to the bottom of each rubber band. The figures should agree tolerably well with those in the third column on *f. 107v*.

Galileo's calculations to the right show how he actually got the numbers entered in the third column. In each calculation he multiplied 60 by some integer and then added a number less than 60. That is just what someone would do if he were measuring distances with a short ruler divided into 60 equal parts. Galileo certainly had such a ruler, because several diagrams in his notes on motion were drawn so that a principal line contained 60 times the number of units in which all the other lines were measured. In one note he gave "180 points" as the length of a line I measured as 174 millimeters, so that we may take Galileo's unit as being 29/30ths of a millimeter. That is almost exactly the shortest graduation (which he called a "point") measured on his own proportional compass at the Museum of the History of Science in Florence.

Since Galileo actually measured the distances listed on *f. 107v*, it is safe to say that when he started, he did not yet know the times-squared law. Otherwise his natural procedure would have been to tie frets at appropriately calculated distances, saving many runs of the ball and a lot of adjusting. In that way a single run would have confirmed the law and there would have been nothing to write down. Hence the actual calculations, combined with evidence that the first column was written after the third, indicate that Galileo did not have the law of free fall when he began *f. 107v*.

Still further evidence is provided by the series of alternate odd numbers (1, 5, 9, 13, 17, 21) he wrote down on *f. 107v* and then canceled. That represents his first guess at a rule for the growth of the figures in the third column. It may seem a strange guess, but (as I shall explain below) Galileo had good reason to expect some kind of regular additions to the speeds and distances in successive equal times. Since the actual rule of growth is 1, 4, 9, 16, ..., Galileo's canceled figures were not far wrong up to the fifth, and he stopped at the sixth.

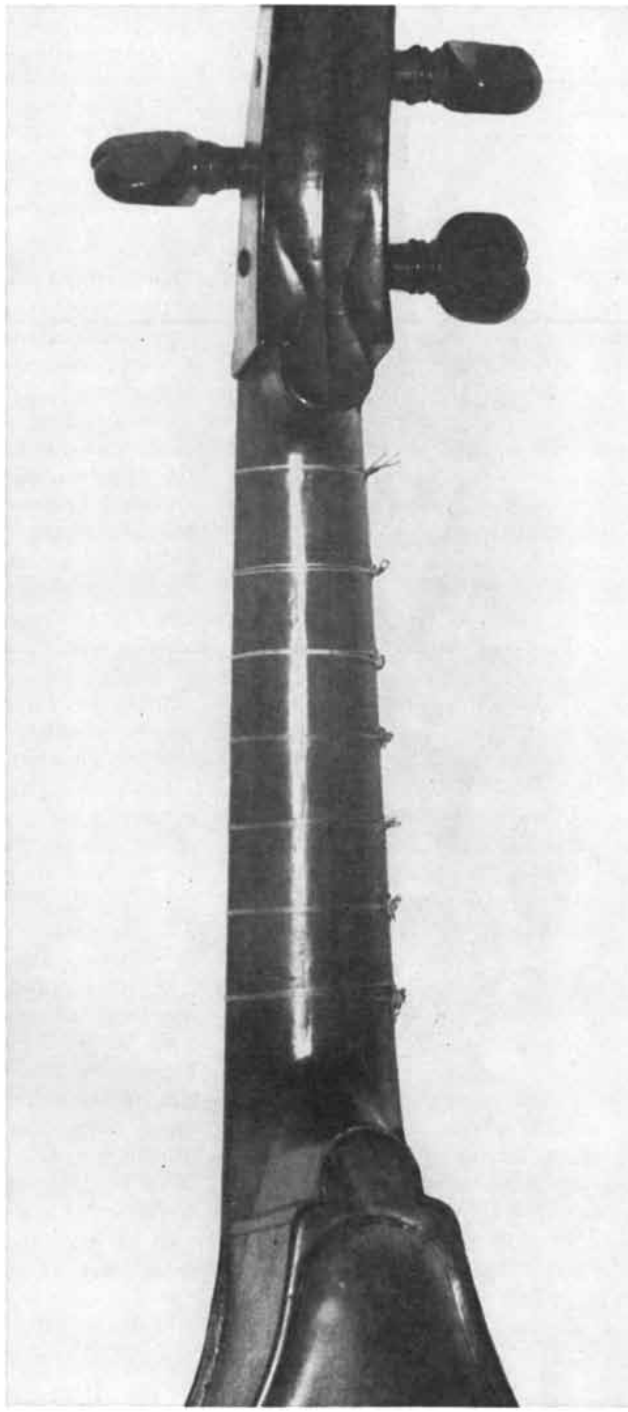
If my reconstruction is correct, this experiment will surprise other historians of science as much as it did me. We all believed that until considerably later than Galileo's lifetime, experiments were made only to test some preconceived rule, not to seek a rule from measurements. The only exception previously known to me was a series of experiments by Galileo's father, Vincenzio Galilei. He discovered by measurement that the weights loading similar strings of equal length that sound a given musical interval are as the inverse squares of the string lengths sounding that same interval under equal load. Music thus appears to have had a dual role in the beginnings of experimental physics, since both pitch and time played their part.

The slope of plane and unit of time Galileo used to get the data recorded on *f. 107v* cannot be exactly specified, but there are some practical limitations on both. A plane much steeper than two degrees will produce an inconveniently high speed too quickly, whereas a much gentler slope causes the ball to hesitate at the first few frets. Intervals of less than half a second or more than a second cannot be judged to be equal as accurately as those within that range, perhaps because the heartbeat (and its half) supply unconscious rhythms. A probable incline for the experiment is reasonably suggested by the length of the plane used (2,100 points) and the length of Galileo's ruler (60 points). The same plane, but one more steeply tilted, was probably again used for Galileo's experiments leading to the discovery of the parabolic trajectory, in which the greatest length needed was 2,000 points [see "Galileo's Discovery of the Parabolic Trajectory," by Stillman Drake and James MacLachlan; SCIENTIFIC AMERICAN, March]. A ball rolling on a plane elevated 60 points in 2,000 has an acceleration of 21 centimeters per second squared, for which the distances passed in eight equal times out to 2,100 points agree closely with Galileo's data. The distance run in a 64th of a second at each speed yields a probable variance from the ideal placement of frets [*see illustration above*].

Two figures are given in the table for

the final distance, the first being the one Galileo originally measured and the second being his altered figure in the third column. Before correction this distance had been the only one showing a difference greater than can be accounted for by a tolerance of a 64th of a second. It is interesting that after Galileo's correction the last fret was almost exactly in place, whereas the two preceding frets were left about a centimeter off, although they were still within the assumed range of tolerance. Probably after the frets had been placed to Galileo's satisfaction and the distances had been measured and recorded he made a final run or two and decided that the last bump was definitely late. After two of the other figures he wrote a plus mark and after two a minus mark, probably indicating that those sounds seemed a little early or late but not enough to require further adjustments.

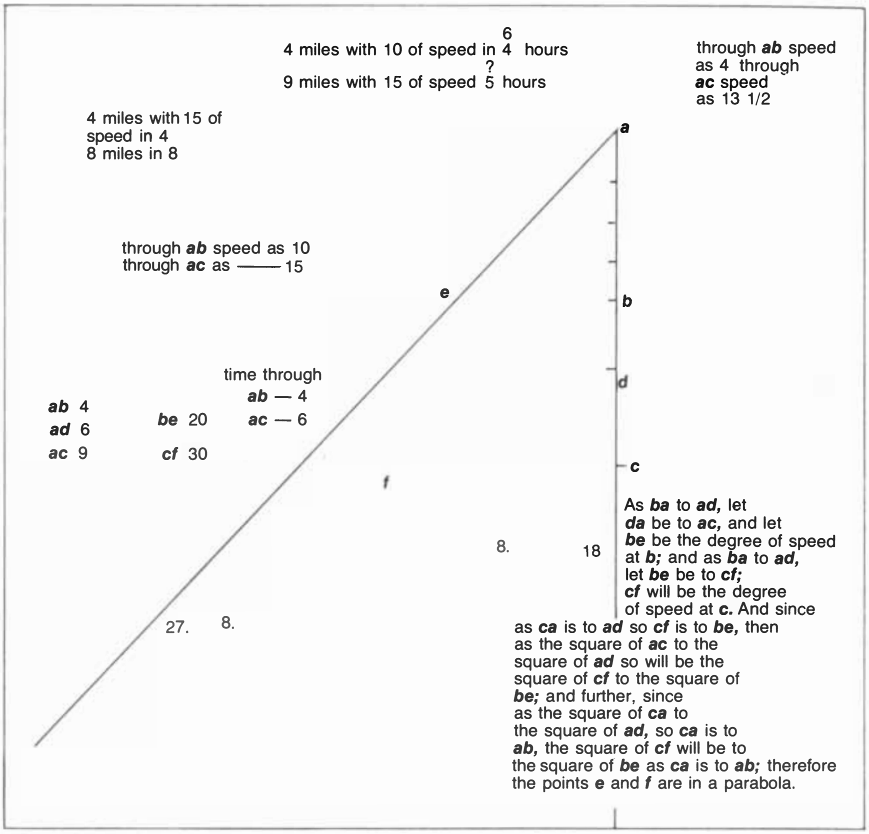
Consistently positive deviations (except for the second fret) are not surprising. A trained observer will judge the instant a star crosses the meridian with great consistency among his own data, but his data will not be identical with those of another equally skilled observ-



ADJUSTABLE GUT FRETS similar to the rubber-band frets employed in the reconstructed experiment shown in the illustration on page 99 would have been quite familiar to Galileo, who came from a family of musicians. Unlike modern stringed instruments, which have fixed frets, old instruments such as the lute and the viola da gamba had movable frets around the neck of the instrument. The photograph at left shows a viola da gamba belonging to the author; close-up at right shows how gut frets were tied.

er. Astronomers have long known of this phenomenon, which is technically called the personal equation. If my reconstruction happens to be right for the slope and time used by Galileo, it means that he judged each beat to be simultaneous with a bump when the bump was in fact a little after the beat. (A different slope or time would at most reverse the sign, without changing the explanation.) The negative difference in placing the second fret, on the other hand, has a physical explanation rather than a psychological one. In practice unless the ball is very massive it is delayed in climbing the first fret, so that the second fret must be moved a little closer than it is in theory in order to get the second time equal to the first. No such delay is noticeable after the ball is really under way.

Thus everything about the data on *f. 107v* conforms with the foregoing experimental procedure, except perhaps that the high degree of accuracy seems incredible. Yet once one abandons the notion that precisely calibrated instruments are absolutely necessary it is apparent that Galileo had more than one means available to him for the equal division of small times. He knew about the isochronism of the pendulum by 1602, so that he could have used returns to the vertical of a pendulum hung from a point on a vertical line. Or he could have used the sound of drops of water falling into a resonant basin from a small hole in a large tank. In practice, however, it is harder to judge the agreement of a sound with a position of a moving pendulum, or even the agreement of two different sounds coming from outside sources, than it is to decide when a single set of external sounds agrees with a strong internal rhythm. Those who have studied music with a professional performer know how remarkably the last ability can be developed. That is why the very accuracy of Galileo's data persuades me that his experiment was of the kind I first described.



ENGLISH TRANSCRIPTION of a portion of *f. 152r*, another manuscript page of Galileo's notes uncovered in recent years, can now be reinterpreted in the light of *f. 107v*. This document, which represented Galileo's first attempted logical derivation of speeds in free fall, was first published in full in the May 1973 issue of SCIENTIFIC AMERICAN. Galileo started with the false hypothesis at top (that overall speeds are the sums of consecutive numbers beginning with 1), which caused him to obtain conflicting ratios. After he found the correct time-distance relation on *f. 107v*, he resumed work on *f. 152r*, finishing with derivation at right, which assumes that overall speeds are proportional to speeds acquired in free fall.

So much for the musical and experimental implications of *f. 107v*. Its other parts—the notations in the right-hand margin—are equally interesting for the light they shed on Galileo's return to *f. 152r* for its successful completion and on his first attempted logical derivation of the law of free fall. Those matters were discussed in my 1973 article, but on the assumption that mathematical persistence and ingenuity alone accounted for Galileo's discovery of the times-squared law for distances in free fall. With a little repetition of what was said there, more steps can now be filled in from *f. 107v*, as follows.

From the Middle Ages to the time of Galileo the usual way of thinking about actual acceleration in free fall was in terms of a series of small and rapid spurts of uniform speed, accumulating as though by quantum jumps. The reason for this tradition was that early physicists were primarily interested in causal explanations. Increments of added impetus provided a satisfactory cause for successive acceleration, but any truly continuous change of speed seemed to require different effects within a single cause, contrary to certain of Aristotle's principles. (It made sense for the early physicists to study change, but not change of change.) When Galileo began his search for the mathematics of acceleration on *f. 152r*, he likewise assumed successive increments, and he sought the "overall speed" from rest by adding up separate "degrees of speed," putting $1 + 2 + 3 + 4 = 10$ "degrees" as the overall speed for four units of time and $1 + 2 + 3 + 4 + 5 = 15$ "degrees" for five units of time. This hypothesis led him to contradictory meanings for "15 degrees of speed," however, and so he put *f. 152r* aside for a time.

The experiment of *f. 107v* left Galileo no room for doubt that distances from rest are related as the squares of the times. He went on to reason that since in this experiment the individual times were equal, the successive speeds (each of these being thought of as a brief overall speed) must be proportional to the successive distances passed. And since the overall distances from rest went up as the square numbers, the successive partial distances must go up as the odd numbers. (Mathematicians used to call the odd numbers the "gnomon" of the squares, because when one adds in turn 1, +3, +5, +7, . . . , one gets 1, 4, 9, 16, ...) Therefore in order to represent the increasing speeds Galileo wrote the series of odd numbers seen at the right-hand margin. To the left he drew a series of uniformly increasing parallel vertical lines, in the way that Archimedes always represented an arithmetical progression.

The series of odd numbers has the unique property among arithmetical progressions that the ratio of the first number to the second is also the ratio of the first two numbers to the second two, and so on. Thus 1 is to 3 as $1 + 3$ is to $5 + 7$ and as $1 + 3 + 5$ is to $7 + 9 + 11$. This means that if the initial time unit had been any multiple of the one used, the relations would still have been the same. That thought had probably struck Galileo when he wrote down a 3 over a 1 near the odd numbers on *f. 107v*. Then he drew the little triangle at the right to represent overall speeds as being built up of successive speeds in small equal times, which speeds he represented by lines parallel to the base. In effect this was merely enclosing in a triangle the Archimedean diagram for an arithmetical progression. There is a logical oversight in this step, however, because the triangle actually comes to a point, whereas the other diagram always begins with a line, no matter how short a line. This particular triangular diagram was accordingly destined to lead Galileo astray several times in the ensuing years, whenever he sought a purely logical foundation for the law of free fall. Having represented speeds by lines parallel to the base and distances of fall by the altitude, he thought he had somehow to make "speeds" proportional to distances from rest. Actually what we call speeds are proportional to the square roots of those distances. Galileo eventually treated them as such, but not in 1604.

If I am correct about the order in which Galileo's notes were written, he quickly derived the square-root rule for speeds and almost as promptly rejected it. This rule was derived on *f. 152r* from the simple (and correct) assumption that the speeds at two given points in fall are proportional to the overall speeds of a body in reaching those points from rest. The two distances Galileo happened to select for his original hypothesis on *f. 152r* were 4 and 9, and by a coincidence the two overall speeds he had then assigned were 10 and 15, as I have explained above. The square roots of those distances are in the same ratio as the next two "triangular" numbers used for speeds, but that is not usually the case. (For example, the roots of 9 and 16 are not in the same ratio as 15 and 21, the next two triangular numbers.) Galileo had before him two "overall speeds" based on the traditional discrete-addition concept, which happened to fit the square-root relation that was really valid only for continuous acceleration from rest. He duly noted on *f. 152r* that the square-root relation would make the speeds plotted against distances fall along a parabola. Since there was no way to reconcile that with the triangle of "speeds" already drawn on *f. 107v*, a choice had to be made.

It seems that Galileo put his faith in the diagram he had arrived at (by a logical oversight) from his experiment on *f. 107v* rather than in the correct assumption he had made on *f. 152r*. Thus in the demonstration of the times-squared law that he went on to work out for Sarpi he decided to make his "overall speeds" proportional not to the speeds acquired but to their squares. He made the acquired "degrees of speed" in turn proportional to the distances from rest. That came about because when Galileo sought a physical meaning for "speed," he hit on "the effects of machines that work by striking," as he said in the 1604 derivation. A pile driver doubles its effect by dropping the weight from a doubled height. Galileo reasoned that the weight stays the same and all that changes is the speed, so that it was reasonable to make "speed" proportional to distance fallen. Galileo changed this to time of fall when he discovered the parabolic path of projectiles five years later.

Einstein might have been surprised at the role played by precise experiment in Galileo's discovery of the law of free fall, although as a musician Einstein would have particularly liked its origin. Galileo's boldest speculation bridged the gap not between empirical data for lack of means to deal with times shorter than a second but between the discrete and the continuous. It occurred to Galileo only after he had made several more attempts to reconcile his first notion of additive speeds with mathematically continuous acceleration from rest. We still cannot prove experimentally, any more than Galileo could, that continuous change really underlies the changes we can measure only in discrete units of some kind. That assumption took an act of faith and resulted in a bold speculation that put Galileo at odds with some of his ablest contemporaries. In accepting a mathematically continuous change of speed from rest, Galileo consciously abandoned the traditional search for a physical cause of acceleration in free fall and contented himself with its mathematical description. The elimination of cause as a necessary concept in physics was as distasteful to most of Galileo's contemporaries as the elimination of force was later to many of Einstein's.

It's hard for one computer to be all things to all people.

So if your big computer can't do what somebody wants done, get a small computer to do the job.

A Data General computer.

At Lowe's Companies (a group of 130 building material stores) the central computer knew the current prices and stock levels of the 7,000 items in stock at each store. But the salesmen didn't.

So Lowe's data processing management put Data General computers in their stores to give the salesmen all the information they needed. Our computers cost a third less than the IBM 370 it would have taken to do the job. And ours do the job better.

At the end of every day, our computers automatically report to the computer center in the home office. And by the beginning of the next day, the central computer sends each store the latest prices and new stock information.

Data General computers not only make order-taking easier and more accurate, they also cut the time needed to take an order by 30% and reduce inventory carrying costs, pricing errors, and bad debt losses.

RCA Records put our computers in the hands of their operating management so they could do order processing and inventory control without getting caught up in the complications of a big computer.

One of the largest banks in the Northeast put a Data General computer in their trust department to give their portfolio officers access to vital investment information in their large IBM computer. And our computer does the job for much less than it cost for the large computer to do it alone.

So instead of giving people reasons why your big computer can't do something, write for our brochure, "The Sensible Way to Use Computers."

Then when somebody starts complaining about your computer, you'll know enough to tell him where to go.

SCI-675

NAME

TITLE

FIRM

ADDRESS

CITY STATE ZIP

Data General, Dept. G, Route 9, Southboro, Mass. 01772 (617) 485-9100. Data General (Canada) Ltd., Ontario, Data General Europe, 15 Rue Le Sueur, Paris 75116, France. Data General Australia, Melbourne (03) 82-1361/Sydney (02) 908-1366.

DataGeneral

MATHEMATICAL GAMES

Games of strategy for two players: star nim, meander, dodgem and rex

by Martin Gardner

Consider a two-person game with the following characteristics: (1) It is a game of perfect information, that is, both players have complete knowledge of the game's structure after every move. (2) The players move alternately. (3) Decisions are not made by chance. (4) The game ends after a finite number of moves with a win by one player. (No draw is possible.)

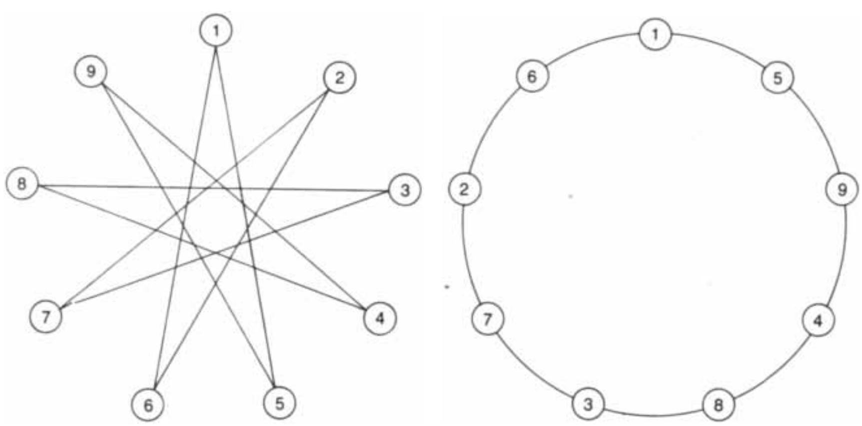
It is not hard to see that there must be a winning strategy for either the first player or the second. If the first player (henceforth called *A*) does not have a winning strategy, he must lose. This means that the second player, *B*, has a winning strategy. Does the argument apply if we rescind the requirement that the game end in a finite number of moves?

Curiously, it all depends on whether or not one accepts the "axiom of choice." This notorious axiom says that from any collection (finite or infinite) of nonempty sets, with no elements in common, you can form a new set by taking one element from each set. In the 1930's Stefan Banach, Stanislaw Mazur and Stanislaw Ulam discovered a type of infinite game in which neither *A* nor *B* has a winning strategy if the axiom of choice is accepted. Someone argued that this proves the Unitarian dogma that there is "at the most" one God, because if two gods could play such a game, neither could know a winning strategy and therefore neither could be called omniscient!

That, however, is by the way. Here we shall examine some new two-person nonchance games for which the rules are extremely simple and for which a winning strategy is either known or capable of being known. All but one of the games are played with counters on boards that are easily drawn on cardboard. Two differently colored sets of counters, such as go stones or small poker chips, will be useful for any reader who wants to play or to analyze the games.

An example of an almost trivial game of the nim type, but one with a strategy that is not immediately apparent, is played on the star pattern shown in the illustration below. Put a counter on each of the star's nine points. *A* and *B* take turns removing either one counter or any two counters joined by a straight line segment. The player who takes the last counter wins.

B can always win at star nim by a strategy based on the board's symmetry. Imagine that the black lines are strings. The pattern can be opened up to a circle that is topologically equivalent to the star. If *A* takes one counter from this circle, *B* takes the two counters that are directly opposite. If *A* takes two counters, *B* takes one counter that is directly opposite. In each case two sets of three counters are left. Now, whatever *A* takes from one set, *B* takes the corresponding counter or counters from the other set. Obviously *B* will get the last counter. If the reader plays a few games on the circle, translating each move to its equivalent on the star, he will soon see how to use the star's symmetry for playing the strategy.



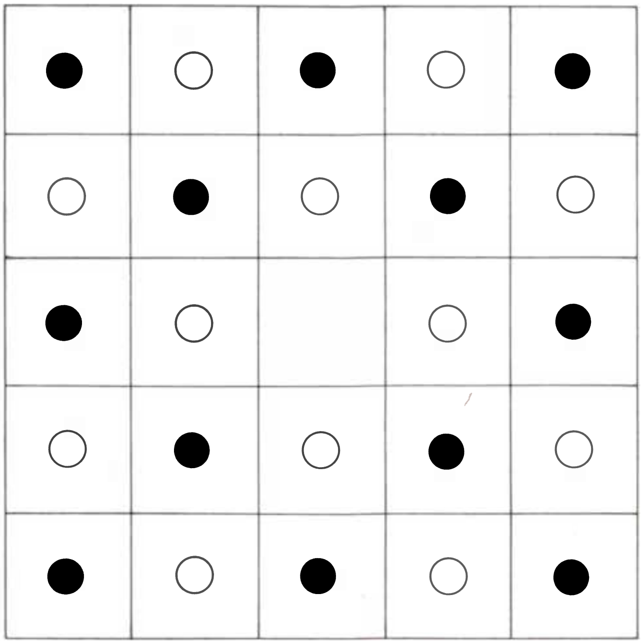
***Star nim* (left) *and its winning strategy* (right)**

In the late 1960's G. W. Lewthwaite of Thurso in Scotland invented a delightful game with an artfully concealed "pairing strategy" that gives the second player a sure win. On a 5-by-5 square matrix place 13 black counters and 12 white counters in alternating checkerboard fashion. Any one of the black counters, say the one in the center, is removed [*see top illustration at left on opposite page*].

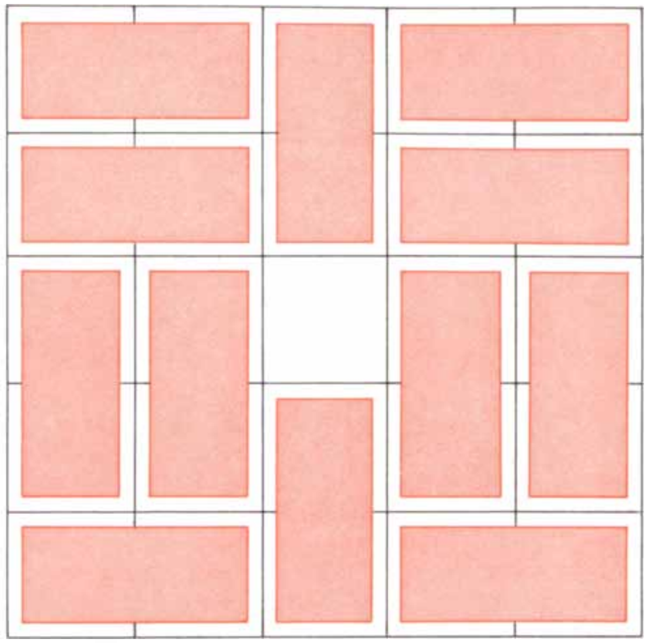
Player *A* controls the white counters and *B* the black. They take turns moving one of their counters orthogonally to the vacant square until a player loses by being unable to move. If the board is colored like a checkerboard, it is obvious that on each move a counter goes to a square of different color and that no counter can be moved twice. The game therefore cannot go beyond 12 moves for each player. It may end before then, however, in a win for either player unless *B* plays rationally.

B's strategy is to imagine that the matrix, except for the initially vacant cell, is covered with 12 nonoverlapping dominoes. It does not matter how they are placed. The top illustration at the right on the opposite page shows a sample pattern. Whenever *A* moves, *B* simply moves his counter that is on the domino *A* has just vacated. Since this ensures that *B* always has a move to follow a move by *A*, *B* is sure to win in 12 or fewer moves.

The game can be played not only with counters but also with square tiles or cubes that slide within a matrix surrounded by a rim. Suppose the rules are amended to allow either player at any time to move a row or a column of three counters (or tiles), provided that the two end counters are of his color. This is a splendid example of how an apparently trivial alteration of a rule can enormously complicate a game's analysis. Lewthwaite was unable to find a winning strategy for either player in this variant of his game.



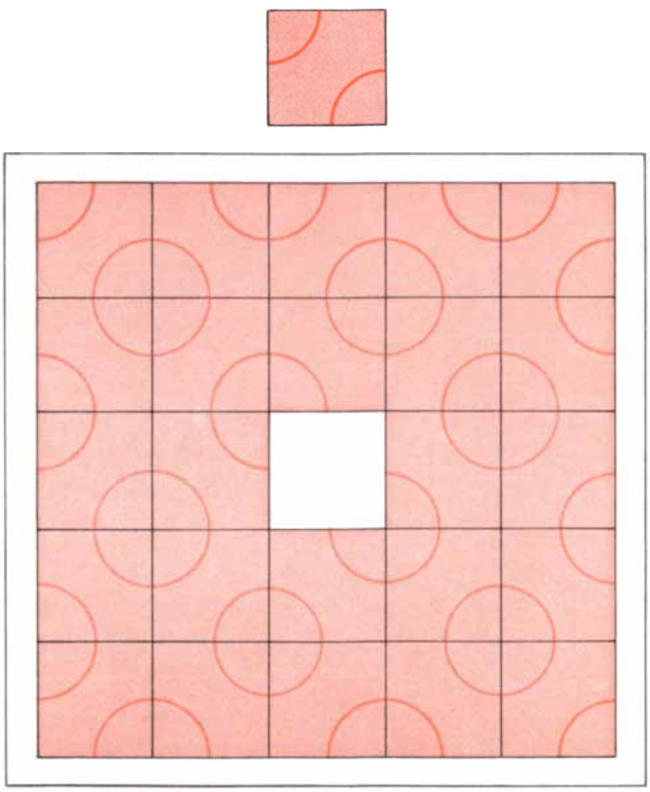
G. W. Lewthwaite's counter game



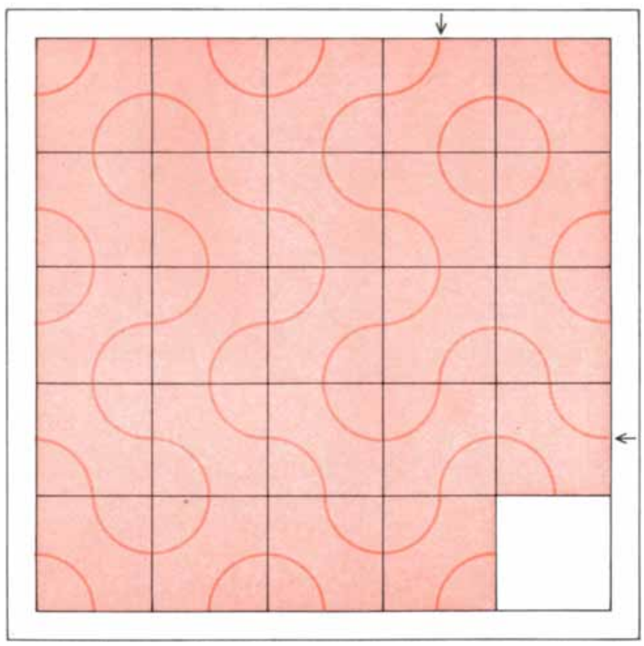
A pairing strategy for Lewthwaite's game

Games based on the sliding of unit squares within a square matrix offer a plethora of unexplored possibilities. Lewthwaite proposes an attractive game that he calls meander. It uses 24 identical tiles placed in a 5-by-5 tray to form the pattern shown in the illustration at the left below. The players take turns sliding a single counter or a straight row or column of two, three or four counters. The play continues until a player wins by creating a pattern in which at least three tiles form a continuous line or path that joins two edges (opposite or adjacent) of the tray. The illustration at the right below shows a winning pattern. The game is probably too complex for solving without a computer program, and perhaps too complex for solving even with one.

In 1972, when Colin Vout was a mathematics student at the University of Cambridge, he invented an intriguing counter game that he calls dodgem be-



Meander, with example of pattern on a tile at top



A possible winning pattern in meander

cause it is so often necessary for a piece to dodge around enemy pieces. It is playable on a checkerboard of any size. Even the game on a 3-by-3 board is complicated enough to be interesting.

Two black counters and two white ones are initially placed as shown in the upper illustration below. Black sits on the south side of the board and White sits on the west. The players alternately move a counter one cell forward or to their left or right, unless it is blocked by another counter of either color or by an edge of the board. Each player's goal is to move all his pieces off the far side of the board. In other words, Black moves orthogonally north, west or east and attempts to move both of his pieces off the north side of the board. White moves east, north or south and tries to move his pieces off the east side of the board.

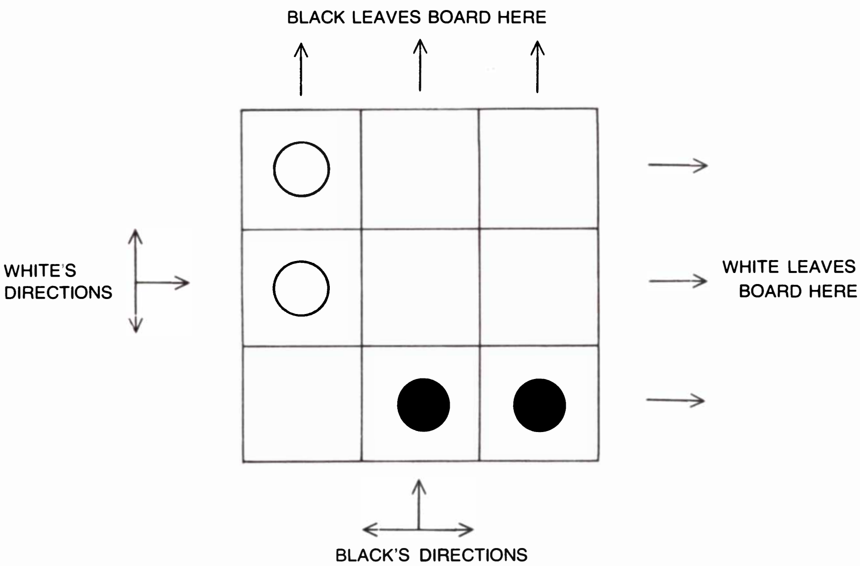
There are no captures. A player must always leave his opponent a legal move or else forfeit the game. The first to get all his pieces off the board wins. The lower illustration on this page shows a typical game won by Black.

Vout assures me that the first player has the win on the order-3 board, but as far as I know no games on higher-order boards have yet been solved. On a board of side n each player has $n - 1$ pieces placed on the west and south edges, with the southwest corner cell vacant. Played with seven checkers or pawns of one color and seven of another color on the standard order-8 checkerboard or chessboard, it is a most enjoyable game.

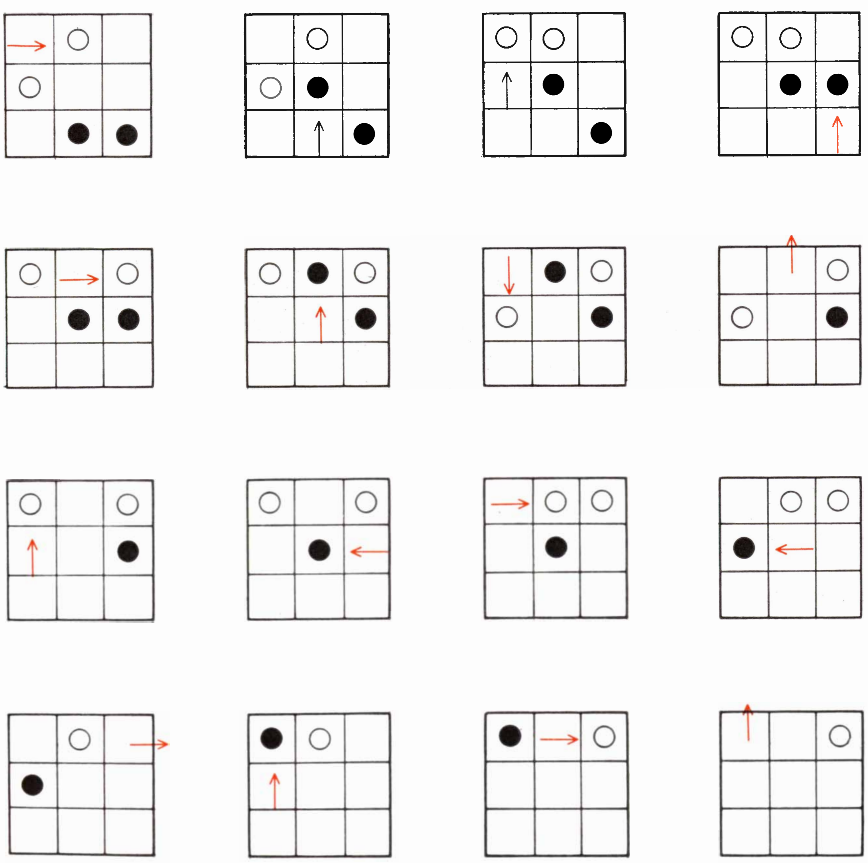
Piet Hein's now classic game of hex (see Chapter 8 of *The Scientific American Book of Mathematical Puzzles & Diversions*) remains unsolved except for small boards. For readers unfamiliar with the game, it is played on an n-by-n rhombus of hexagons such as the order-4 board shown in the top illustration on the opposite page. White opens by placing a white counter on a cell. Black follows with a black counter. They take turns placing counters on vacant cells (there are no moves or captures) until a player wins by forming a chain of adjacent counters that joins his side of the board to the opposite side, White by joining the north and south edges, Black by joining the east and west edges.

It is easy to see that no draw is possible. There is a famous proof by John F. Nash (who independently invented hex) that on a rhombus of any size the first player has a winning strategy, although the proof gives no hint of what the strategy is. Suppose White allows Black to tell him where he must make his first move. Can White still always win if he plays rationally? This modified version of hex has been called Beck's hex after Anatole Beck, who both proposed and solved it. Writing on hex in Chapter 5 of *Excursions into Mathematics*, by Beck, Michael Bleicher and Donald W. Crowe (Worth Publishers, 1969), Beck proves that if White opens in an acute corner cell, Black has a winning strategy if he plays next to it, on his first row. As Bleicher comments in a footnote, this "wrecks Beck's hex."

What about *misère*, or reverse, hex, known as rex, in which the first player to join his sides loses? As is so often the case in two-person games, the reverse



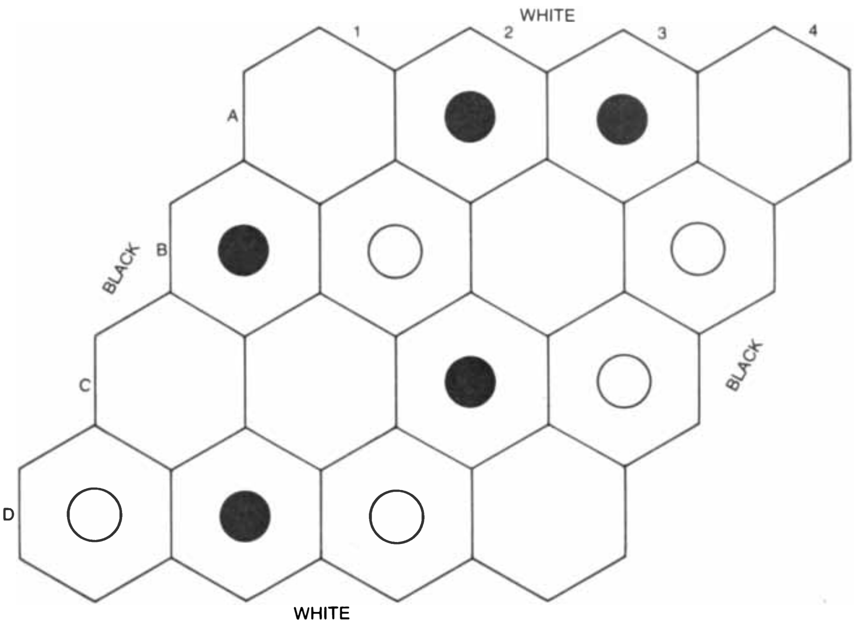
Colin Vout's dodgem



A dodgem game won by Black

game proves to be much harder to crack. No general strategy is known, although Robert O. Winder, in unpublished arguments, has shown the existence of a first-player winning strategy in rex of even order and a second-player winning strategy for all odd orders. More recently Ronald J. Evans has carried Winder's arguments a step further by showing that on even-order boards there is a winning strategy if White opens in the acute corner (see Evans' paper "A Winning Opening in Reverse Hex," in *Journal of Recreational Mathematics*, Volume 7, Summer, 1974, pages 189–192).

Rex on the order-2 rhombus is trivial, and it is not difficult to analyze exhaustively on the order-3 rhombus. Play on the order-4 rhombus is so complicated, however, that even though it is known that an acute-corner opening initiates a win, the strategy itself remains unformulated. The position shown in the top illustration on this page is an order-4 rex problem composed by Evans. Can the reader determine White's only correct move? It will be given next month.



Rex, a reverse hex game, with White to play and win

Here is an even simpler game for which no general strategy is known. It is played on a 1-by-n board (a single row of n squares) with counters that are all alike. A and B take turns placing a counter until one player wins by getting three counters adjacent. Could anything be simpler? A can always win when n is odd by first taking the center cell, then playing symmetrically opposite his opponent thereafter. For even n, however, things are not so simple. On most even rows A seems to have the win, but not necessarily, and the exceptions follow no known rule. Take $n = 6$, for example. The reader may enjoy working it out to see who has the win.

John Horton Conway has pointed out that this game is equivalent to a game I called 1-by-n cram in this department for February, 1974, except that it is played with trominoes instead of dominoes. It is easy to see the isomorphism. In playing the game as described above it is obviously disastrous to place a counter either next to another counter or one cell from it, since either move gives the opponent an instant win. Hence we might as well prohibit both moves. An easy way to do it is to require that each play consist of a triplet of adjacent counters, which is the same as placing a tromino on the field. (The middle of the triplet corresponds to placing a single counter, and the ends of the triplet enforce the two new rules.) The winner is the player who places a tromino last. (To complete the equivalence we must allow the placing of a tromino at either end of the field so that it extends one cell beyond the end.) Of course, the game can also be played by forming a row of n counters and removing them by alternate moves of taking three adjacent counters.

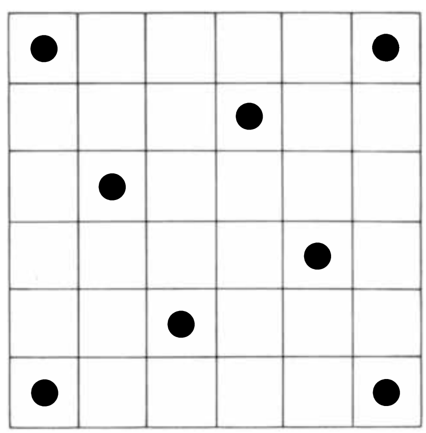
This triplet version of cram is considered in a classic paper by Richard K. Guy and Cedric A. B. Smith, "The G-Values of Various Games" (*Proceedings of the Cambridge Philosophical Society*, Volume 52, July, 1956, pages 514–526), where it is coded as Game .007. Elwyn Berlekamp has computer-analyzed the game to very high n without finding any periodicity in the Grundy numbers, which means that no one is even close yet to a general rule. The *misère* version of 1-by-n tromino cram, like its domino counterpart, is also unsolved.

Ulam has proposed extending the counter form of tromino cram to a square matrix. The players take turns placing single counters until one player wins by getting three in a row orthogonally or diagonally. As before, odd-order fields are trivial because the first player wins by taking the center and then playing symmetrically until his opponent offers a win. On even-order boards the order 4 is trivial, but no one yet knows who has the win on orders 6 or 8. The bottom illustration on this page, supplied by Ulam, shows a position on the order-6 board for which the next player must lose.

Here again we can play an equivalent game by alternately placing polyominoes, in this case squares of nine cells, but it is not very convenient because in addition to allowing the pieces to extend into a unit border around the field we must also allow them to overlap one another by just two cells (corner and side). No one has even begun to find a general strategy for the game in standard or reverse form.

Last month's problem concerned nine girls. Each day for 12 days a triplet of girls goes out for a walk. How can the triplets be formed so that each pair of girls appears in exactly one triplet?

H. S. M. Coxeter, in the much revised 12th edition of W. W. Rouse Ball's *Mathematical Recreations and Essays* (University of Toronto Press, 1974), shows how the unique solution (known



Stanislaw Ulam's triplet game

as a Steiner triple system of order 9) is obtained from the following matrix:

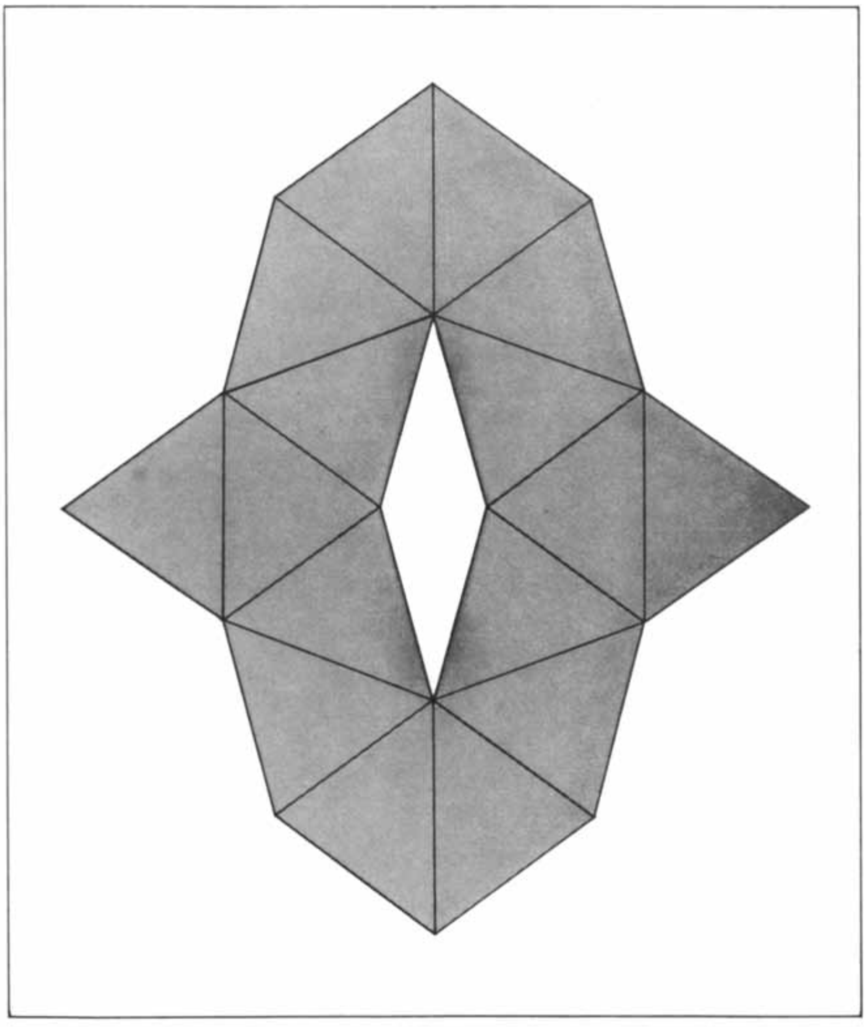
$$\begin{matrix} a & b & c \\ d & e & f \\ g & h & i \end{matrix}$$

The rows (*abc, def, ghi*) provide three triplets. The columns (*adg, beh, cfi*) provide three more. The main diagonals (*aei, ceg*) and the broken diagonals (*bfg, cdh, afh, bdi*) complete the list of 12 triplets.

The orders of Steiner triple systems are numbers that have a remainder of 1 or 3 when they are divided by 6. The next higher system, order 13, has two basic solutions. Order 15 is known to have 80 solutions.

A model of Kurt Schmucker's one-hole toroid with 48 congruent equilateral triangle faces, mentioned last month, is easy to make. It consists of a ring of eight regular octahedrons joined by their faces [*see illustration below*]. Schmucker found that rings could be made by joining eight replicas of each of the Platonic solids except the tetrahedron. No matter how many tetrahedrons are joined by their faces, no ring is possible even when the solids are allowed to intersect one another. A proof is given by J. H. Mason in his paper "Can Regular Tetrahedra Be Glued Together Face to Face to Form a Ring?" in *The Mathematical Gazette*, Volume 56, October, 1972, pages 194–197.

So many letters were received on the short problems in the March issue that I can report on only a few of them. Scores of readers solved the problem of the worm, some using the discrete model, others assuming continuous stretching of the rubber rope. It is assumed, naturally, that the worm is an ideal worm of point size and eternal life span. Some notion of the enormous length of the rope when the worm finally reaches the end can be gained from an observation by H. E. Rorschach. If the rope starts with a cross-sectional area of one square kilometer, it ends up as a single line of atoms, the space between each adjacent pair of atoms being many times the size of the known universe. The time lapse is comparably greater than the age of the universe. (The length of the rope was erroneously given in centimeters. It should have been kilometers.)



Ring of eight octahedrons

For playing the integer-choice game I suggested a spinner for randomizing selections from the set 1, 2, 3, 4, 5. Walter Stromquist proposed using a pair of dice as follows:

"After a few beers you cannot be expected to distinguish the fives and sixes, so that if either of these numbers appears on either die, you will have to roll again. Also, since you are really using only 2/3 of each die, it is only natural to multiply the total by 2/3 (dropping all fractions) before writing it down. For example, the largest number you can roll with two dice (without rolling again) is 8, so that the largest number you would ever choose is 2/3 of that, or 5. Out of 16 plays you should expect to choose with these frequencies: 1 once, 2 five times, 3 four times, 4 five times, 5 once." These are precisely the desired frequencies for playing the best strategy.

C. Stanley Ogilvy wrote to say he had included the three-circle problem in his book *Excursions in Geometry*. He too assumed that the proof by way of the three cones took care of all cases, but one day, after giving the problem to his class at Hamilton College, a student pointed out that the proof does not apply when a small sphere is between two larger ones. In such cases it is not possible for the two intersecting planes to be mutually tangent to all three spheres.

Many readers sent in other ways of proving the theorem. Richard I. Felver, Clyde E. Holvenstot, David B. Shaw and Radu Vero each proposed turning the drawing so that the line (on which the intersections of the tangent pairs lie) is horizontal and above the circles. The circles can now be viewed as being equal spheres inside three mutually intersecting pipes of identical circular cross section, seen in perspective. The tangent lines become the parallel sides of the three pipes. Since the pipes all must rest on a plane, their parallel sides seen in perspective will all have vanishing points on the horizon line.

It is not necessary that the circles be nonintersecting; indeed, the theorem can be stated in a more general way, in terms of "centers of similitude" instead of tangents, to hold for circles that lie entirely within one another. I am indebted to Donald Keeler for explaining this, as well as pointing out that the theorem is known as Monge's theorem after the French mathematician and friend of Napoleon, Gaspard Monge, who described it in a 1798 treatise. R. C. Archibald, writing in *American Mathematical Monthly*, Volume 22, 1915, page 65, traced the theorem (says Keeler) back to the ancient Greeks.

Daniel Sleator found that the theorem has an analogue with four spheres in space. Each of the four triplets will have the apexes of its three cones on a straight line. Because these four lines intersect one another at six points the four lines must be coplanar. Therefore the vertexes of the six cones all lie on a plane. Perhaps the theorem generalizes to all higher spaces.

The chess problem, to my surprise, has a second basic solution that was found by so many readers I cannot list all their names. Black opens with P-Q4, then follows with either P-KN3, P-KR4 or N-KB3. (Black's first two moves can be interchanged.) On his third move he checks with Q-Q3, then mates with Q-KB5.

The discussion of D. R. Kaprekar's self-numbers contained three errors: 101 is generated by 100 (not 101), 100,000 is generated by 99,959 (not 99,950) and the years of this century that are self-numbers begin with 1906 and 1917.

Donald R. Woods of Princeton University was the first to report on a computer program for John Harris' cube-rolling problem. It produced an unexpected result. The minimum-move solution is 36, two fewer than the one given last month. Moreover, aside from its reversal and rotations and reflections, it is unique:

URD LLD RRU LDL URD RUL
DLU URD RUL DRD LUL DRU

This was confirmed by the computer programs of John L. Couch, Robert Harris, Ken Jackman, John MacDonald, David Nixon, Jerome J. Sjostrand, John Sweeney and David Vanderschel. I shall report later if any readers found the 36 solution by hand or by other computer programs before its publication here. Harris writes that he has lowered the number of moves for his second problem (mentioned in April) from 84 to 74.

So many letters have been received on the six "April Fool" hoaxes that appeared in this department in April that it has been impossible to reply to any of them. In the next issue, however, I hope to comment in a general way on how the hoaxes were received, and to make some observations on the jokes themselves.

THE AMATEUR SCIENTIST

Diverse topics, from enhanced astronomical photographs to an ingenious electric motor

Conducted by C. L. Stong

Most amateurs who make photographs of astronomical objects use color film. Color emulsions record more information than black-and-white film does and cost only a little more. Not all amateurs who make such photographs extract the maximum information from their color slides, however. Subtle differences in hue can be easily overlooked. John Sanford (2215 Martha Avenue, Orange, Calif. 92667) has hit on a photographic technique for enhancing those subtleties. He makes black-and-white prints of the color transparencies with both ordinary and panchromatic projection papers exposed through certain color filters. Differences in color that would ordinarily escape detection in a color print stand out as patterns of strong contrast in the black-and-white prints.

"A great deal of information frequently escapes detection in color slides made of the night sky," Sanford writes. "Much of it, however, can be recovered in the darkroom with black-and-white photographic paper and a few filters. Essentially a black-and-white projection print is made of the color slide. The resulting print is a negative image of the slide.

"The slide is placed in the enlarger to expose the black-and-white paper to make a paper negative on which stars appear as dark spots against a white background. Many laymen have seen photographs of this type; typical are the prints from the Palomar Sky Survey that appear occasionally in magazines. Professional astronomers prefer reproductions of this kind because of the ease of marking objects (for duplicating the photographs with office copiers) and because faint detail on the original plate appears on the negative print as distinct regions of gray or black against the contrasting white background of the sky.

"Certain slide films have a high sensitivity to red light. They are excellently suited for recording the reddish light from H II (ionized hydrogen) regions of the galaxy, where one finds most of the well-known diffuse or galactic nebulas including M42 (the Great Nebula in Orion), M8 (the Lagoon Nebula) and NGC 7000 (the North American Nebula). My color slides have been made with Fujichrome 100, which is available in most camera stores. This film responds well during processing to 'pushing,' which is controlled overdevelopment to compensate for underexposure. The film can be developed with the standard E-4 process, as Ektachrome is, either by a commercial processor or by the photographer himself with a kit. My films are processed by a commercial laboratory. I usually specify a one-stop push to ASA 200. A 10-minute exposure with Fujichrome 100 is equivalent in my experience to a 30-minute exposure with High Speed Ektachrome, assuming that both films are pushed one stop during processing.

"The projection prints are made on either regular or panchromatic paper, depending on whether the observer wants to enhance the blue parts of the spectrum or the red. Blue hues can be enhanced by making the print with regular enlarging paper, which is sensitive to blue-green light. This selectivity explains why the darkroom can be illuminated with a yellowish safelight. Regular enlarging paper is even less sensitive to red light.

"Blue hues can also be enhanced by exposing panchromatic projection paper through a blue filter, such as a Wratten 38A. This filter is available inexpensively in the form of gelatin film at Eastman Kodak stores. Expose the print just to the point at which the background begins to take on a perceptible tone of gray. That exposure will maximize the detail of nebulas. The time of exposure may seem overlong to anyone accustomed to making enlargements of conventional black-and-white photographs. Be patient. To enhance reddish detail use panchromatic projection paper, such as Eastman's Panalure, and expose it through a Wratten 23 or 25 filter.

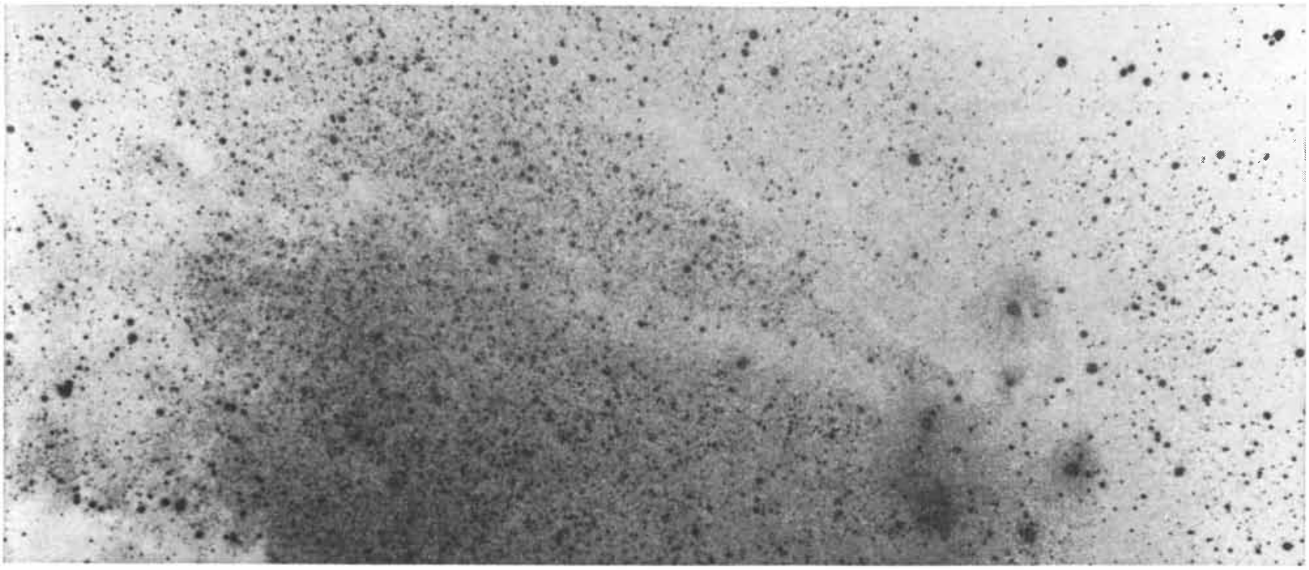
"A photograph of the center of Scorpius made by a 20-minute exposure of Fujichrome 100 color film with a 105-millimeter Nikkor lens at $f/3$ apparently failed to disclose differences in the density of dust lanes that characterize this region of space. The lanes do show clearly, however, in the enhanced negative print of the slide [*see top illustration on opposite page*]. A similar enhancement of blue hues disclosed many details of the North American Nebula that are barely perceptible to the eye when the color slide is examined [*see bottom illustration on opposite page*]."

Casual users of reflecting telescopes rarely fit their instrument with a setting circle or any other elaborate device for centering the eyepiece on an object. Instead they take aim with a gunsight or with crossed wires in a short length of tubing. These expedients require the observer to cock his head at awkward angles when he wants to scan certain parts of the sky. A more efficient and convenient accessory, an elbow finder telescope, has been improvised by Harry F. Kale (1800 Galaxy Drive, Newport Beach, Calif. 92660). He describes the device as follows.

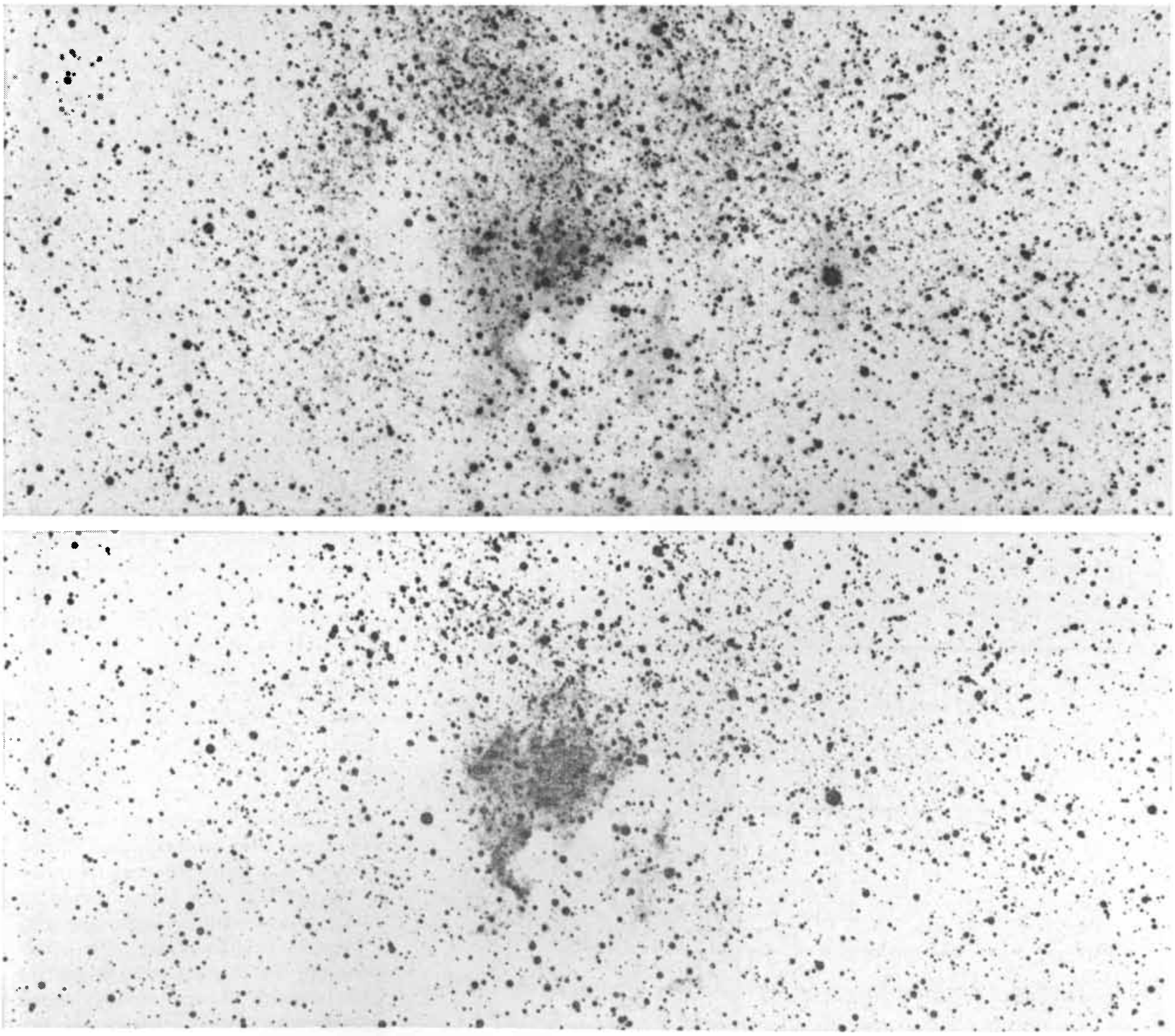
"My finder consists of an eight-inch right-angle tube three inches in diameter. Between thin metal rings at its ends it supports a pair of three-inch magnifying glasses of four-inch focal length. I bought the magnifying glasses for $1 each.

"Two tin cans from which the ends have been removed can be soldered together to form the tube. Before I assemble the tube one can is cut in half at an angle of 45 degrees, rotated and resoldered to make a right-angle elbow. The elbow section is then soldered to the other can.

"A cylindrical block of wood cut to form a 90-degree prism was cemented inside the elbow to support a front-surface mirror at an angle of 45 degrees [*see illustration on page 114*]. Two lengths of No. 18 stovepipe wire were



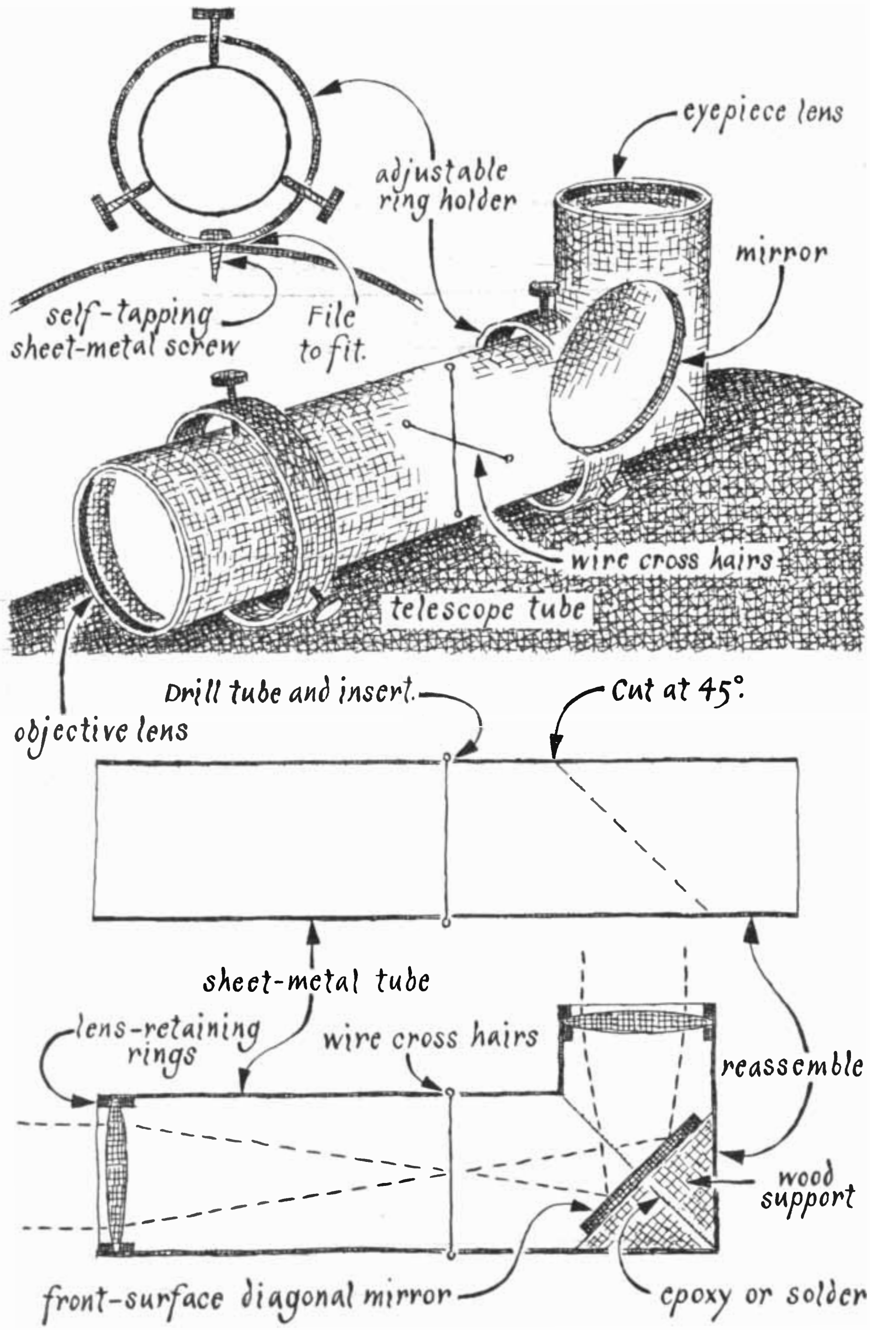
John Sanford's enhanced negative print of the center of Scorpius



The North American Nebula without enhancement (top) *and as enhanced* (bottom) threaded at right angles through the tube four inches from the magnifying glass that functions as the objective lens. In this position the crossed wires are sharply focused by the lens that functions as the eyepiece.

"A front-surface mirror is essential as the diagonal element for bending the light rays to avoid multiple reflections and hence multiple star images. I bought the mirror for $3 from a mail-order house that sells optical supplies. I reshaped the mirror with a grinding wheel to fit inside the elbow. The mirror was fixed to the supporting prism by silicone-rubber cement.

"Support rings for adjustably mounting the finder telescope to the main telescope were cut from a discarded aluminum ashtray. Three equally spaced holes were drilled in each ring and threaded for radial adjustment screws that clamp the finder. A single bolt attaches each ring to the tube of the main telescope at positions where the optical axes of the two instruments are approximately parallel. Final alignment can be made during the daylight hours by first centering a distant object in the eyepiece of the main telescope and then centering the cross hairs of the finder on the same object by manipulating the adjustment screws of the aluminum rings. Although the finder does not magnify the star field, the light-gathering power of the three-inch objective lens significantly increases the apparent brightness of all celestial objects. The cross hairs can be seen clearly in silhouette against the background of stars, and the elbow scheme greatly increases the comfort of examining the night sky."



Harry F. Kale's finder for a telescope

D. G. Prinz (37 Parkville Road, Manchester M20 9TX, England) recently noticed that the reflected image of a drafting instrument made of clear plastic appeared highly colored in a shaft of sunlight that had bounced off the painted side of a filing cabinet. The observation started Prinz, who is a computer consultant, on an investigation that led him to the report of a similar observation made late in the 18th century by Étienne Malus, an engineer in the French army. Malus had seen the same colors when he looked through a calcite crystal at light reflected from the windows of the Luxembourg palace. He explained the effect as the consequence of polarization, that is, the rotation of the direction in space in which light waves of differing color vibrate.

The same effect had subsequently attracted the interest of the Scottish physicist Sir David Brewster. Brewster found that maximum polarization occurs in light reflected from a substance at an angle with a trigonometric tangent numerically equal to the index of refraction of the substance. For example, light reflected at an angle of 53 degrees 7 minutes 46 seconds from water is more strongly polarized than light reflected at greater or lesser angles. The trigonometric tangent of this angle is 1.3333, which is also the index of refraction of water at the temperature of 20 degrees Celsius. (The index of refraction expresses the ratio of the speed of light in a vacuum to its speed in water.)

The index of refraction of most ordinary plastics is approximately 1.45. Prinz reasoned that light reflected from the polished surface of a plastic should be strongly polarized at the angle (about 55 degrees 24 minutes of arc) that corresponds to the trigonometric tangent of 1.45. To check his conclusion he hinged together with flexible adhesive tape two sheets of polished black plastic, much as

On Father's Day, think well of the man who always thinks of you.

When the thought is genuine the Scotch should be too.

Dewar's never varies.

DEWAR'S®
"White Label"®

BLENDED SCOTCH WHISKY • 86.8 PROOF • ©SCHENLEY IMPORTS CO., N. Y., N. Y.

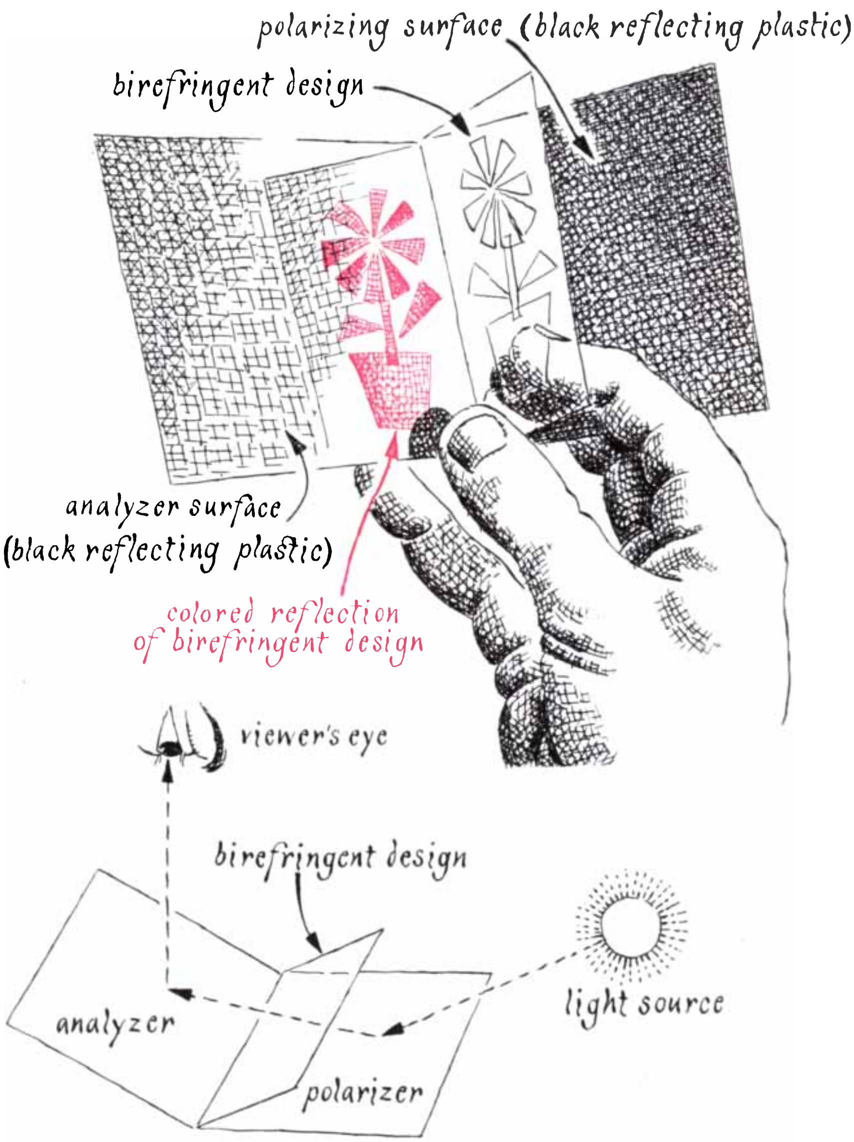
the covers of a book are joined. He laid the assembly in sunlight on a windowsill with the hinge facing into the room. When he raised the upper sheet so that it made an angle of about 110 degrees with its mate, its inner surface was flooded with sunlight reflected by the lower sheet.

Prinz hinged between the sheets of plastic a few pieces of transparent cellophane. He held them so that they would intercept the polarized rays. The cellophane displayed dazzling colors.

Cellophane, which consists of reconstituted cellulose, has the property of birefringence. It splits incident light into two components that propagate at different speeds, which vary with the length of the light waves. As a result some waves of polarized light interfere destructively in birefringent substances, depending on the thickness and the orientation of the substances. Such waves are not seen by the observer.

The remaining light appears as the complementary color. The hue that is seen depends on the thickness of the cellophane and its orientation with respect to the plane in which the polarized light vibrates. Prinz kept a record of the number of sheets of cellophane and the orientation required to generate each color. The data enabled him to employ the cellophane to create full-color pictures that resemble miniature stained-glass windows.

The polarizing sheets need not be made of polished black plastic. Prinz achieves the same effect by blackening white business cards with India ink and, when the ink dries, making them reflective by covering the surface with shiny transparent adhesive tape. A coating of quick-drying acrylic varnish can be substituted for the tape. Images reflected from such surfaces may appear grossly distorted because strips of plastic tape on cardboard do not make an optically true mirror. Nonetheless, vivid colors appear. The blackened material merely improves the contrast in the image.



The simple polariscope devised by D. G. Prinz

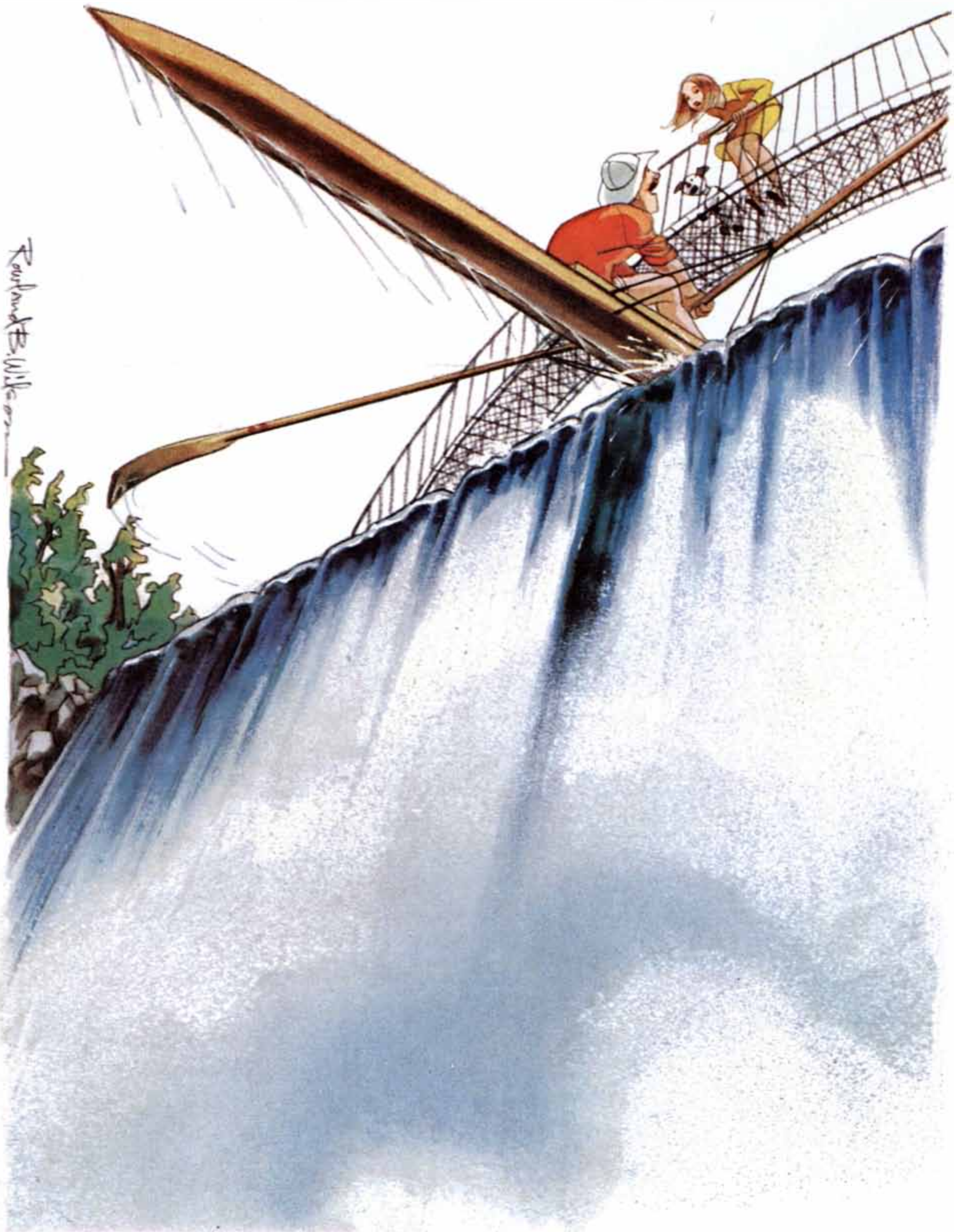
Frederick S. Duncan (P.O. Box 212, Pima, Ariz. 85543) reports a dramatic experience that in 1969 introduced him to a wildly esoteric field of art. Duncan is a systems engineer and was stationed at Thule in Greenland at the time. "For two years in Thule," he writes, "I whiled away the long winter nights by trying, among other things, to find all the solutions to the tangency problem of Apollonius of Perga, that is, to draw a fourth circle so that it just touches each of three given circles. Another of my diversions was experimenting in a similar way with a Steiner chain. In this chain each circle of a finite number of circles must touch two other circles in the series as well as two that are not part of the series.

"Once I happened to rotate one of my sketches about an axis perpendicular to the paper, hoping thereby to find a point of view that would clarify the problem. Suddenly I became aware of an effect that made me feel like an inhabitant of Flatland viewing his village and its inhabitants for the first time from a balloon. Some of the circles rose out of my sketch and floated in the air! Others sank below the surface. Their positions interchanged again and again as I watched. I could only conclude that my senses were reflecting our overlong stay in Greenland.

"I later learned that the effects had been investigated in a preliminary way half a century ago by two Italian psychologists, V. Benussi and C. L. Musatti. Benussi pointed out that when a circle is drawn as a line of constant thickness on a disk, it will appear motionless if the disk is rotated about a center that coincides with the center of the circle. Displacing the center of the circle from the center of rotation creates the illusion that the circle is rising above or sinking below the surface of the disk on which it actually rests. The effect can be enhanced by closing one eye.

"Replacing the circle with an ellipse further confounds the eye. The pattern may revert to a circle that gyrates wildly in three dimensions. It may warp and stretch with an amoeboid motion. These phenomena are known as Musatti effects.

"After my return to the U.S. in 1971



"My insurance company? New England Life, of course. Why?"

Our crew also offers mutual funds, variable annuities and investment counseling.

I made a number of designs in color: bright reds, whites, blues, greens and blacks. The patterns were drawn on heavy disks of cardboard 18 inches in diameter. The disks can be rotated by a variable-speed turntable at rates between two and 30 revolutions per minute. I regard the designs [*see illustration below*] as a kind of kinetic 'op' art that gives aesthetic satisfaction to the viewer and that may have interest for those who investigate the mechanisms of visual perception. The patterns are not difficult to draw. An adequate turntable can be improvised from the mechanism of a discarded phonograph. For people who prefer just to look, kits of 12-inch disks in color complete with a motor are available from Research Media, Inc. (4 Midland Avenue, Hicksville, N.Y. 11801)."

During a mathematical investigation of electromagnetic forces that arise locally in various electric circuits, Harry E. Stockman (75 Gray Street, Arlington, Mass. 02174) has discovered a number of novel motors. Some of them, such as the tunnel-diode motor and the parametric motor, have been described in these columns [see "The Amateur Scientist," October, 1965, and January, 1973]. Stockman's most recent construction is a variable-speed motor that operates on alternating current. Unlike most

Art forms generated by Frederick S. Duncan

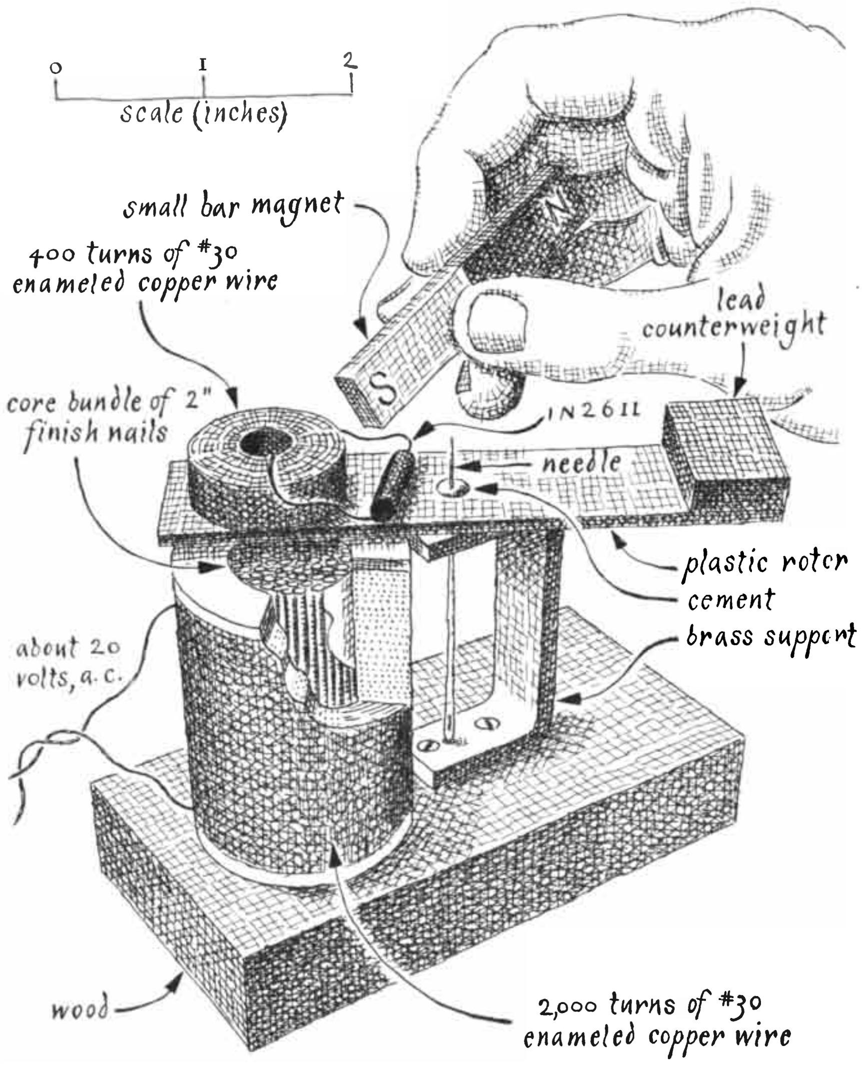
variable-speed motors, his version requires no commutator or brushes. Stockman describes the motor.

"A demonstration model of the asynchronous motor includes a fixed electromagnet energized by a potential of about 20 volts at a frequency of 60 hertz. The speed of the motor can be altered by varying the applied potential. The moving element consists of a nonmagnetic rotor that supports at one end a coil of insulated wire connected to a silicon diode. The rotor is balanced mechanically by a counterweight and turns on a vertical shaft, as shown in the accompanying illustration [*below*].

"A bar magnet is supported near the electromagnet. As the small coil approaches the electromagnet it cuts through the field of the magnet at approximately right angles. The design is not critical. All dimensions can be altered.

"An alternating potential applied to the electromagnet causes a force of magnetic repulsion to act on the small coil. Simultaneously the alternating potential induces a unidirectional current in the small coil connected to the rectifying diode. The resulting unidirectional magnetic field of the small coil can interact with the field of a bar magnet to minimize the repulsive force exerted on the coil by the electromagnet.

"By experiment one can find a position of the bar magnet that enables the small coil to approach the electromagnet and move freely into alignment with it. At this position, however, the coil is violently repelled. As the coil recedes from the electromagnet it acquires angular momentum that carries it into the field of the bar magnet in preparation for the next revolution. The speed of rotation is a function of the forces and hence of the applied potential."



Harry E. Stockman's asynchronous motor

SCIENCE/SCOPE

Indonesia's domestic communications satellite system -- first in the Eastern Hemisphere -- will be built by Hughes. It will include two satellites like those now in service for Telesat Canada and Western Union, a master control station, and nine earth stations (30 additional earth stations will be built by other contractors). They will link the 5,000-island republic with telephone, telegraph, television, and teletype service. Future plans include a national radio network.

Oil and mineral exploration crews in the Canadian wilderness can now have immediate communications with company offices and families by telephone or teletype via Telesat Canada's Anik satellites and a compact portable terminal developed by Hughes. The terminal can be erected in four hours and operates in temperatures as low as -70°F. Telesat Canada has leased five of the new mini terminals.

The National Weather Service is now testing a prototype of AFOS (Advanced Field Operating System), its proposed $40-million all-electronic weather reporting network. Key elements of AFOS are the on-site minicomputers and TV-type displays that will replace teletypewriter and facsimile equipment. The displays feature the Hughes Conographic™ terminal which, because of its unique ability to convert contour data to conic curves, requires significantly less data than conventional x-y plotting systems. This results in faster transmission and greater capacity for the network, lower storage requirements for the terminals.

Weather maps will be transmitted 20 times faster, printed matter 30 times faster than by present methods. The increased speed and capacity of AFOS will be particularly valuable for warnings of tornadoes, hurricanes, and floods. The Weather Service hopes to have about 275 of its offices automated by 1980.

R&D project leader needed. Responsibilities will include advanced device development, customer interface, presentations to top management, and establishing manufacturability. Must have PhD in solid-state physics or electrical engineering and five years of developmental work in MOS, CMOS, bipolar, and CCDs, with recent experience in silicon devices and integrated circuits. U.S. citizenship required. Please send resume and salary history to: P.A. Schneider, Hughes Aircraft Company, 500 Superior Avenue, Newport Beach, CA 92663. An equal opportunity M/F employer.

A new engineering data reduction service offered by Hughes features fast problem resolution and quick turnaround times. Inputs include real-time data from radio or telephone, analog or digital magnetic tapes, film, strip charts, punched cards, and paper tape. Outputs include quick-look form, published presentations, and computer tapes. For information, please call or write Bob Margolies, 213/391-0711, Ext. 7988, Hughes Aircraft Company, Flight Test Division, Culver City, CA 90230.

Creating a new world with electronics

HUGHES

HUGHES AIRCRAFT COMPANY

BOOKS

Illusion: a clue to how the visual system interprets the stochastic rain of photons

by Philip Morrison

ILLUSION IN NATURE AND ART, edited by R. L. Gregory and E. H. Gombrich. Charles Scribner's Sons ($19.95). CONCEPTS AND MECHANISMS OF PERCEPTION, by R. L. Gregory. Charles Scribner's Sons ($40). VISION: HUMAN AND ELECTRONIC, by Albert Rose. Plenum Press ($15). The illusory triangle glows whiter than white against the blank page; it is not there at all, yet you see it! The two huge Hercules beetles stand on the color plate full-sized; one almost matches the yellow banana it crawls on, the other is a sober beetle black. Picasso's *Bull's Head* is too convincing as sculpture to see as the bicycle seat and handlebars it is. From the initial essay, a clear brief account of the bases of the neurophysiology of vision, to the final one, a clarifying chronicle of the illusory devices of the painters of our time from Braque to Vasarely, *Illusion in Nature and Art* is plain pleasure.

Coeditor Gregory's own essay makes once again his compelling case that the act of perception resembles less the random snapshot than the careful experiment. It is active, it relies on some inner hypothesis and it proceeds relatively slowly, drawing unconsciously on rules of inference both innate and learned.

Gregory's Bristol colleague, zoologist H. E. Hinton, presents next a very fresh account of plant and animal mimicry. His essay—the longest of the six in the book—vigorously defends "what was considered to be one of the most extravagant theories of the old-fashioned biologists": the real selective advantage of such ingenious tricks. The naturalists were right, because man and the chief vertebrate predators see in similar bands of wavelengths. "I have studied insects for some forty years, and I am of the unshakable opinion that if a beetle can deceive me into thinking that it is a wasp, it will certainly also deceive a lizard or bird." Certain caterpillars of distinct families in many parts of the world attach up to 40 little empty cocoons to the silk surface of their own larger pupal housings. Why? Because some parasitic larvae emerge after consuming the pupal host and spin just such little outer cocoons of their own; birds avoid the parasitized cocoons because there is nothing left in them but the pupal skin, and the individually packaged parasites on the outside are too small to be worth the birds' trouble. Certain chrysalids take a form that closely resembles the face of a monkey; an Indian butterfly pupa imitates the locally common rhesus and an African one mimics a common African monkey. The pupa is measured in millimeters whereas the monkey face covers as many inches, and so the resemblance has been held a matter of mere chance. Not so: many birds utilize binocular cues very little, and their rapid peering is subject to distance confusion once they get a sudden close-up view of the illusory monkey. There are butterflies that flap as flashily in ultraviolet as the blue Morphos do to our dazzled eyes. There are beetles with fine diffraction gratings that blaze out warning colors during their rare forays into direct sunlight and yet allow the beetles to remain drab in the dim, diffuse light of their usual grass-roots habitat. Visibly bright flowers evolved, we were told, to attract the birds; the fact is they were red and yellow much earlier—to dissuade the lizards, warning them (either honestly or deceptively) away from many flowers noxious to eat. It was the oldest use of such bright color in nature. The birds came later; they were first attracted to flowers by the presence of insects and then gradually reversed their antipathy to the bright warning colors, as some birds have also done for well-marked bees and wasps. The flowers have solved an ancient dilemma of design: they had to attract pollinators and at the same time repel undesirable visitors.

Another essay treats of the cultural context of illusion. The author, a psychologist from the University of Aberdeen, reviews cross-cultural studies of the impact of optical illusion, particularly the illusion of depth in pictorial representation, which is perhaps visual illusion's most general form. The data are still thin, but the conflicting evidence is most interesting.

Concepts and Mechanisms of Perception presents us with the selected papers of Gregory himself, papers uniting philosophy, electronics and experimental psychology and extending over 20 years. His writing is so lively, his thought so richly connected to the hardware experience of engineer and physicist while not overlooking the software subtleties of meaning and language, that it is next to impossible to skip any paper, however limited or special its topic. Here is the rare case of a man who recovered vision at 52 through a corneal graft after having been fully blind since infancy. He had been a curious, intelligent and well-trained blind man, a Braille reader, with "a long-standing interest in tools and machinery." He saw no depth in perspective, nor much effect in the usual optical illusions. There is evidence suggestive of some true cross-modal transfer from touch to vision. The subject could immediately read uppercase letters, which he had learned by touch, but he never learned lowercase ones. At a screw-cutting lathe he complained that he could not see the cutting edge or the metal being worked, and he "appeared rather agitated." Then he was allowed to touch the lathe. The result was startling: "Now that I've felt it I can see." He died within two years, in a psychological crisis. He had borne a great handicap with energy and enthusiasm, but once his handicap was suddenly removed "he lost his peace and his self-respect." Gregory continues: "We may feel disappointed at a private dream come true; S. B. found disappointment with what he took to be reality."

There is a series of papers on the judgment of weight differences. The level of discrimination of weight is optimal with objects of normal density, but when objects are packaged to appear lighter or heavier than they actually are, illusion makes discrimination of weight harder. Processes of the central nervous system must be involved—how otherwise can the look of a weight affect the arm muscles? It is suggested that the "software" within sets up some central standard and the rest is comparison. The analogy of a bridge circuit is tempting: a bridge has its maximum sensitivity with equal arms. Some clever devices are described. There is a solid-image microscope that scans in depth and projects onto a synchronously scanned screen, so that you see not a fly entombed in amber but a living fly in a block of light. (The instrument has never been marketed, but it works.) There are nine or 10 others: color photometers, telescopes that defeat the "seeing" disturbances and still others, a few of them more ingenious than promising. The entire corpus is engaging, and it illuminates the mind of a man who writes philosophical papers full of engineering metaphor.

One stance seems wrong. Gregory thinks very little of the localization of brain function as determined by removing or making a lesion in some portion of the structure. He tries to compare the classic lore on brain function with the "removal of any of several widely spaced resistors." That might cause a radio set to howl, but the resistors are not to be regarded as howl-inhibitors. Gregory fails to show us engineering examples as subtle as the lesions that destroy, say, speech but leave writing unimpaired. One suspects that it is not a radio that is the analogy but a much more complex machine with differentiated subsystems, more like a computer and its peripherals.

Such engineering is merely human; the living world has evolved more remarkable designs. They are not neglected. Here are an illustration and description of a copepod, *Copilia quadrata,* that Gregory and his co-workers fished out of the Bay of Naples. It had been described (by the zoologist Exner in 1891) as having an eye in "continual and lively motion." Gregory writes; "We had no drawings to go on, ... and so we did not know just what to expect. We examined a gallon of water a day, drop by drop, with three microscopes until, though fascinated by what we did see, we began to despair of finding *Copilia.* Then suddenly there she was! Incredibly beautiful: perfectly transparent—so no veil hid the secrets of her eyes." The little creature is about five millimeters overall. There are two fixed eye lenses far forward on the head. Well behind each one, clearly seen deep within the transparent head, lies another lens, attached to an orange filament that contains the photosensitive element. The optic nerve leads plainly out of the orange fiber into the ganglia. The inner lenses continually approach each other and recede, moving laterally several diameters in perfect synchrony up to five times per second, in a sawtooth time pattern. Surely those eyes are scanning, reading out what lies in the image plane of the forward lenses.

Speculate. Multiply the scanning channels to provide more bandwidth. Then let the now compound eye turn static; with enough facets it needs no scan. In that step you will have recapitulated the fundamental sensory relations of touch. Passive touch means many neural channels bringing information in parallel, and it maps a two-dimensional world pattern sensed by the skin. Its complement, active touch, is the single-channel pattern of an exploring tentacle, which codes in time its news of a three-dimensional world. Vision took over the central processors from touch: first the simple eye, using the passive touch system; next the scanning eye; finally the compound eye, taking over the time-dependent system. If we are lucky, in science as in the stock market, says Gregory, speculation can pay off. "The developed brain entertains many possibilities, only some of which are true." One can be sure of this: the book is full of tentative models, each one a stimulus to thought and a gift of novelty.

Vision: Human and Electronic, by Albert Rose, a man who for a generation has given much to the modern television system and its photon-handling kin, is a smaller text. Rose deals in a clear and deft way with the first illusion of all visual systems: the appearance of a continuous and coherent world out of a stochastic rain of photons and their atomic interactions. He makes plain how the signal arises out of the photon rain, ruled by the laws of large numbers. All useful systems—eye, photographic film and various television cameras—have evolved to extract the most information possible from the photon flux. The book essentially makes clear how well the three systems compare in photon-counting, and how they do it. "In brief, the eye is able to count each absorbed photon." That implies an efficiency of between 5 and 10 percent on incoming quanta, allowing for retinal absorption and for losses before the quanta reach the retina. The performance is maintained within a factor of perhaps three over a range of 10^8 in light intensity, from threshold up. Noise analysis shows that quantum efficiency does not change much in dark adaptation; the big changes by three orders of magnitude must be in amplifier gain. Television camera tubes match the eye; indeed, the newer solid-state self-scanned arrays, which use quite a different method of reading out the electron-charge pattern induced by the light, promise an efficiency approaching one count for every photon. Photographic film shows a quantum efficiency of about 1 percent, with a chemical "gain" on development of some 10^9. It has been tied tightly to silver bromide—a century of ingenious and successful improvement involving a single compound. Image multipliers, both vacuum and solid-state, offer further opportunities. The text here does not try to survey the full state of this art; its references are five years old or more except in the domain of the author's own laboratory colleagues, the self-scanned solid-state image sensors.

There is a real need to exceed the eye in performance, not only for astronomical purposes but also for the commonplace needs of marketed illusion. The eye refocuses constantly; we scan the world piece by piece; what we look at is always in focus. The television camera must give us the entire scene in focus at once, however, because it cannot know where our attention is fixed. Moreover, the scene is presented at a level of illumination that can be higher than the level at the time of recording, so that the viewer demands less noise to match the perceived brighter scene.

From photon rain through Plato to perspective depth in the Florentine painters these books—nowhere complete—describe a still-wild realm of human understanding, fascinating for the familiarity of the subject, the subtlety of its structure and the catholicity of its interest. Here we see what best we mean by the grand old phrase natural philosophy.

CLIMATE: PRESENT, PAST AND FUTURE. VOLUME I: FUNDAMENTALS AND CLIMATE NOW, by H. H. Lamb. Methuen & Co Ltd, distributed by Barnes & Noble Books ($37.50). THE WEATHER MACHINE, by Nigel Calder. The Viking Press ($14.95). Hubert Lamb, a scholarly old party, is shown in color at a handsome desk with an elegant aneroid as he "consults a warship's weather log of 150 years ago." He is in a way the hero of these British books. The Calder book is a gossipy, up-to-date, handsomely illustrated and entirely popular account that flows from a BBC-produced television program first screened last fall. In Britain the book had the scarier subtitle still carried by its final chapter, "The Threat of Ice." Lamb's clear and graceful book, a heavy volume that summarizes a distinguished career, is rich with quantitative graphs, data and tables: sunspots, worldwide weather maps, power spectra of such diverse time series as rainfall and Lapland tree rings. It seeks to outline the technical, physical and historical bases of the study of long-range climate change. (The book was published in 1972; a second volume is promised, focusing on the past and the future.)

The weather is of course the work of a big heat engine: sun-fed, sea-coupled, mountain-deflected, gravity- and spin-controlled. Calder attends also to the people who study that engine with means far from the scholar's desk: satellite photographs, the mesh on the world map that marks where the British Meteorological Office daily computes its numerical forecast (3,000 points, each at 10 altitudes), the drill crew high on the Greenland ice cap taking the deep core to sample the snows of autumns long past. It is not easy to weigh judiciously the evidence of old manuscript against the solution of the simultaneous difference equations of modern hydrodynamics by computer. The numerical modelers "doubt the rather vague meteorological arguments of the investigators of past climates," whereas those workers in turn cannot see the computers ever turning up the "slow and subtle changes" for which the past bears sure witness. Such divisions of the discipline have "become scandalous in Britain," where it appears that Lamb cannot find government funds with which to carry on his work.

The issue is of course tangled. The blue-and-white box of the sky is complicated enough to put out many surprises, even if one has the right input. Error enters with the data, however, and amplification can be so strong that we cannot be certain that the flutter of a moth will never grow into a hurricane. Worse yet, for the long run—and possibly even with earlier impact—we cannot specify all the appropriate inputs.

Lamb has made the best study of one of the most striking unexpected external effects: the dust veil thrown up in many large volcanic eruptions. The dust scatters the incoming shortwave solar radiation much more than it does the outgoing infrared, and so the surface temperature goes down a little, the effect persisting for several years. The notorious Krakatoa explosion of 1883 serves as a unit, but comparable effects have occurred some five times per century since A.D. 1500. Indeed, the largest dust veil we know of was thrown up in 1835 from the volcano Cosigüina in Nicaragua. The red sunsets of the mid-1960's from an eruption in Bali pointed to a dust veil only a fifth smaller than Krakatoa's. Yet it is impossible to ascribe enduring changes to such a sporadic source. The man-made volcano of plows and furnaces may latterly have become important too, with carbon dioxide effects acting as well. And thermonuclear war would multiply Krakatoa a thousandfold. The sun, with its dim spots and bright faculae, must be considered. The changes in sunspots alone, a popular proposal, will not do; any deeper long-range changes are simply unknown. The orbit of the earth undergoes small but important drifts over hundreds of millenniums; the proposal that these are cumulatively adequate to cause the great shifts of ice is widely believed today. Indeed, Calder himself graphs a version of this theory to compare with the oxygen-isotope composition of the shells of marine bottom dwellers. This last surprising quantity appears to correlate well with the mass of water locked up in ice. Glacier ice is low in heavy oxygen compared with seawater. Organisms of the sea bottom, which grow in water of relatively constant temperature, show an oxygen-18 content whose small changes seem mainly to reflect the store of that more sluggish atom, which preferentially remains out of water vapor and hence falls less as snow. The fit to the orbital changes is suggestive but less than compelling.

By now we can see in the cores the mark of eight or more maximums of the ice during the past 750,000 years. The ice today, mainly in the South Polar cap, has about half the mass of the ice found at a glacial maximum. In the last maximum there was a North Polar cap across Canada, Scandinavia and northwestern Siberia that was about equal to the present Antarctic ice cap in volume. The sea dropped more than 300 feet below the present level. On the shallow coastal fringes there were glaciers; the London geologist Geoffrey Kellaway, Calder reports, has shown that the ice spread up the English Channel from the west, a small part of the great ice sheet that covered the whole of the British Isles and their sea approaches. (A small stream of ice may have carried heavy bluestone boulders to Salisbury Plain, where long afterward the builders of Stonehenge would make use of them—erratic glacial gifts, not the wares of a heroic Neolithic stone trade!) Such lowland ice fits the new proposal by Lamb and his co-workers that ice can form not only on mountain slopes, to ooze slowly onto the plains, but also in the lowlands. One year snowflakes invade flat ground. They cover entire counties in a single winter; that they "have come to stay" does not appear until the next summer. Then the snow on the ground locally reflects the summer sun, chills the air and guarantees more snow next winter. Such was Fimbulvetr, three winters on end, that would herald the twilight of the Norse gods. On this view territory can be seized by the glacier in a few seasons, or perhaps over a hundred, not over the long millenniums during which northern highland ice flows southward. Late last year Nicholas Shackelton of the University of Cambridge found that it took only 5,000 years for the ice to build up from a minimum to a half-maximum about 110,000 years ago. It appears that ice eras have been more rapid and more frequent than we thought; our moderate climate is found only a tenth of the time during the past million years. Yet something holds final control: the planet is now ice-free and again somewhat icy, but it has never been superheated like Venus or entirely frozen, a ball of ice.

It is plain now that there was a "little ice age." From about the time of the Black Death to that of Charles Dickens the winters of Europe were generally cold and snowy and the summers cool. How often they skated on the frozen Thames! Since then it has been warmer, although perhaps it has turned cooler again in the past 20 years. We do not know what will come next. We can be sure, however, that there is no simple solution. Human beings are linked to uncertain climate by complicated chains. Nor can we expect easy control. The typhoons bring havoc, but they also bring needed rain—a fourth of all Japanese rainfall. Our species evolved in the time of ice. We shall encounter a test—whether in this decade or a few millenniums hence we cannot say, but it will come—of whether all we have earned and done since we left off hunting the mammoth has made our kind more or less fit to survive the coming of the ice.

HANDBOOK ON HUMAN NUTRITIONAL REQUIREMENTS, by R. Passmore, D. L. Bocobo, B. M. Nicol and M. Narayana Rao in collaboration with G. H. Beaton and E. M. DeMaeyer. Monograph No. 61, World Health Organization ($4.80). RECOMMENDED DIETARY ALLOWANCES. National Academy of Sciences ($2.50). We lay waste our powers at quite different rates. Getting and spending are not tabulated, but the numbers stand. At very light work (laboratory work, typing or fiddling) men use up to 2.5 kilocalories per minute; tree felling and basketball demand some eight to 12 kilocalories per minute; in comparison indolent sleep takes about one kilocalorie per minute. (The rates reckoned for the "58-kilogram standard woman" run about 25 percent less.) That is the kind of information that crowds the pages of these two quasi-official brochures, whose common intent it is to explain and justify the closing table in each book, where for many conditions of man (by age, sex and weight) our daily bread is quantified.

The WHO book is the simpler; half the length, it is considerably less medical in language and style. The U.S. book contains 25 pages of up-to-date references categorized by topic, whereas the UN publication makes do with only a dozen footnoted citations. Both are aimed not at physicians but at planners, teachers and working nutritionists. The WHO authors have an international frame of reference; catholicity gives a tone to their work that our national example of course lacks.

Everyone knows about vitamin A, now known as retinol, and its colorful precursors, particularly beta-carotene. Within the past few years the weight of the pure crystalline alcohol, the vitamin itself, has become the standard of measure, supplanting the older biological assay quoted in arbitrary units. An adult needs 750 micrograms (the carotene yields about a sixth of its intake weight as retinol) per day according to WHO standards, or 1,000 micrograms as the U.S. experts see it. Such variations are to be expected; the recommendations include safety factors judged to allow for individual variations. Overall consistency is nevertheless marked. Experiments on large groups made over decades showed that about 400 micrograms per day maintains night vision; curing the skin lesions produced by deficiency requires about 600 micrograms per day. The rest is for good measure. The Geneva authors go on: "One of the worst scandals of our time is that some twenty thousand young children go permanently blind every year because of the lack of minute amounts of retinol in their diets." Epithelial cells flatten and dry out for lack of retinol; malnourished children often suffer from a form of conjunctivitis. If the condition spreads to the cornea, an opening can appear in that transparent layer; the iris and even the lens protrude through the gap and blindness follows. "Each of these tragedies could have been prevented by a little knowledge and timely care." Cereals and vegetable oils are generally poor in carotene—except for red-palm oil, which is very rich. The savory palm-oil chops of West Africa, deep orange next to the white starchy staples, once more display the adaptive values of classical cuisines.

The protein requirement shows good agreement between the WHO and the NAS. It was not always so: in 1968 the world experts set a protein requirement that was distinctly ungenerous. Bad data on nitrogen loss under a protein-free diet lay behind that result; there was both inattention to such minor losses as those in sweat, hair and nails and an assumption of full utilization of food nitrogen. Now there is agreement on a maintenance protein intake of .47 gram per kilogram of body weight per day as an average, assuming the amino acid composition to be about the optimum: that of cow's milk or egg protein. The amount is increased to accommodate those with above-average need (about 20 percent by the WHO experts and 30 percent by the U.S. experts). The requirement is further increased in the NAS tables to allow for the typical mixed-protein intake of the U.S. diet, to end as .8 gram per kilogram of body weight per day. The famous mixed traditional dishes—black beans and rice, tortillas and frijoles, pasta and fagioli—represent schemes empirically evolved to provide intake less limited by the shortage of a single protein "letter," say lysine, which is low in all three of the great cereals but high in the legumes.

The recommendations include eight or nine vitamins and half a dozen mineral elements. Beer drinkers should welcome the remark that "one litre daily almost meets the recommended intake" of riboflavin. No other vitamin is present in beer in significant amounts. Calcium is a particularly interesting mineral atom. The adult body contains about 1.2 kilograms, 99 percent of it in the skeleton, where crystals of the phosphate are held in a protein matrix together with much amorphous material. In the U.S. it is dairy products that provide nearly all the calcium taken in. All natural foods contain some of it, "fish consumed whole" being very rich, and a millet, ragee, yielding tenfold the amount in milk, gram for gram. "In countries where there is little milk the people are usually small," because the size of adults is in general related to the amount of milk they drank as children. That, however, is not the result of a lack of calcium. Bone samples from individuals on low-calcium or high-calcium diets are of equal quality. The low-milk examples are smaller because the better diet in youth promotes skeletal growth by providing protein for the matrix in which bone-mineral crystals are deposited.

The relation of calcium absorption to vitamin D (cholecalciferol) and to sunlight is complicated. Rickets is still a major problem in a few parts of the world; normal American adults can meet their requirements without any dietary vitamin D, natural or synthetic, unless their exposure to sunlight is too limited. Those who are growing bone—infants, children, adolescents, pregnant and lactating mothers—need an extra intake. Even a minor case of rickets in young girls can be enough to narrow the pelvis and render eventual labor long and hard.

These two expert groups curiously give the same garbled and misunderstood definition of the joule, the unit of mechanical work. This makes no material difference to their topic, but it is gloomy reading for teachers of physics and for all who value the unity of science.

Specialized Communications Techniques for the Radio Amateur. The American Radio Relay League, Inc., Newington, Conn. 06111 ($3). Those beeping dot-dash or tight-squeezed-voice transmissions that crowd the amateur radio bands satisfy nearly all who practice the half-magical art of direct communication across city or continent with a deskful of gear, store-bought or homemade. This small book, a carefully edited anthology of technical articles with many photographs, curves and diagrams, is aimed at the "stalwart fellows" who want something more. Their role is as secure as it is historical; the high frequencies were pioneered by just such amateurs when vacuum tubes were novelties. Can their future echo the past?

In the era of integrated circuitry and dial-direct-to-London it is not easy to pioneer. The last chapter of the book indeed suggests that only lasers and digital communications systems "hold any real promise" for amateurs of the future, although even that rather workaday promise is still speculative. Current but unusual amateur schemes are what is here described, with detailed information of a "how and when to do it" kind. ("Replace the stylus with carbon-steel wire only. A wire brush is a common source.") The book cannot help beginners much, although it may well seduce them. Its detailed hints are elaborated against a background of experience and feel that years of trial, some study and a good deal of help from one's friends alone supply. The level is rather like that of the more demanding examples from "The Amateur Scientist" in this magazine.

Three novel forms of amateur communications are the content of these 200 pages. The first is video-image exchange, the second is typing by radio link and the third is space communications. An amateur television station can be put on the air for a couple of hundred dollars "or less," using a converted FM taxi transmitter and a secondhand closed-circuit Vidicon camera, once the shoplifter's deterrent at some local store. Quality pictures can be exchanged on the wide 70-centimeter band, under good circumstances for "several hundred miles up and down the Atlantic Seaboard, or 1,000 or so across the Gulf." Much more difficult is the challenge of slow-scan television, an ingenious adaptation to amateur conditions. The standard video bandwidth requirement is here reduced by a factor of a couple of hundred. Such transmission demands no more spectrum space than a voice transmission, some three kilocycles, and is then legally and morally fitted for the narrow long-distance bands, giving the amateur the chance to look across oceans. The pictures no longer move but become a succession of stills, each lasting for seven or eight seconds. This departure from television standards makes heavy demands on the builder of such equipment, and the slow scan requires some form of storing the image as it builds up. The storage can be photographic (at the price of the frustrating delay for processing) or can make use of a long-persisting oscilloscope screen, video storage tubes or even digital memories.

Facsimile systems are related but even slower. They exchange graphics in permanent form, using waves only in the six-meter band and below it. The most common means is a cunningly converted Western Union Telefax transceiver, available used for about \$20. Other systems capture fascinating up-to-the-hour weather photographs from satellites. (Compare "The Amateur Scientist" for February, 1974.)

Oscar 6 orbits the earth 900 miles up. Since its launch in late 1972 it has offered a repeater service open to all, with an uplink at two meters and a downlink at 10. Amateurs can communicate by means of this spaceborne relay system while it remains within mutual range, a matter of some tens of minutes for distances of continental size. Antennas and careful orbit location are the main topics described here.

Finally there is moonbounce. At the margin of their possibilities ambitious amateurs exchange faint signals worldwide by way of that inconstant, irregular and passive satellite. Two meters is usually used, although the 70-centimeter band is gaining in popularity. Code signals at low band speed are standard; moonbounce has been achieved from a Brooklyn apartment-house rooftop antenna. A few stations sport big parabolic dishes (sometimes "borrowed" from another user). So far no two-way voice or teleprinter contact via the moon has been reported, but come next time of perigee, which offers a two-decibel advantage, you can be sure that some people will be trying.

BIBLIOGRAPHY

Readers interested in further reading on the subjects covered by articles in this issue may find the lists below helpful.

AGRICULTURE IN CHINA

THE TECHNICAL TRANSFORMATION OF AGRICULTURE IN COMMUNIST CHINA. Leslie T. C. Kuo. Praeger Special Studies in International Economics and Development, 1972.

MAKING GREEN REVOLUTION—THE POLITICS OF AGRICULTURAL DEVELOPMENT IN CHINA. Benedict Stavis. Rural Development Monograph No. 1, Cornell University, 1974.

THE AGRICULTURAL SITUATION IN THE PEOPLE'S REPUBLIC OF CHINA. Economic Research Service. ERS-Foreign 362, U.S. Department of Agriculture, May, 1974.

TRIP REPORT: PLANT STUDIES DELEGATION TO THE PEOPLE'S REPUBLIC OF CHINA—AUGUST 27–SEPTEMBER 23, 1973. National Academy of Sciences, in press.

HOW THE LIVER METABOLIZES FOREIGN SUBSTANCES

PHARMACOLOGICAL IMPLICATIONS OF MICROSOMAL ENZYME INDUCTION. A. H. Conney in *Pharmacological Reviews,* Vol. 19, pages 317–366; 1967.

MICROSOMES AND DRUG OXIDATIONS. Edited by R. W. Estabrook, J. R. Gillette and K. C. Liebman. The Williams & Wilkins Company, 1972.

MICROSOMAL HYDROXYLASES: INDUCTION AND ROLE IN POLYCYCLIC HYDROCARBON CARCINOGENESIS AND TOXICITY. Harry V. Gelboin, Nadao Kinoshita and Friedrich J. Wiebel in *Federation Proceedings,* Vol. 31, No. 4, pages 1298–1309; July–August, 1972.

METABOLIC INTERACTIONS AMONG ENVIRONMENTAL CHEMICALS AND DRUGS. A. H. Conney and J. J. Burns in *Science,* Vol. 178, No. 4061, pages 576–586; November 10, 1972.

SLAVERY IN ANTS

SLAVE-RAIDS OF THE ANT *POLYERGUS LUCIDUS* MAYR. Mary Talbot in *Psyche, Cambridge,* Vol. 74, No. 4, pages 299–313; 1967.

THE INSECT SOCIETIES. Edward O. Wilson. The Belknap Press of Harvard University Press, 1971.

CHEMICAL COMMUNICATION AND "PROPAGANDA" IN SLAVE-MAKER ANTS. F. E. Regnier and E. O. Wilson in *Science,* Vol. 172, No. 3980, pages 267–269; April 16, 1971.

LEPTOTHORAX DULOTICUS AND THE BEGINNINGS OF SLAVERY IN ANTS. E. O. Wilson in *Evolution,* in press.

ELECTRON-POSITRON ANNIHILATION AND THE NEW PARTICLES

THE WORLD OF ELEMENTARY PARTICLES. Kenneth W. Ford. Xerox College Publishing, 1963.

PHYSICS IN THE TWENTIETH CENTURY: SELECTED ESSAYS. Victor F. Weisskopf. The MIT Press, 1972.

DISCOVERY OF A NARROW RESONANCE IN e^+e^- ANNIHILATION. M. L. Perl, B. Richter, W. Chinowsky, G. Goldhaber, G. H. Trilling et al. in *Physical Review Letters,* Vol. 33, No. 23, pages 1406–1408; December 2, 1974.

EXPERIMENTAL OBSERVATION OF A HEAVY PARTICLE *J*. Samuel C. C. Ting et al. in *Physical Review Letters,* Vol. 33, No. 23, pages 1404–1406; December 2, 1974.

DISCOVERY OF A SECOND NARROW RESONANCE IN e^+e^- ANNIHILATION. G. Goldhaber, G. H. Trilling, M. L. Perl, B. Richter et al. in *Physical Review Letters,* Vol. 33, No. 24, pages 1453–1455; December 9, 1974.

PULSATING STARS

VARIABLE STARS. P. Ledoux and Th. Walraven in *Handbuch der Physik: Vol. 51,* edited by S. Flügge. Springer-Verlag, 1958.

A STUDY OF PULSATION IN RR LYRAE MODELS. Robert F. Christy in *The Astronomical Journal,* Vol. 144, No. 1, pages 108–179; April, 1966.

PULSATING STARS. John P. Cox in *Reports on Progress in Physics,* Vol. 37, pages 563–698; 1974.

VISUAL MOTION PERCEPTION

PERCEPTION OF MOTION AND CHANGING FORM. Gunnar Johansson in *The Scandinavian Journal of Psychology,* Vol. 5, No. 3, pages 181–208; 1964.

WHAT GIVES RISE TO THE PERCEPTION OF MOTION? James J. Gibson in *Psychological Review,* Vol. 75, No. 4, pages 335–346; July, 1968.

VISUAL PERCEPTION OF BENDING MOTION. G. Jansson and G. Johansson in *Perception,* Vol. 2, pages 321–326; 1973.

VISUAL PERCEPTION OF MOTION IN DEPTH: APPLICATION OF A VECTOR MODEL TO THREE-DOT MOTION PATTERNS. Erik Börjesson and Claes von Hofsten in *Perception & Psychophysics,* Vol. 13, No. 2, pages 169–179; April, 1973.

VISUAL PERCEPTION OF BIOLOGICAL MOTION AND A MODEL FOR ITS ANALYSIS. Gunnar Johansson in *Perception & Psychophysics,* Vol. 14, No. 2, pages 201–211; October, 1973.

PELAGIC TAR

PETROLEUM LUMPS ON THE SURFACE OF THE SEA. Michael H. Horn, John M. Teal and Richard H. Backus in *Science,* Vol. 168, No. 3928, pages 245–246; April 10, 1970.

OIL POLLUTION AND MARINE ECOLOGY. A. Nelson-Smith. Plenum Press, 1973.

PELAGIC TAR FROM BERMUDA AND THE SARGASSO SEA. James N. Butler, Byron F. Morris and Jeremy Sass. Special Publication No. 10, Bermuda Biological Station for Research, 1973.

THE ENVIRONMENTAL FATE OF STRANDED CRUDE OIL. M. Blumer, M. Ehrhardt and J. H. Jones in *Deep-Sea Research,* Vol. 20, No. 3, pages 239–259; March, 1973.

THE ROLE OF MUSIC IN GALILEO'S EXPERIMENTS

DIALOGUE CONCERNING THE TWO CHIEF WORLD SYSTEMS. Galileo Galilei, translated by Stillman Drake. University of California Press, 1967.

GALILEO STUDIES: PERSONALITY, TRADITION AND REVOLUTION. Stillman Drake. University of Michigan Press, 1970.

RENAISSANCE MUSIC AND EXPERIMENTAL SCIENCE. Stillman Drake in *Journal of the History of Ideas,* Vol. 31, No. 4, pages 483–500; October–December, 1970.

MATHEMATICAL GAMES

YOUR MOVE. David L. Silverman. McGraw-Hill Book Company, 1971.

MATH, WRITING & GAMES IN THE OPEN CLASSROOM. Herbert R. Kohl. Random House, 1974.

SETS, NUMBERS AND UNIVERSES: SELECTED WORKS. Stanislaw Ulam. The MIT Press, 1974.

THE AMATEUR SCIENTIST

HALF-HOURS WITH GREAT SCIENTISTS. Charles G. Fraser. Reinhold Publishing Corporation, 1948.

SATURDAY SCIENCE. Edited by Andrew Bluemle. E. P. Dutton & Company, 1960.

How did Dodge Colt put so much in such a little car?
Reclining bucket seats
Tinted glass
Carpeting
Sorry, you'll have to provide your own driver.
Thrifty four-cylinder engine. 30 mpg*
Four-speed transmission (five-speed in the GT)
1975
Locking gas cap
Bumper guards —front and rear
Adjustable steering column
Front disc brakes
Flow-through ventilation
NO.1 FOR THE MONEY IS DODGE
Introducing the '75 Dodge Colt Carousel hardtop. It comes with all the good things listed here, as do all the other Colt models: 2-door coupe, 4-door sedan, 4-door wagon and the Colt GT.
Prices start at $2,945.**
*Based on EPA test results for 1975 Dodge Colt, 1600 cc engine. 30 mpg highway cycle and 20 mpg city cycle.
**Manufacturer's suggested retail price for a '75 Dodge Colt coupe. Not included are state and local taxes, destination charge, license and title fees, and Dealer preparation charge, if any.
CHRYSLER MOTORS CORPORATION
Dodge
Dodge Trucks

It took a lot of technology to change a name like United Aircraft.

Because we're now a multi-market company, our old one-market name no longer fits.

The name our company has borne since 1934 is hardly descriptive of the activities in which we are now engaged.

While our traditional aircraft and aerospace businesses continue to grow, we've tapped the vast technology bank that's evolved from these operations to enlarge the company's business base in industrial and commercial fields.

Through the selective exercise of our abilities and skills over a wide spectrum of high technologies, we're now a multi-market corporation. But one with the same solid, dependable virtues. A corporation with 1974 sales in excess of $3.3 billion, substantial financial strengths, a 39-year record of consecutive dividend payments, a truly international business with representation in some 120 countries of the world.

We're also a corporation with a promising future. Because when all those technologies are United, there's no limit to our powers of invention. United Technologies Corporation, Hartford, Conn. 06101.

	1974	1964
Total Sales	$3,321,106,000	$1,235,918,000
Net Income	104.705,000	29,084,000
Business Backlog	3,577,000,000	1,200,000,000

UNITED TECHNOLOGIES

Hamilton Standard's Autosense is the world's most sophisticated computerized diagnostic unit for automotive engine analysis.

Electric utilities use our gas turbines in single- and multi-engine plants for peak demand. These units come in modular packages for fast on-site installation.

Our Essex subsidiary produces cable like this for 1600 telephone companies, and has pioneered in moisture-proof cable for underground installation.

We'll always be a major force in flight—both military and commercial. Our Pratt & Whitney Aircraft engines power aircraft for 188 airlines around the world.