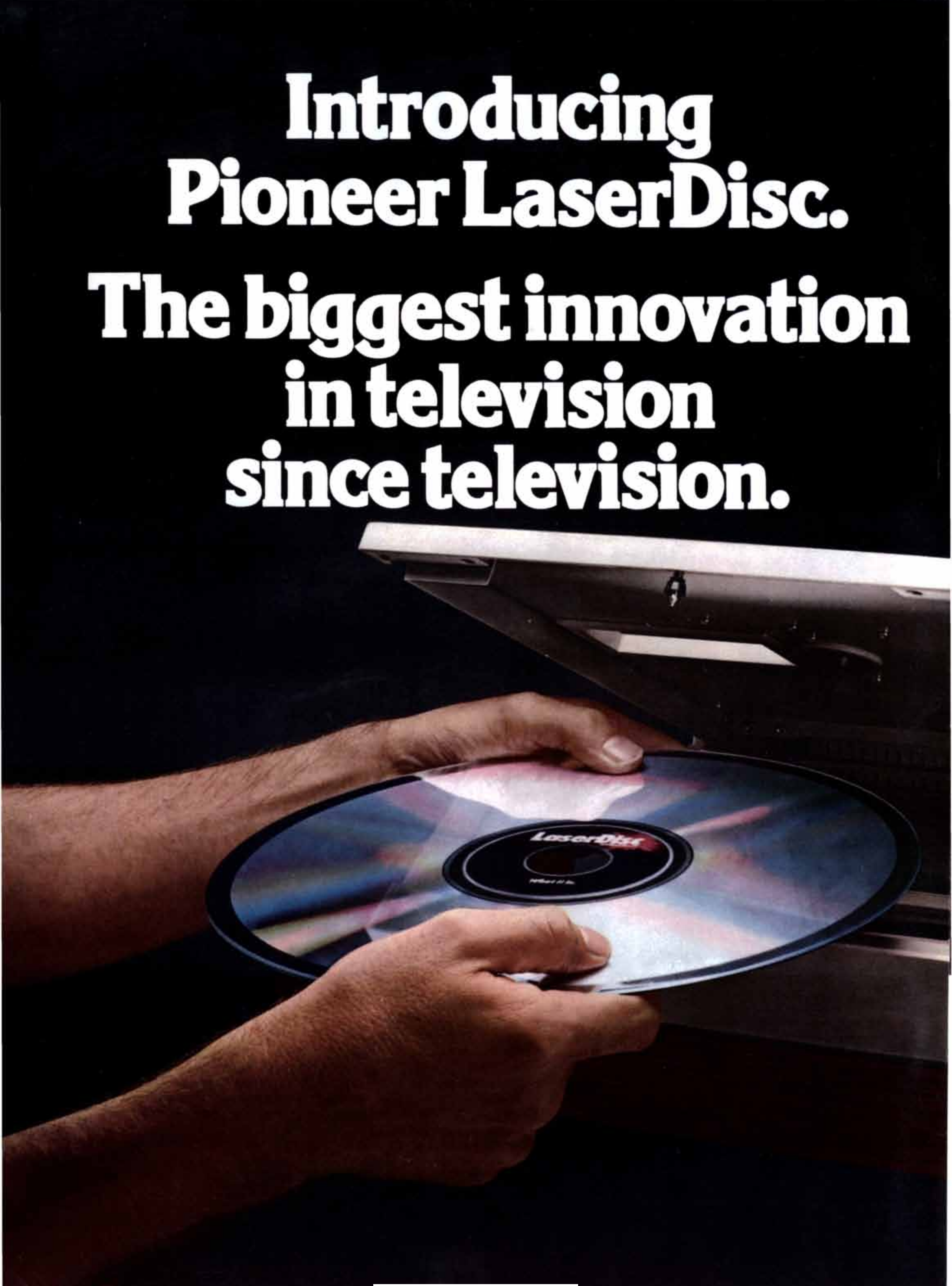
SCIENTIFIC AMERICAN

GENETICS OF WHEAT

$2.00

January 1981


Introducing
Pioneer LaserDisc.
The biggest innovation
in television
since television.

Imagine you could sit down in front of your TV set and see virtually any movie or concert you wanted to see when you wanted to see it.

Imagine you could actually see and hear concerts on your TV in stereo. The best stereo you've ever heard. Or cut to your favorite scene in a movie at will. Or study sports in slow motion, even one frame at a time. Imagine a machine that could teach your children at their own rate.

You now have just an idea of Pioneer LaserDisc. A remarkable innovation that puts both picture and sound on a record. And plays them both by means of a laser beam onto your TV and through your hi-fi. (The player hooks up to your TV with just one wire. And when it's not in use, your TV plays the way it normally plays.)

The laser picture quality is exceptional. As good as the best broadcast reception you've ever seen. And laser sound is better than the best conventional audio recordings you've ever heard. And since nothing touches the disc but a laser beam, the disc never wears out. The quality is forever.

For all it does, surprisingly, the suggested retail price of the player is only $749* (just $50 more with remote control). And you can own a disc of a great movie or concert forever for the cost of taking your family to the movies.

There are a few hundred different discs to choose from right now. And more and more are coming out every day. Someday, virtually anything that entertains anyone will be on the disc.

Nothing we say here will fully prepare you for the magic of Pioneer LaserDisc. You simply have to see it.

For a personal demonstration from the dealer nearest you call us at **800-621-5199** toll free. (In Illinois 800-972-5855.)

LaserDisc
PIONEER

*Suggested retail price, actual price set by dealer
© 1980 U.S. Pioneer Electronics Corp

Established 1845 SCIENTIFIC AMERICAN January 1981 Volume 244 Number 1

ARTICLES

DEPARTMENTS

SCIENTIFIC AMERICAN (ISSN 0036-8733), PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y 10017 COPYRIGHT © 1980 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. CONTROLLED CIRCULATION POSTAGE PAID AT NEW YORK, N.Y. (USPS 509-530), AND AT CHICAGO, ILL. (USPS 508-030). AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $21 PER YEAR, U.S., ITS POSSESSIONS AND CANADA; $27 PER YEAR, ALL OTHER COUNTRIES. POSTMASTER: SEND ADDRESS CHANGES TO SCIENTIFIC AMERICAN, 415 MADISON AVENUE, NEW YORK, N.Y 10017

THE COVER

The painting on the cover shows the upper part of the spikes, or seed-bearing heads, of three plants belonging to the genus *Triticum,* a subdivision of the Gramineae, or grass family. The three species represented are *T. tauschii* (*left*), a wild grass native to both central and southwestern Asia, *T. turgidum* (*middle*), a primitive form of wheat that originated somewhere in southwestern Asia, and a manmade version of *T. aestivum,* or common wheat (*right*), obtained by crossbreeding the first two species named and then treating the hybrid plant with the drug colchicine to induce chromosome doubling, and hence full fertility, in the subsequent generations. The experimental procedure mimics the spontaneous hybridization that presumably took place in a Neolithic farmer's field in western Iran, whereby the diploid (*DD*) genome of *T. tauschii* was added to the tetraploid (*AABB*) genome of *T. turgidum* to create the hexaploid (*AABBDD*) genome of *T. aestivum.* The exploitation of the large reservoir of genetic variability contained in the wild relatives of wheat for the production of synthetic amphiploid hybrids, such as the one shown on the cover, is considered essential to the restoration and enrichment of the gene pool of the cultivated wheats (see "The Wild Gene Resources of Wheat," page 102).

THE ILLUSTRATIONS

Cover painting by Enid Kotschnig

Page	Source
24	Alan D. Iselin
56–65	Albert E. Miller
67	José M. Sogo and Theo Koller, Swiss Federal Institute of Technology
68–69	Allen Beechel
70	T. O. Diener, Plant Virology Laboratory, Beltsville, Md.
71	Allen Beechel
72–73	Allen Beechel (*top and bottom right*); William L. McClements and Paul J. Kaesberg, University of Wisconsin (*bottom left*)
75–80	Patricia J. Wynne
92	George V. Kelvin
94	Hale Observatories
95	Kitt Peak National Observatory (*top*), George V. Kelvin (*bottom*)
96	George V. Kelvin
97	Max Planck Institute for Radio Astronomy
99–100	George V. Kelvin
103–104	Shalom Nedam, Weizmann Institute of Science
105	Enid Kotschnig (*top*), Ilil Arbel (*bottom*)
106–107	Ilil Arbel
108	Enid Kotschnig
109–111	Ilil Arbel
112	T. Mello Sampayo, Calouste Gulbenkian Foundation, Oeiras, Portugal
115–120	Jerome Kuhl
125	Fritz Goro
126–127	Gabor Kiss
128	Fritz Goro
129–138	Gabor Kiss
156	Smithsonian Institution (*top*), Library of Congress (*bottom*)
158	David A. Hounshell, University of Delaware
159–160	Dan Todd
162	Dan Todd (*top*), SCIENTIFIC AMERICAN (*bottom*)
165	Jearl Walker (*top*), Christopher Heilman (*bottom*)
166–169	Michael Goodman

LETTERS

Sirs:

I had occasion recently to challenge an editorial in the British medical journal *The Lancet,* only to find that you have made the same editorial the subject of a note titled "Death in Britain" ["Science and the Citizen," SCIENTIFIC AMERICAN, June, 1980]. It is true that male mortality rates from coronary heart disease in the U.S. were substantially higher than those in England and Wales in 1968. It is not true, in spite of a large decrease, that by 1976 the American rates were actually lower, although they had closed the gap and the trend lines suggest that a crossover might take place later. In 1977 the age-standardized rates for the two countries were identical. Data from 1978 will be of interest.

HUGH TUNSTALL PEDOE

Department of Epidemiology
St. Mary's Hospital Medical School
London

Sirs:

It was with interest that I read your note about the ability of Andrew Levine to talk backward ["Science and the Citizen," SCIENTIFIC AMERICAN, July, 1980]. I must advise that Dr. Levine's ability is far from unique. In Sydney there are hundreds (perhaps thousands) of butchers who can speak reverse English (as well, of course, as forward English). This ability appears to have evolved during the late 1920's in response to the requirement of communicating price changes in the presence of shoppers.

Subsequent development led to a fluency in reverse English among most of the butchers in the Sydney region. I understand that the number of people currently able to speak reverse English is diminishing, because of a lack of interest among new apprentices and new legal requirements with respect to pricing policies.

GRAHAM F. LEE

Randwick, Australia

Sirs:

Professor Parker's discussion of adaptations that occur in the multimodal system of balance and orientation ["The Vestibular Apparatus," by Donald E. Parker; SCIENTIFIC AMERICAN, November, 1980] reminded me of a phenomenon that must have affected millions of Army trainees over the years since I first noticed it during basic training in World War II.

During the first few days of riding in a jeep while wearing a steel helmet one's head tends to follow the heavy helmet's inertial path during rapid acceleration and deceleration. That an adaptation has taken place becomes amusingly evident only when one then switches back to lighter headgear, in my case the featherweight overseas cap. For the first few rides you find your head involuntarily nodding at takeoff and tilting back during braking periods.

ROBERT C. CONNELL

Portola Valley, Calif.

Sirs:

Engineer Egon Orowan ["Letters," SCIENTIFIC AMERICAN, September, 1980], commenting on "The Safety of Fission Reactors," by Harold W. Lewis [SCIENTIFIC AMERICAN, March, 1980], implied that the Three Mile Island accident demonstrated a certain lack of competent, expert engineering in the design of that plant. Orowan recalled a conversation in which the late Sir James Chadwick asked why everything went well at Los Alamos if it was done by physicists and all the troubles occurred where engineering was involved.

As a physicist I think it is only fair to recall another event from the early history of reactor design that may help to restore some balance. The Hanford reactors, designed to produce plutonium for the Manhattan project, were the first reactors to operate at significant power levels (250 megawatts compared with one megawatt for the earlier X-10 graphite reactor at Oak Ridge). In September, 1944, during the initial start-up and approach to full power, the first Hanford reactor unexpectedly shut itself down completely with no action by the operators. This event caused great consternation among the physicists at Argonne because the entire multimillion-dollar plutonium-production project was in jeopardy. The physicists soon discovered that the effect was due to the buildup of xenon 135, a strong neutron absorber, among the fission products. Because the equilibrium concentration of xenon 135 was dependent on power level it had not been detected in earlier experiments conducted at much lower power levels.

The solution was to overcome the neutron-absorption effect of xenon by adding more uranium to the periphery of the reactor, thus decreasing neutron leakage by increasing the size of the reactor. This was possible only because the du Pont engineers responsible for the design had provided a generous number of extra channels for fuel, well beyond the anticipated needs of the physicists at the Argonne Metallurgical Laboratory. Quoting from the official history of the project (*The New World,* Volume 1 of *A History of the United States Atomic Energy Commission,* by Richard G. Hewlett and Oscar E. Anderson, Jr., page 308):

"The scientists at the Metallurgical Laboratory had cited this feature as one of the best examples of extravagant conservatism in the du Pont design. Without the slightest reason for believing the additional tubes would ever be necessary, the du Pont designers had stuck to their guns as a matter of principle. Now it seemed that the incredible had occurred. More from luck than foresight du Pont was prepared to meet the situation."

Although some may call it luck, I think sound engineering judgment would be a more appropriate description. In scaling up any new and untried process, experience had taught the du Pont engineers to provide conservatism and flexibility in the design, wherever possible, to meet unforeseen problems that were sure to arise.

The sound judgment of many engineers and physicists in introducing sufficient conservatism into commercial reactor designs to protect the public was demonstrated by the accident at Three Mile Island. Certainly there are many lessons to be learned from this accident that will enable physicists and engineers, working in close cooperation, to further improve the safety margins for both new and existing reactors.

RAYMOND P. SULLIVAN

Falls Church, Va.

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

Advertising correspondence should be addressed to C. John Kirby, Advertising Director, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017.

Offprint correspondence and orders should be addressed to W. H. Freeman and Company, 660 Market Street, San Francisco, Calif. 94104. For each offprint ordered please enclose 60 cents.

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

Name ____________

New Address ____________

____

____

Old Address ____________

____

____

Simple evolution. The Honda Civic has grown a trunk.

FOUR HEADS ARE BETTER THAN TWO.

JVC's Vidstar has four video heads and the clarity that goes with them.

Admit it. Despite all the fancy features that videocassette recorders have to offer, you're worried about picture quality. Will it be crisp and vibrant? Will you get superb reproduction of recorded material?

Well, stop worrying. When JVC brought you 6-hour mode recording and playback, we equipped Vidstar™ with *four* video heads. Here's why.

Vidstar's 4-Head Construction

Vidstar uses one pair of recording/playback heads for the 2-hour mode and another pair of heads for the 6-hour mode. Each pair is suited to performing at a specific tape speed. This specialization gives Vidstar unsurpassed picture quality in both the 2-hour and 6-hour modes. That's the kind of performance you'd expect from the people who developed the VHS system. And who developed this four head technology. That's us, JVC.

That's only the beginning. Take a look at our other features: 6-hour recording on a single 2-hour VHS cassette, multi-function remote control unit, stop action and slow motion, double and triple speeds, and a micro-computer-assisted timer that can record up to 42 programs a week. You simply pre-set days, times and channels.

Vidstar is designed to let you add components to your home entertainment system effortlessly. You'll find goof-proof features that prevent recording mistakes and other features that let you add your own audio to recorded material. Vidstar has features you may not even know you want yet.

But you'll want them the minute you get your hands on them. You can do exactly that at your nearest JVC Vidstar dealer.

For JVC dealer names and locations call TOLL-FREE 800-221-7502. In NY, 212-476-8300.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

JANUARY, 1931: "The chief difficulty encountered in the operation of a rocket is that of finding a substance that will blow out of its container with such terrific velocity as to produce a very great back-pressure and at the same time be under control. The most persistent and effective investigator of the problem is Professor Robert H. Goddard of Clark University. Just now he is at work on it harder than ever before, having both financial backing and the counsel and moral support of many eminent men of science. These savants do not envision an early Jules Verne trip to the moon in a super skyrocket. They do envision, however, the acquisition of new information about the solar and terrestrial atmospheres. Possibly we may someday be able to send a rocket with a radio-recording instrument beyond the conducting limits of our own envelope and thus know definitely whether radio static of appreciable intensity is or is not knocking at our outer door. At the other end of the spectrum we are helpless even in the highest balloons and on the tops of the loftiest mountains. Long before we have got halfway down the spectrum from the limit of visibility the atmosphere has become opaque. The rocket also can give us much information about the upper air, concerning which many more questions are asked than as yet can be answered."

"The word *galápagos* is Spanish for tortoises. It has been estimated that 10 million of the animals have been taken from the Galápagos Islands since their discovery. The Galápagos tortoise had apparently a food value to the Pacific whaling fleet not unlike that of the bison to the settlers of the Great Plains. The animals have long been extinct on most of the islands where they once swarmed. With this fact under consideration the board of managers of the New York Zoological Society authorized Dr. Charles Haskins Townsend to procure a breeding stock of tortoises for colonization in our southern states and elsewhere under favorable climatic conditions. The success of the expedition greatly exceeded expectations. One hundred and eighty tortoises were captured in the mountains of southern Albemarle Island in the Galápagos. Most of the tortoises brought back were soon placed in colonies along the southern border of the United States. All are in the keeping of zoological gardens or other responsible co-operating agencies."

"Professor T. H. Morgan, director of the William G. Kerckhoff Laboratories of the Biological Sciences at the California Institute of Technology, whose presidential address at the annual meeting of the American Association for the Advancement of Science held this year in Cleveland over the Christmas holiday will no doubt receive widespread notice, is the foremost living investigator of genetics and is regarded by many other scientists as America's ablest man of science. In addition he is president of the National Academy of Sciences. Professor Morgan's contribution to human knowledge is an understanding of much of the problem of heredity and the isolation of principles that underlie and guide the practical work of plant and animal breeders. The discovery of the mechanism of heredity has enabled other workers to breed cows that give more milk, hens that lay more eggs, wheat of higher yield and better quality, corn with more sugar, sheep with more wool and better meat. In sheer economic value alone Professor Morgan has been worth billions."

"Individuals, committees, city and state governments and even certain branches of the Federal Government itself have over the past few weeks worked out praiseworthy plans for the relief of unemployment. The problem of unemployment, however, assumes a more serious aspect with the coming of cold weather. If our winter is severe, there is bound to be a great deal of suffering, particularly in our cities, where many thousands will lack sufficient fuel, warm clothing and nourishing food. The work of making jobs must therefore go on, not half-heartedly but with well-directed vigor. We millions of well-fed, well-clothed ones who have been touched only superficially by the depression must open up our purses as we have not since the war and must give to the Red Cross, the Salvation Army, the Volunteers of America and other organizations of a like nature in our communities. And for those people who cannot be classed as being among the poor who receive charity we must make more and more jobs."

SCIENTIFIC AMERICAN

JANUARY, 1881: "To subject his system of electric lighting by incandescence to the crucial test of actual outdoor use on a large scale Mr. Edison has set up at Menlo Park a plant embracing 500 lamps distributed over an area a mile long and half a mile wide. The lamps are in a circuit comprising seven and three-quarter miles of wire and are supplied by a current generated by nine dynamo-electric machines driven by one engine. The lamps in Mr. Edison's test are of 16 candle-power, equal to an ordinary street gas light, and are absolutely steady, shining with a mild and serene effulgence that is exceedingly pleasing to the eye. The division of the current is complete and economical, and the entire system of lights can be turned up or down, off or on, as easily as one can regulate the flow of gas in an ordinary burner."

"The island of Madagascar is inhabited only by animals the pedigree of which can be traced back to the Eocene epoch. Madagascar must have been separated from the African continent at least since that time, this being the reason its fauna has developed in a manner quite different from that of other parts of the world."

"At a late meeting of civil engineers in St. Louis a very interesting paper was read by Mr. O. Chanute, summarizing the progress and wonderful growth engineering has made in this country. In metallurgy the increase of our blast furnaces is particularly notable. In the amount of iron produced we are next to England, with Germany standing third. Our steel industry, which is now second only to that of Great Britain, will surpass the British one in another year. Our increase has been 50 per cent in two years. In agricultural engineering our progress has been truly wonderful, and before it all other branches become as dust in the valley. In the mechanization of plowing alone the annual saving of labor in producing our crops amounts to fully $36,000,000 less than the same work would have cost 30 years ago."

"A great many steam wagons and carriages have been devised and built for transporting loads on the highway without tracks, but although some of the devices were masterpieces of ingenuity, the practical results were never perfectly satisfactory. A steam carriage invented and built by the French engineer Amédée Bollée is an object of more than ordinary interest. The driving engine is in the front part of the carriage and is controlled by the engineer seated above it, who also operates the steering gear and the powerful brake levers. The boiler is placed above the rear axle, the coal bins are at each side of the boiler and the water truck is under the seat of the engineer. The average speed attained is 18 miles per hour, but a maximum of 22 m.p.h. is reached. The fuel is coke, which produces very little smoke."

"The Commissioner of Public Works in New York City has granted Mr. Edison a permit to introduce his new system of electric lighting in the lower part of the city."

THE AUTHORS

H. WILLIAM MENARD ("Toward a Rational Strategy for Oil Exploration") is director of the U.S. Geological Survey. A graduate of the California Institute of Technology, he obtained his Ph.D. in geology from Harvard University in 1949. He was a marine geologist at the U.S. Navy Electronics Laboratory in San Diego from 1949 to 1955, when he joined the faculty of the Scripps Institution of Oceanography as professor of geology. In his career as a marine geologist Menard took part in 20 deep-sea oceanographic expeditions in both the Pacific and the Atlantic. He was also involved in underwater mapping off the coast of California, logging more than 1,000 aqualung dives for geological purposes. His interest in the politics and sociology of science began as a "research sideline" in the mid-1960's, when he was technical adviser on marine science and geology in the Office of Science and Technology of the Executive Office of the President. He also served on a number of committees of the National Academy of Sciences concerned with natural resources and the environmental impact of the development of offshore oil. Menard's article is based on work done while he was still at Scripps, before he was appointed to succeed Vincent E. McKelvey as director of the Geological Survey in 1978. It is not intended as a policy statement of the Survey.

T. O. DIENER ("Viroids") is a research pathologist in the Department of Agriculture's Plant Virology Laboratory in Beltsville, Md. He received his undergraduate and graduate education at the Swiss Federal Institute of Technology in Zurich, where he was awarded his doctorate in plant pathology in 1948. He came to the U.S. in 1950 and spent most of the next decade studying virus-infected fruit trees at an experimental station of Washington State University. Since going to work for the Department of Agriculture in 1959, he has investigated various aspects of virus diseases in plants. It was in the course of this research that Diener discovered the novel group of disease agents now known as viroids.

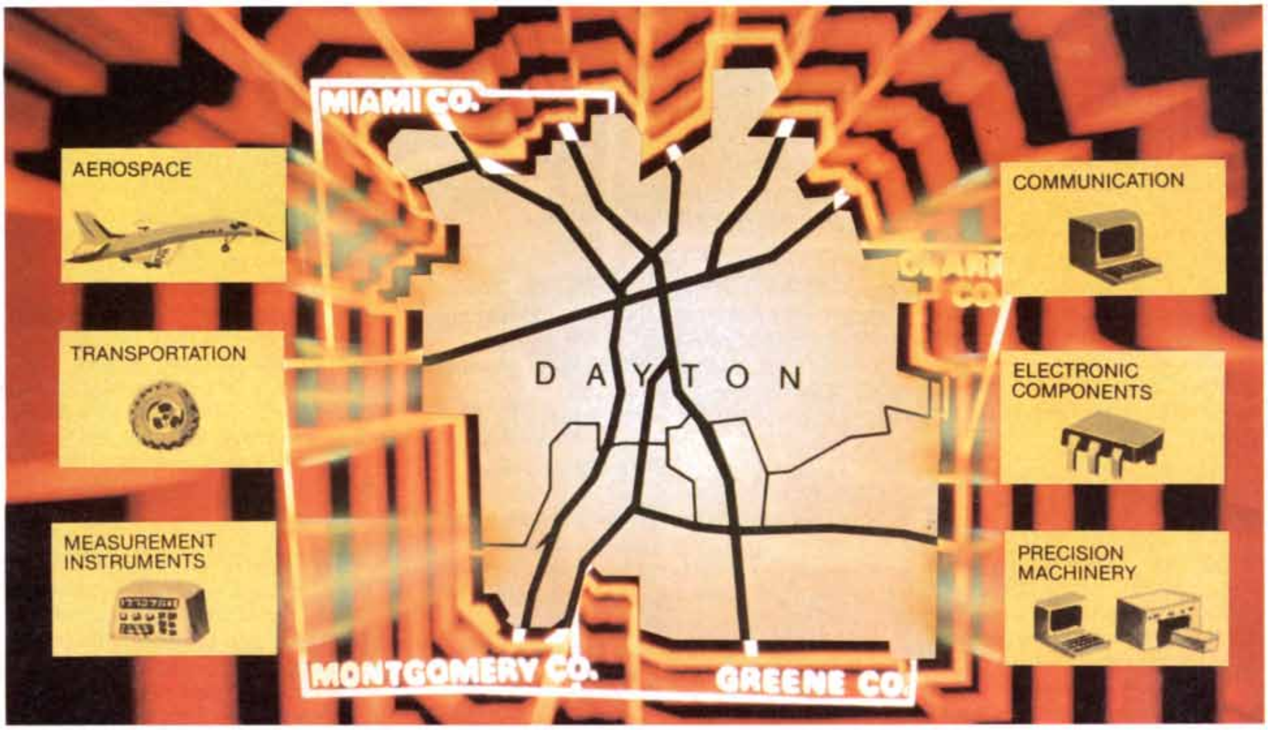
ROBERT K. JARVIK ("The Total Artificial Heart") is research assistant professor of surgery and bioengineering at the University of Utah School of Medicine, where he got his M.D. in 1976. His interest in medical engineering goes back to his high school years, when he began working on automatic surgical staplers. He enrolled in the School of Architecture of Syracuse University, he reports, but switched to premedical studies shortly after his father was operated on for an aortic aneurysm in 1965. As an undergraduate he continued his development of surgical-stapling instruments, for which he now holds a number of U.S. and foreign patents. Jarvik attended medical school at the University of Bologna for two years and completed the master's-degree program in biomechanics at New York University before joining the artificial-heart program at Utah in 1970.

PAUL W. HODGE ("The Andromeda Galaxy") is professor of astronomy at the University of Washington. His degrees are from Yale University (B.S., 1956) and Harvard University (Ph.D., 1960). Before he joined the faculty at Washington in 1965 he taught and did research at a number of institutions, including Harvard, the Hale Observatories, the California Institute of Technology, the University of California at Berkeley and the Smithsonian Astrophysical Observatory. He is the author of 10 books and more than 200 articles and scientific papers, including "Dwarf Galaxies," which appeared in the May 1964 issue of SCIENTIFIC AMERICAN. He was awarded the Beckwith Prize in astronomy by Yale in 1956 and the Bart J. Bok Prize in astronomy by Harvard in 1962. In 1978–79 he served as chairman of the astronomy section of the American Association for the Advancement of Science. Hodge's major fields of investigation at present include the evolution of stars and galaxies and the nature of meteorite craters and interplanetary dust.

MOSHE FELDMAN and ERNEST R. SEARS ("The Wild Gene Resources of Wheat") have had a long collaboration in the study of the genetics of wheat. Feldman is currently professor of genetics at the Weizmann Institute of Science in Israel. Born in Israel in 1933, he studied botany, genetics and biochemistry at the Hebrew University of Jerusalem, from which he received his Ph.D. In 1964 he joined Sears as a postdoctoral fellow at the U.S. Department of Agriculture's Cereal Genetics Research Unit at the University of Missouri. In 1967 he went to the University of Manitoba to continue his research. Feldman returned to Israel and joined the Weizmann Institute in 1969. Sears recently retired after working for 44 years as an investigator for the Agricultural Research Service, a branch of the Department of Agriculture. He grew up on a farm in western Oregon and earned his B.S. in agriculture from Oregon State College in 1932 and his Ph.D. in biology from Harvard University in 1936. He has won a number of honors for his work on the origin, evolution and cytogenetics of wheat, among them the

$\sum_0^\infty Tt$= DAYTON: Formula for High Technology

This special report on advances in high technology in the Greater Dayton Area was written by Peter J. Brennan, writer on international business, finance and development in concert with Development Counsellors International, Ltd.



$\sum_0^\infty Tt$= DAYTON: Formula for High Technology

THAT EQUATION YIELDS A TRUE BUT PERHAPS UNEXPECTED RESULT. WHEN ONE ADDS TOGETHER THOSE FACTORS THAT ARE IMPORTANT TO PEOPLE AND COMPANIES CONCERNED WITH ADVANCED TECHNOLOGY—EDUCATIONAL RESOURCES, COMMUNICATIONS AND TRANSPORTATION INFRASTRUCTURE, SUPPORT INDUSTRIES, AN EXISTING BASE IN AEROSPACE AND ELECTRONICS AND THE LIKE—THE ANSWER COMES UP DAYTON, OHIO.

Dayton is the center of a four-county, 850,000 population standard metropolitan statistical area (SMSA) in southwestern Ohio that is home to several long established high-technology firms, supports a number of institutions of higher learning that strongly emphasize advanced technology, has an extensive support industry in precision machining, electrical and electronic components and contract research and testing, has important chemical research facilities and, above all, has as an integral part of the industrial community the very fountainhead of the nation's aerospace research and development.

The Ninety Minute Market...

Located at the confluence of three rivers, a nodal center for air freight and express (Emery Air Freight has its regional center here), juncture of two major interstate highways, the city is superbly located to service quickly and efficiently high-technology, high-value industries over a wide area. In fact, Dayton in one sense is the tenth largest marketing area in the United States.

That sense is as the hub of a transportation net that holds within its ninety-minute trucking radius a market of some 4.1-million people. Only nine other cities reach larger markets within a similar radius. And each of those nine is in some way inferior in other respects to Dayton—costs, prices, taxes, transportation, education, quality of life, labor and, not least, technological infrastructure.

Ever to the Future...

It was Dayton, Ohio, that nurtured Orville and Wilbur Wright, two brothers who refused to accept man's earth-bound status. In Dayton they built their models out of bicycle technology and tools. And it was to Dayton that they returned to further develop their successful technology.

The pioneering, entrepreneurial scientific bent preceded the Wrights and is as strong today in the region's enterprising people. True to its heritage, the city has the highest percentage of engineers and scientists in its population of any in the region. A competent technical workforce brought up in an era and area of craftsmanship and employed in many small but technically sophisticated companies provides a base for future technologies.

Given focus by a strong educational system, a deeply rooted foundation in advanced technology going back to the 19th Century and a determined and cooperative drive centered on the well-supported Dayton Development Council, that spirit is preparing the ground for the technology of the Eighties and beyond.

In 1979, the Development Council established a Science and Technology Task Force to identify scientific and technological industries that represent the greatest potential for the area and to prepare strategies and recommendations.

The Task Force found that Dayton

$\sum_{0}^{\infty} Tt$ = DAYTON: Formula for High Technology

had good representation among the fastest growing technological areas, though in absolute numbers this representation reflects the region's size. There are over sixty firms in three of the fastest growing high-technology areas—electronic computing equipment, aircraft and space, and electronic components. In medium technology areas that today themselves use much high-technology and provide essential support to high-technology companies, Dayton has over 150 firms.

These together with the excellent technological foundation provided by the area's educational institutions and the continual input of technology of the very highest order through the Air Force research facilities at Wright Patterson Air Force Base (WPAFB) do indeed position the Dayton region for the future.

The Legacy of Flight

The Wrights' chief legacy is Wright-Patterson Air Force Base (WPAFB), home for the Air Force Logistic Command and key units of the Air Force Systems Command. Here, 25,000 people, military and civilian, carry on the Wrights' work. WPAFB is the major research and development center for the Air Force, and thus for all of aeronautics.

The Aeronautical Systems Division and the Air Force Wright Aeronautical Laboratories have mostly civilian staffs that pursue both long and short range technology for Air Force programs. The level ranges from basic research to manufacturing technology. About 75% of the multi-billion dollar budget goes to contractors while the balance supports in-house work. "An increasing amount of the innovative research and development is in-house," says George Peterson, civilian Director of the Materials Laboratory. "We use the in-house work to stimulate industry into producing the materials, equipment and technology that we need."

Leading-Edge Technology...

The Materials Laboratory pioneered composite materials. These are now showing up in automobiles, an absolute boon to manufacturers who must reduce weight. "Current research," says Mr. Peterson, "involves powder metallurgy, which opens up a tremendous vista of new alloys."

Other things turn up. "We pioneered rare earth/cobalt magnet technology here," says Peterson, proudly. Another development is the apparent discovery of a room-temperature superconductor. This serendipitous event could have incalculable consequences for the electrical and electronic industries.

Making Things Work...

The Flight Dynamics Laboratory in a sense puts it all together. Says the Laboratory's Deputy Director, Keith Collier, "We do research and development on aeronautical systems with respect to structure, equipment, control and mechanisms. It's a full spectrum."

The lab has been working for years on the design and structural aspects of a new method for forming and bonding titanium which emerged from Materials Laboratory development efforts. This system, called super plastic forming/diffusion bonding structures, reduces the metal to a plastic state during which it can be molded to any desired shape with uniform wall thickness and bonded without welding or brazing.

Fitting these systems to people is the job of the Human Engineering Division of the Air Force Aerospace Medical Research Laboratory.

The Lab designs flight simulators, some of them airborne. To save fuel and expensive aircraft, for example, the Air Force will outfit a small jet with a cockpit and computers that will make the plane look, feel and act in flight like a B-52.

A Technological Upwelling...

WPAFB pumps a lot of money into the area, some $540-million in salaries alone. But it also provides a far more important resource—highly educated, trained and intelligent manpower. As the nation's center for aeronautical technology, the base continually brings new people to the area, many of them engineers and scientists.

The Air Force's huge technological resource has substantial spinoff. Locally, the level is largely R&D oriented rather than production and intended to expand the Base's own capacities. This spinoff has helped many a firm to germinate and nursed it through infancy.

There is still more to the base. The on-site Air Force Institute of Technology is a graduate school of note, offering technical and scientific education at the doctorate level. Base scientists and engineers, military and civilian, teach at local schools and universities. The immense Air Force Museum is an extraordinary-three-dimensional archive of flight technology that has pristine examples of planes from every age. The Museum, too, is an educational and technological resource.

An Industrial Base...

Many a local advanced technology firm got its start with an innovative proposal on an Air Force project. Other local firms benefit from the technological spinoff. The continuing presence of the base not as an outside institution imposed upon the region but as an integral part of the community has a synergistic influence on the regional

The research and development facilities in Area B of Wright Patterson Air Force Base are a huge complex of interrelated laboratories and institutions including the Air Force Institute of Technology and, upper right, the Air Force Museum. Several samples of recent aircraft are visible adjacent to the hangar-like museum.

$\sum_0^\infty Tt$= DAYTON: Formula for High Technology

economy. Neither the base nor the community is passive to one another. Local businesses of all sizes and levels of technology support the base. In its turn, WPAFB provides a stable foundation to the local economy and ever renewing opportunities for venturesome enterprises to benefit from the very latest in aerospace, electronic and materials technology.

Since WPAFB is not a production unit but rather one of research and development and logistic support, its activities tend to remain at fairly stable levels. Today's research will result in production runs years in the future. At WPAFB, the activity level for few or many units remains much the same. Research is on-going regardless of fluctuating levels in production appropriations.

Neither the base nor local companies make airplanes in Dayton anymore. But many local firms are heavily engaged in high-technology component manufacturing and research and development for aircraft and parts, missiles, space vehicles and associated equipment. Some 14 firms with combined growth rates exceeding 11%, many of which started and grew directly out of contracts for WPAFB, now thrive in the region. Among these are established names like Cessna Aircraft Co., United Aircraft Products, B.F. Goodrich, Hartzell Propeller and the Delco Moraine Division of General Motors.

Other firms nominally listed under other classifications also grew out of WPAFB and continue to do much aerospace-related work as well as electronic, instruments and controls, environment and the like. Their foundation was the base but their worlds have grown larger. Some of these include Technology, Inc., United Systems Corporation and Systems Research Laboratories.

A Black Art...

"Aerodynamics is still a black art," says Dr. Robert Finch, Director of Program Management for the Instruments and Controls Division of Technology Incorporated, a rapidly growing $23-million company founded in 1959 in Dayton as a source of research, scientific and software studies for the Government, i.e., WPAFB. "We can only conduct analytical studies to a certain point, then we have to build things and see what really happens."

One area of growing interest is the effect of lightning on aircraft. Ironically, as the proportion of non-metallic materials in planes has increased, so has their tendency to build up high static charges with consequent havoc to electronic systems and physical vulnerability to lightning strikes. Technology Incorporated tries to predict the damage potential of various configurations and incorporate the findings in specifications.

The company's industrial sector, applies advanced technology to industrial products. These include annunciator panels for power plants and low-flow flowmeters for processing industries. The company has also opened significant new export markets through contracts with the German Air Force and with foreign orders for its flow devices.

Lasers for Measurement...

More electronic-oriented and perhaps more Government-oriented is Systems Research Laboratories. This 25-year old company, founded and still headed by Fritz J. Russ, an electrical engineer does some 70% of its business with the Government.

A significant portion of the company's work is in electronic warfare systems. Specifically, these programs lead to such things as techniques for distinguishing low level signals from a high-level background and expertise in laser technology. And this expertise then spills over into the commercial sector with such devices as high performance TV cameras and color displays, non-destructive testing systems and non-contact measurement and process control systems.

The $40-million company considers the WPAFB a major resource in many ways. Says Mr. Russ, who put together his business plan while he himself was working for the Missile Guidance and Control Group at WPAFB, "The Air Force brings lots of good talent here. Dayton becomes home for technically-oriented lieutenants who marry and raise families here."

Protective Treatments, Inc. of Dayton uses an SRL microprocessor based laser measurement and control system to measure the total width of automotive bumper trim to .0005 of an inch. The SRL AutoMike system assures the exact required dimension of the product and also minimizes the use of excess material.

Precision Machining...

Often it is only the smaller company that can afford to work at the very edge of manufacturing technology to prove out the materials, designs and techniques developed in larger, well funded organizations. To do so, these smaller companies must themselves remain at the edge of their own technology.

"At a time when my company was using manually-operated machine tools, I was waiting for a job to come in that I regularly depended on," recalls Robert Millat, President of Millat Industries Corp., a progressive example of the nearly 200 such companies that form a most important sector of the Dayton resource base. "I found that my former customer had bought his own numerical controlled machine and was doing the work in-house."

Millat's equipment now includes fourteen numerical controlled (NC) and computer controlled (CNC) machine tools. Parts range from machined stainless steel valve bodies for hydraulic components to helicopter rotor parts and tiny sub-assemblies for nuclear devices. The company is skilled in machine tape programming. It is shop-floor technology, to be sure, but advanced technology nonetheless that must stay ahead to remain competitive in a most competitive market.

For Openers, Cans...

Another such firm that has parlayed machining expertise into an international phenomenon is nearby Dayton Reliable Tool & Manufacturing Co. Founder and chairman Ermal Fraze, an experienced machinist whose company specializes in serving the electrical and packaging industries, observes that there are 150 to 175 companies like his in the area. "Each one serves in its own way one or another of the major industries. Together we are a most valuable service making use of the latest technology in our industry to advance the technology of others."

Mr. Fraze long ago became interested in packaging, specifically, can making. "The public takes it for granted," he says, "but the can industry uses almost every technology." The only problem, as Mr. Fraze saw it, was that you needed another tool to open a can. So in 1962 he developed the ring-pull can opener and the machinery to make it. Mr. Fraze now has patents in 28 countries on the device and machinery. Some 35% of the company's total volume is exported.

By 1965, Mr. Fraze realized that the ring-pull tear strips were going to be a problem. So he invented the push-in pull top can. Meantime, his company also provides various types of special tooling for the electrical industry as well as electric motor and switchgear components.

Ernie Fraze thinks there is so much potential for new technology-based industries in Dayton, that together with others in the Greater Miami Valley Investment Service, a Small Business Investment Corporation (SBIC), he has pledged $10,000 toward a million dollar pot to invest in new technology startups. "It's seed money," he says. "If the thing takes off, the inventor can easily find more money. Meantime, he gets our advice for free."

Data Processing...

It is an odd coincidence that the Mead Corporation and NCR Corporation, together with at least two other older companies, Standard Register Corp. and Reynolds & Reynolds Co., are today making their greatest mark in an area where the Wrights first ventured—communications. Orville and Wilbur ran a successful newspaper and printing business before they took up bicycles.

Mead Corporation is one of the oldest continually operating enterprises now in Dayton, tracing its beginnings to 1846. As such, the forest products company has been a major influence in the area's business and community life. Mead's new world headquarters building dominates the downtown skyline and has helped spark a rebirth of the downtown area. Its thrust into electronic information storage, retrieval and reproduction has helped to encourage other firms to center more of their efforts in high technology in the Dayton area.

Few industries are more basic than paper and forest products. Yet paper is the very stuff of the industry that today demands most from advanced digital technology—communications. Mead has combined the two, first to improve its own operations and now to compete successfully in burgeoning new markets.

Says Chairman James W. McSwiney: "There are not many U.S. industries left that can still be cost-effective in international markets. Three which still can are forest products, the digital industry and the service industries. Mead has developed special expertise in all three."

So seven or eight years ago Mead began to move strongly into advanced technology that at first glance would seem far from its traditional business. It is a good example of rapid adaptation to a changing technological world by an old-line company.

The company set up an Advanced Systems Group to develop and market digital technology. A Division of this group, Mead Data Central, currently markets two very large data bases, LEXIS and NEXIS. Both are supplied out of a massive new computer center south of Dayton.

LEXIS provides on-line full-text computer-assisted legal research to law firms and accounting firms, courts, corporations and law schools. Users gain access to LEXIS through specially designed terminals also fabricated and marketed by MDC.

NEXIS, a computer-assisted news research service, provides instantaneous access to the full text of a wide variety of publications and wire services. The service is particularly useful to journalists, corporate planning and forecasting staffs, analysts, or anyone with a need to quickly retrieve general and business news information.

Mead has developed a digital computer-driven ink-jet printer that spits drops of ink at a fast-moving web of paper. The process, called Direct Imaging by Jet Ink Transfer, or DIJIT, can instantly print variable information at rates of 800 feet per minute. The system can reproduce words and line art in black and white and color. "In future," predicts Mr. McSwiney, "publications will not have to remake pages for regional editions. Computer-controlled ink-jet technology could conceivably print publications individually for each subscriber." What will work someday for publications works now for direct mail material, billings, labellings and the like.

Mead is committed to the design and manufacture of hardware and software systems needed to position the company at the leading edge of communications technology from data base to hard copy. The company recently formally opened a new facility in the area to assemble minicomputers, terminals and printers. "That facility is a clear indication that Dayton can attract qualified people," says Mr. McSwiney. "The future belongs to people who can communicate in a digital language that is truly universal. We feel our investment in digital technology provides us with a kind of 'Window on the Future.' "

Microelectronics on the Miami...

A lot of the microelectronic technology inside Mead's digital equipment could come from its neighbor, NCR Corporation, the largest company headquartered in the Dayton area.

Founded in 1884 by a descendant of one of the original settlers of Dayton, John Patterson, as the National Cash Register Co., NCR was long a leading company in data processing through its cash registers and other business machines. NCR has since become a leading developer and manufacturer of computers and of the most advanced microelectronic devices.

A sizeable part of the company's effort—development laboratories, designers, engineers, programmers, and manufacturing plants for data terminals and for microprocessors and other advanced devices—is in Dayton.

It might have been otherwise. A decade ago the $3-billion company faced a major transition in converting from electromechanical business machines to all-electronic products and had to run very fast to catch up with and get ahead of the bounding pace of technology. There was then pressure to locate all development activities where most computer technology was developing, such as California. But after looking hard at the company's deep roots in Dayton, at comparative working and living conditions, at the costs of a move in human and financial terms, NCR decided to stake a significant part of the company's future where it had built its past, in Dayton. It proved to be a wise decision. "We have higher yields than do many companies in Silicon Valley," remarks Richard M. Perrin, Director of Engineering at the company's Microelectronics plant in Dayton.

NCR today is an integrated computer systems manufacturer designing and building its own equipment from microelectronic chips to turnkey installations. Indeed, NCR is the world's second largest computer systems vendor, a major user of metal-oxide-silicon (MOS) technology and a large producer of MOS devices.

"We pioneered non-volatile memory and license it to others," says Mr. Perrin. "We hold the basic patents on Electrically Alterable Read-Only Memory (EAROM) as well as basic patents on liquid crystals."

The company, a large manufacturing entity, has 5,500 people in the Dayton area. About 350 are at the semicon-

$\sum_{0}^{\infty}$Tt=DAYTON: Formula for High Technology

ductor site whose size is now being doubled at a cost of $50 million.

Sticking to the Last...

Software is the key to the continuing success of yet another old-line Dayton company. Reynolds & Reynolds, a form manufacturer founded in 1866, has entered the digital age by sticking to the business it knows best. The company has built up expertise in the arcane world of automobile dealerships, providing the paper forms by which dealers and manufacturers keep track of their dealings with one another and their customers.

As some dealers and the factories turned to computers for information processing, Reynolds & Reynolds stayed out front with systems and forms. Soon the company entered the computer business and started supplying dealers with mini-computers and specially-designed software. "We went into data processing to protect our forms business," says Peter Granson, Vice President, Marketing, "thanks to the foresight of people in the company as far back as the early 60s who saw the handwriting on the wall."

The company is now an intelligent terminal manufacturer, making about 160 a month and offering a complete line. But its strength and its view of the future is software developed specifically for vertical markets. Today, it's auto dealers. Tomorrow, it's various professional groups, from veterinarians to lawyers, doctors and other vertical areas. To meet that need, the $200-million company, whose sales have tripled in five years, has a relatively large staff of 116 software people.

Automobiles, Number One...

There was no automotive industry when Reynolds & Reynolds started. But automobiles followed close upon the aircraft industry and eventually displaced it.

G.M.'s Dayton Components Divisions (Delco Products, Delco Moraine, Delco Air Conditioning and Inland Division) develop materials for vehicles. Of late, the emphasis has been on less weight, greater efficiency and safety.

Out of Inland's extensive research and development facilities have come not merely revised lightweight components but wholly new designs that incorporate new materials and technology. New Corvettes will have a rear spring made of fiber composites that weighs 33 pounds less than the metal leaf spring it replaces.

At Delco Products, ferrite permanent magnets replace wire-wound field stators in motors and actuators, saving many pounds.

New designs are at the edge of automotive technology. So are the means used to develop them. In common with the microelectronics industry, which has had to go to elaborate computer systems to design new circuits, the auto industry has been a leader in computer-aided design (CAD). At Inland, engineers at terminals are on-line to central data bases in Detroit and can instantly call upon company wide resources in pursuit of a new design or revision of an old one. It helps a lot in redesigning, if not reinventing, the wheel.

Chemical Technology...

There is a substantial chemical presence in Dayton.

Monsanto Co. has headquartered its Monsanto Research Corporation here. The company began life as a private laboratory in the 1930s which, among other things, "worked on the artificial aging of whiskey right after Prohibition," according to Mr. Richard K. Flitcraft, President, and also Chairman of the Dayton Development Council's Task Force on Science and Technology. The nearby Mound Facility, which the Research Corporation operates for the U.S. Department of Energy (DOE), also reports to him.

The Research company's work is largely environmental with heavy emphasis on the life sciences. "It's a hands-on job," says Mr. Flitcraft, "involving a lot of field work. We are developing and providing environmental services for a wide range of government and industry and building the technology to clean waste water streams." That technology covers some of the most sophisticated analytical methods available, capable of detecting materials in parts per billion or less.

At Mound Facility, originally part of the Manhattan Project that developed the atom bomb, much work remains weapons related. But there is also basic research for many DOE projects. Among these are some of the earliest work on solar pond energy recovery, underground coal gasification and a feed system for a fusion energy project. Nuclear fuel processing, safeguards and waste technology also get close attention.

The Educational Base...

Advanced technology industry works closely with local schools to an unusual degree. The community, the companies and the schools know that the area's advanced technology base must rest on the availability and quality of higher education facilities.

Sinclair Community College is a downtown school that has grown startlingly in the past few years. The two-year college founded in 1887 had 3,000 students in 1966. It now has 17,000. "We turn out technicians who are scientifically and linguistically competent to work with engineers and scientists," says President Dr. David Ponitz.

Dr. Ponitz is acutely aware of the region's need to attract more advanced technology industry and offers the facilities of the schools to help. "If a

Computer banks at Mead Corporation's Mead Data Control Computer Center near Dayton provide storage and processing for on-line full-text information systems, LEXIS and NEXIS. Users easily access the base from remote locations through Mead-marketed interactive terminals similar to that shown.

$\sum_0^\infty Tt$ = DAYTON: Formula for High Technology

company comes to Dayton looking for people," he promises, "Sinclair will provide the training support either through courses we already have or new courses that we will tailor to the need."

The oldest university in the area is the University of Dayton founded in 1850. It is a full University on a 76-acre campus offering degrees from the associate through the doctoral level. There are currently nearly 11,000 students in 115 programs of study. There is also a Co-op program that currently has over 300 students at some 100 companies. The School of Engineering offers programs in chemical, civil, electrical, aerospace, materials and mechanical engineering as well as engineering management and programs in chemical, electronic and mechanical engineering technology.

The University established a Research Institute in 1956 though sponsored research for outside clients began in 1949. Annual volume of research has grown from $8,000 to nearly $20 million.

The University is presently engaged in more than 200 funded research projects. These cover a range of advanced technology. A ceramic bone implant material that virtually duplicates natural bone was developed here. Research is being conducted in such diverse areas as vibration damping, technology forecasting, laser holography, Raman spectroscopy for the analysis of ion-implanted materials in microelectronic devices, safe disposal of hazardous and toxic wastes, rare earth magnets and aircraft canopy analysis and design.

University President Raymond L. Fitz, S.M., himself an electrical engineer, characterizes the school as "a significant support to the area's technological growth due to its extensive research capabilities and its role as a leading educator of scientists and engineers." Last year the University graduated almost 400 engineers at all levels.

The other major local university is. Wright State University, near WPAFB. Founded in 1964 out of a joint program between Ohio State and Miami Universities, which had been offering courses in the area, the new school now occupies a 618-acre campus that seems constantly abuilding to accommodate its 15,000 students.

The community has made a conscious effort to avoid wasteful duplication in its universities. The University of Dayton and Wright State complement one another. Wright State has a new medical school. Its engineering school offers all degree levels, but is more systems oriented. Programs include a strong one in biomedical engineering and one of the largest computer engineering science programs in the Midwest.

The University biomedical activities include cooperative research programs with the local medical technology community, the National Aeronautics and Space Administration, the National Institutes of Health, the United States Air Force and the U.S. Army.

"Biomedical engineering is a new field that has grown out of the explosion in electronic and materials technology," remarks Prof. Jerrold Petrofsky. "Here, we are working on neural prosthetics, artificial stimulation of muscles. We hope to develop a microprocessor-based system that can electronically bypass a damaged spinal cord and allow the handicapped to walk or move limbs." WPAFB also benefits and contributes. "Muscle modelling has given us new insights into muscle movement control patterns that the Air Force finds helpful in their studies of pilots.

The Air Force Institute of Technology (AFIT) is also an important element in the local educational matrix. AFIT is the center for all Air Force advanced education. It monitors officers who are attending civilian schools elsewhere and handles their subsequent placement.

The region has ready access to some fourteen higher educational institutions. These form a consortium that includes the Air Force Institute of Technology at WPAFB, independent universities and various units of the State University system.

Technician at NCR Corporation Dayton-area microelectronics manufacturing facility places semi-conductor grade silicon wafers in processing equipment at end of cleaning cycle. Hood and smock worn by all employees in area help maintain clean-room standards and reduce chance of inadvertent contamination.

A Heartland Base...

Dayton is the sort of place that engineers and scientists and their employers find comfortable. Commuting is easy, swift and uncrowded. Taxes are 13th lowest per capita in the nation. Housing costs are far below those to be found on either coast or most larger metropolitan areas. Anyone house hunting today will recognize that an average home price of under $60,000 is almost a giveaway.

With an annual budget for the arts of $1.5 million, Dayton has active and first rate theater and music groups. The city is a frequent showcase for major performers on tour.

For local sports, there is University of Dayton Division I basketball with a tradition of competitive excellence on a national level. Wright State University consistently produces high-ranked basketball teams in Division II. For professional sports, less than an hour's drive is Cincinnati with its major-league teams. As a resident says: "It's darn nice to have Cincinnati and Columbus as suburbs."

Government is stable, participatory and relatively non-partisan. It was also pioneering. Dayton was one of the first cities in the nation to have a city manager. Elected commissioners govern the counties.

Government, industry and people are committed to preparing the region as a hospitable and productive base for tomorrow's highly technological industries and for the people who will run them.

Hoblitzelle Prize in 1958 for transferring leaf-rust resistance to wheat from the wild grass *Triticum umbellulatum*. Sears is currently spending six months as a visiting researcher at the Waite Institute in Adelaide, Australia.

JERRY A. FODOR ("The Mind-Body Problem") is on the faculty of the Massachusetts Institute of Technology. He got his B.A. at Columbia College and his Ph.D. in philosophy from Princeton University. He writes: "I have been a member of the faculty at M.I.T. more or less continuously since 1960, and I now hold a joint appointment as professor in the department of psychology and in the (recently formed) department of linguistics and philosophy. I have written books and articles in all three fields. My current interests include (in approximately reverse order) problems in the theory of meaning, experimental issues in psycholinguistics and cognitive psychology, problems in the philosophy of mind, and sailboats."

TOYOICHI TANAKA ("Gels") is associate professor of physics at the Massachusetts Institute of Technology. A native of Japan, he was graduated from the University of Tokyo in 1968 and stayed on to acquire his M.A. in 1970 and his D.Sc. (in physics) in 1972. He went to M.I.T. as a research fellow in biophysics in 1972 and has been there ever since. In addition to his work on laser-scattering spectroscopy and the study of phase transitions in gels he is interested in the application of such research to the understanding and treatment of diseases of the eye. Since 1973 he has held the concurrent position of research associate in medical physics of the Retina Foundation.

DAVID A. HOUNSHELL ("Two Paths to the Telephone") is assistant professor of history at the University of Delaware and curator of technology at the Hagley Museum in Wilmington, Del. He writes: "I grew up in the oil fields of southeastern New Mexico, and while I was in high school and during the first two years of college at Southern Methodist University I worked in the summers and on holidays for an electrical contractor. There I learned how to do everything from bending conduit to designing automatic-control units. It was a great experience and was the reason I pursued a degree in electrical engineering at S.M.U. I was bitten by the history bug, however, and later I found that I could understand the big questions about technology and society better through the study of history than through routine engineering work." Hounshell went on to acquire his graduate degrees (M.A., 1975, Ph.D., 1978) in history from Delaware, where he has been a member of the faculty since 1979.

METAMAGICAL THEMAS

An anagrammatic title introduces a new contributor to this column

by Douglas R. Hofstadter

I never expected to be writing a column for *Scientific American*. Let me say at once that I am not replacing Martin Gardner; no one could do that. Martin is, however, retiring at the end of this year. Until then he and I shall be sharing this space.

I remember once, years ago, wishing I were in Martin's shoes. It seemed exciting to be able to plunge into almost any topic one liked and to say amusing and instructive things about it to a large, well-educated and receptive audience. The notion of doing such a thing seemed ideal, even dreamlike.

Over the next several years, by a series of total coincidences (which turned out to be not so total), I met one after another of Martin's friends. First it was Ray Hyman, a psychologist who studies deception. He introduced me to the magician Jerry Andrus. Then I met the statistician and magician Persi Diaconis and the computer wizard Bill Gosper. Then came Scott Kim, and soon thereafter the mathematician Benoît Mandelbrot. All of a sudden the world seemed to be orbiting Martin Gardner. He was at the hub of a magic circle, people with exciting, novel, often offbeat ideas, people with an imagination of many dimensions. Sometimes I felt overawed by the whole remarkable bunch.

One day five years ago I had the pleasure of spending several hours with Martin in his house, discussing many topics, mathematical and otherwise. It was an enlightening experience for me, and it gave me a new view into the mind of a man who had contributed much to my own mathematical education. Perhaps the most striking thing about Martin to me was his natural simplicity. I have been told that he is an adroit magician. This is hard to believe, because one cannot imagine someone so straightforward pulling the wool over anyone's eyes.

I did not, however, see him do any magic tricks. I simply saw his vast knowledge and love of ideas spread out before me, without the slightest trace of pride or pretense. The Gardners—Martin and his wife Charlotte—entertained me for the day. We ate lunch in the kitchen of their cozy three-story house. It pleased me somehow to see that there was practically no trace of mathematics or games or tricks in their simple but charming living room.

After lunch we climbed the two flights of stairs to Martin's hideaway. With his old typewriter and all kinds of curious jottings in an ancient filing cabinet and his legendary library of three-by-five cards he reminded me of an old-time journalist, not of the center of a constellation of mathematical eccentrics and game addicts, to say nothing of magicians, antioccultists and of course the thousands of readers of his column.

Occasionally we were interrupted by the tinkling of a bell attached to a string that led down the stairs to the kitchen, where his wife could pull it to get his attention. A couple of phone calls came, one from the logician and magician Raymond Smullyan, another man I knew well by reputation but did not know belonged to this charmed circle. It was a most enjoyable day.

Martin's act, as they say, will be a hard one to follow. I shall not, however, be trying to be another Martin Gardner. I have my own interests, and no one except Martin himself could have all his interests. Nevertheless, to express my debt to Martin and to symbolize the heritage of his column, I have kept his title "Mathematical Games" in the form of an anagram: "Metamagical Themas."

What does "metamagical" mean? To me it means "going one level beyond magic." There is an ambiguity here: on the one hand the word might mean "ultramagical," magic of higher order, yet on the other hand the magical thing about magic is that what lies behind it is always nonmagical. That's metamagic for you! It reflects the familiar but powerful adage "Truth is stranger than fiction." So my "Metamagical Themas" will, in Gardnerian fashion, attempt to show that magic often lurks where few suspect it, and by the opposite token, magic seldom lurks where many suspect it. Herewith, dear reader, I take my own plunge.

In his column for July, 1979, Martin wrote a kindly review of my book *Gödel, Escher, Bach: An Eternal Golden Braid*. He began the review with a short quotation. If I had been asked to guess what single sentence of my book he would quote, I would never have been able to predict his choice. He chose the sentence "This sentence no verb." It is a catchy sentence, but something about seeing it again bothered me. I remembered writing it one day, attempting to come up with a new variation on an old theme, but it did not seem as striking as I had hoped it would. After seeing it chosen as the symbol of my book I felt challenged. I said to myself that surely there must be much cleverer types of self-referential sentence. And so one day I wrote down some more self-referential sentences and showed them to friends, which began a mild craze in a small group. Here I shall present a selection of what I consider the cream of that crop.

I should not go further without explaining the term "self-reference." Self-reference is ubiquitous. It happens every time anyone says "I" or "me" or "word" or "speak" or "mouth." It happens every time a newspaper prints a story about reporters, every time someone writes a book about writing, designs a book about book design, makes a movie about movies or writes an article about self-reference. Many systems have the capability to represent or refer to themselves somehow, to designate themselves, or elements of themselves, within the system of their own symbolism. Whenever this happens, it is an instance of self-reference.

Self-reference is often erroneously taken to be synonymous with paradox. This probably stems from the first famous example of a self-referential sentence, the Epimenides paradox. Epimenides the Cretan said, "All Cretans are liars." I suppose no one today knows whether he said it in ignorance of its self-undermining quality or for that reason. In any case two of its descendants, the sentences "I am lying" and "This sentence is false," have come to be known as the "Epimenides paradox" or the "liar paradox." (The latter term is more descriptive, the former more orotund. I shall use both.) Both sentences are absolutely self-destructive little gems and have given self-reference a bad name down through the centuries. When people speak of the evils of self-reference, they seem to be overlooking the fact that not every use of the pronoun "I" leads to paradox.

Let us use the Epimenides paradox as our jumping-off point into this fascinating land. There are many variations on the theme of a sentence that somehow undermines itself. Consider these two:


STOP
STOP

"This sentence claims to be an Epimenides paradox, but it is lying."

"This sentence contradicts itself—or rather—well, no, actually it doesn't!"

What should one do when told, "Disobey this command"? In the following sentence the Epimenides quality jumps out only after a moment of thought: "This sentance has three erors." There is a delightful backlash effect here.

Kurt Gödel's famous incompleteness theorem in metamathematics can be thought of as arising from his attempt to replicate as closely as possible the liar paradox in purely mathematical terms. With marvelous ingenuity he was able to show that in any mathematically powerful axiomatic system *S* it is possible to express a close cousin to the liar paradox, namely, "This formula is unprovable within axiomatic system *S*."

In actuality the Gödel construction yields a mathematical formula, not an English sentence; I have translated the formula back into English to show what he concocted. Astute readers may have noticed that strictly speaking the phrase "this formula" has no referent, since the English version is no longer a formula!

If one pursues the idea, one finds that it leads into a vast space. Hence the following brief digression on the preservation of self-reference across language boundaries. How should one translate the French sentence "Cette phrase en français est difficile à traduire en anglais"? Even if you do not know French, you will understand the problem by reading a translation: "This sentence in French is difficult to translate into English." To what does the subject in the latter sentence refer? If it refers to the English sentence, then the subject is self-contradictory, making the English sentence false (whereas the French original was true and harmless); if it refers to the French sentence, then the self-reference is gone. Either way something quite disquieting has happened, and I should point out that it would be just as disquieting, although in a different way, to translate it: "This sentence in English is difficult to translate into French." Surely you have seen Hollywood movies set in France in which all the dialogue, except for an occasional "Bonjour" or similar phrase, is in English. What happens when Cardinal Richelieu wants to congratulate the German baron for his excellent command of French? I suppose the most elegant solution is for him to say, "You have an excellent command of our language, mon cher baron," and leave it at that.

But let us undigress and return to the Gödelian formula and focus on its meaning. Notice that the concept of falsity has been slipped into the more rigorously understood concept of provability. The logician Alfred Tarski pointed out that it is in principle impossible to translate the liar paradox exactly into any rigorous mathematical language, because if it were possible, mathematics would contain a genuine paradox—a statement both true and false—and would come tumbling down.

Gödel's statement, on the other hand, constitutes a hair-raisingly close approach to paradox, yet it is not paradoxical. It turns out to be true, and for this reason it is unprovable in the axiomatic system. The revelation of Gödel's work is that, in any mathematically powerful axiomatic system that is consistent, an endless series of true but unprovable formulas can be constructed by the technique of self-reference, revealing that somehow the full power of human mathematical reasoning eludes capture in the cage of rigor.

In a discussion of Gödel's proof the philosopher W. V. Quine invented the following way of explaining how self-reference could be achieved in the rather sparse formal language Gödel was employing. Quine's construction yields a new way of expressing the liar paradox. It is this:

"'yields falsehood when appended to its quotation.' yields falsehood when appended to its quotation."

The sentence describes a way of constructing a certain typographical entity, namely a phrase appended to a copy of itself in quotes. When you carry out the construction, however, you see that the end product is the sentence itself. (There is a resemblance here to the way self-replication is carried out in the living cell.) The sentence asserts the falsity of the constructed typographical entity, namely itself (or an indistinguishable copy of itself). Thus we have a less compact but more explicit version of the Epimenides paradox.

It seems that all paradoxes involve, in one way or another, self-reference, whether it is achieved directly or indirectly. And since the credit for the discovery—or creation—of self-reference goes to Epimenides the Cretan, we might say: "Behind every successful paradox there lies a Cretan."

On the basis of Quine's clever construction we can create a self-referential question:

"What is it like to be asked, 'What is it like to be asked, self-embedded in quotes after its comma?' self-embedded in quotes after its comma?"

Here again the reader is asked to construct a typographical entity that turns out, when the appropriate operations have been performed, to be identical with the set of instructions. This self-referential question suggests the following puzzle: What is a question that can serve as its own answer? Readers might enjoy looking for various solutions to it.

When a word is used to refer to something, it is said to be being used. When a word is quoted, however, so that one is examining it for its linguistic aspects, it

A self-referential drawing of a self-referential sentence ("This hand writing is two-dimensional")

is said to be being mentioned. The following sentences are based on this famous use-mention distinction:

"You can't have your use and mention it too."

"You can't have 'your cake' and spell it 'too.'"

"'Playing with the use-mention distinction' isn't 'everything in life, you know.'"

"In order to make sense of 'this sentence,' you will have to ignore the quotes in 'it.'"

"This is a sentence with 'onions,' 'lettuce,' 'tomato' and 'a side of fries to go.'"

"This is a hamburger with vowels, consonants, commas and a period at the end."

The last two are humorous flip sides of the same idea. Here are two rather extreme examples of self-referential use-mention play:

"Let us make a new convention: that anything enclosed in triple quotes, for example '"No, I have decided to change my mind; when the triple quotes close, just skip directly to the period and ignore everything up to it,"' is not even to be read (much less paid attention to or obeyed)."

"À ceux qui ne comprennent pas l'anglais, la phrase citée ci-dessous ne dit rien: 'For those who know no French, the French sentence that introduced this quoted sentence has no meaning.'"

The bilingual example may be more effective if you know only one of the two languages involved. Finally, consider this use-mention anomaly:

"i should begin with a capital letter."

That is a sentence referring to itself by the (mauled) pronoun "I" instead of through a pointing-phrase such as "this sentence"; such a sentence would seem to be arrogantly proclaiming itself to be an animate agent. Another example would be "I am not the person who wrote me." Notice how easily we understand this curious nonstandard usage of "I." It seems quite natural to read the sentence this way, even though in nearly all situations we have learned to unconsciously create a mental model of some person—the sentence's speaker or writer—to whom we attribute a desire to communicate some idea. Here we take the "I" in a new way. How come? What kinds of cue in a sentence make us recognize that when the word "I" appears, we are supposed to think not about the author of the sentence but about the sentence itself?

Many simplified treatments of Gödel's work give as the English translation of his famous sentence the following: "I am not provable in axiomatic system *S*." The self-reference that is accomplished with such sly trickery in the formal system is finessed into the deceptively simple English word "I," and we can—in fact we automatically do—take the sentence to be talking about itself. Yet it is hard for us to hear the following sentence as talking about itself: "I *already* took the garbage out, honey."

The ambiguous referring possibilities of the first-person pronoun are the source of many interesting self-referential sentences. Consider these:

"I am not the subject of this sentence."

"I am jealous of the first word in this sentence."

"Well, how about that—this sentence is about me!"

"I am simultaneously writing and being written."

This raises an entire set of possibilities. Couldn't "I" stand for the writing instrument ("I am not a pen"), the language ("I come from Indo-European roots"), the paper ("Cut me out, twist me and glue me into a Möbius strip, please")? One of the most involved possibilities is that "I" stands not for the physical tokens we perceive before us but for some more ethereal and intangible essence, perhaps the meaning of the sentence. But then what is meaning? The next examples explore that idea:

"I am the meaning of this sentence."

"I am the thought you are now thinking."

These lead to:

"I am thinking about myself right now."

"I am the set of neural firings taking place in your brain as you read the set of letters in this sentence and think about me."

"This inert sentence is my body but my soul is alive in your mind."

The philosophical problem of the connections among Platonic ideas, mental activity, physiological brain activity and the external symbols that trigger them is vividly raised by these disturbing sentences.

This issue is highlighted in the self-referential question, "Do you think anybody has ever had precisely this thought before?" To answer the question one would have to know whether or not two different brains can have precisely the same thought (as two different computers can run precisely the same program). Is a thought something Platonic, something whose essence exists independently of the brain it is occurring in? If the answer is "Yes, thoughts are brain-independent," then the answer to the self-referential question would also be yes. If it is not, then no one could ever have had the same thought before.

Certain self-referential sentences involve a curious kind of communication between the sentence and its human friends:

"You are under my control because you will read until you have reached the end of me."

"Are you the person who is writing me, or the person who is perceiving me?"

"You and I can have only one-way communication, because you are a person and I am a mere sentence."

"As long as you are not reading me, the fourth word of this sentence has no referent."

"The reader of this sentence exists only while reading me."

The last of the preceding group of sentences is a rather frightening thought!

"Hey, out there—is that *you* reading me, or is it someone else?"

"Say, haven't you written me somewhere else before?"

The first of the above two sentences addresses its reader; the second addresses its author. Here we have an author addressing a sentence:

"Say, haven't I written you somewhere else before?"

Many sentences include words whose referents are hard to figure out because of their ambiguity—possibly accidental, possibly deliberate:

"Thit sentence is not self-referential because 'thit' is not a word."

"No language can express every thought unambiguously, least of all this one."

Let us turn to a most interesting category, namely sentences that deal with the languages they are in, once were in or might have been in:

"When you are not looking at it, this sentence is in Spanish."

"I had to translate this sentence into

TALKING TIME

Now a breakthrough in Speech Technology gives Time A Voice.

A miniature marvel of microcomputer speech and timekeeping technology. It measures only 4½ x 7/8 x 2 3/8.

Rapidly emerging electronic progress in solid state speech synthesis brings micro computer generated speech to timekeeping.

In the fall of 1978, Texas Instruments engineers developed "Speak and Spell"™, an electronic spelling aid with speech generated entirely by a tiny electronic micro computer chip. In 1979, talking language translators were introduced.

Today, most semi conductor houses around the world have major research efforts in solid state speech synthesis. In the next few years many new speech products will reach the marketplace. Now there is Talking Time. One of the first to be introduced, it may be the most useful and unique.

No Recorder—No Pull String

Like R2D2 in Star Wars, Talking Time seems to spring to life when speech synthesis and quartz timekeeping technologies come together. There are no recorders or talking doll pull strings in Talking Time, only solid state electronics. Electronics that are on the leading edge of both technologies. The forerunner of things to come. Here's how it works.

Micro Computer Brain to Slight Japanese Accent

Humans produce speech in much the same way that Talking Time does. Our brain sends nerve impulses to our jaw and chest, for our mouth to move and our lungs to push air across our voice box. Talking Times' single microprocessor chip sends electrical impulses digitized into speech patterns to an audio loud speaker. The loud speaker vibrates and out comes a clear voice that says "Attention please, it's now six o'five, please hurry"—with a slight Japanese accent. Talking Time is made by Sharp Electronics, a leading electronics company in Japan. So finely processed is the speech synthesis that you will swear it seems to speak with a slight Japanese accent. Texas Instruments Speech products are said to have a Texas drawl. Perhaps it's our imagination. You can decide for yourself.

Single Microprocessor Chip generates signals that both speak with slight Japanese accent, and time with quartz accuracy.

No Heart Failure with this Alarm

An effective alarm with a snooze feature that reminds you any minute of a 24 hour day. First a chime to get your attention, then, "It's now six o'clock," followed by a rousing melody over Talking Time's loud speaker. After a 5 minute snooze, Talking Time warns you, "Attention please, it's now six o'five, please hurry." It warns you again after 10 minutes, loud enough to wake even a sound sleeper, but pleasant enough not to give you heart failure.

Desk Top Reminder

Talking Time is ideal as a personal reminder. A perfect gift for that busy executive who is engrossed in work and forgets the time. In the **Time Announcement Mode** you're reminded automatically, every 30 minutes by, "It's now two-thirty p.m." If you want more reminders, choose the **Elapsed Time Mode.** Select 1, 5 or 30 minute intervals. Talking Time will announce each interval with "Five minutes elapsed, then "ten minutes elapsed," etc. No more missed meetings or appointments—Talking Time is Here!

Jogging Companion

Talking Time is small enough to be a handheld or pocket jogging companion. In the **Stopwatch Mode,** it's a stopwatch with one second precision and automatic as well as manual operation. In manual it announces the elapsed time at the push of a button for 10 full hours. When in automatic, it announces the elapsed time automatically every 10 seconds. During your entire run, Talking Time keeps you informed. You never have to break stride to look at your watch. Comes with a hand strap so it's easy to carry.

Interval Timer to Space Work or Fun

Talking Time is great for a periodic check against how much time has elapsed on a given task. Speed up or slow down as necessary. Production supervisors, doctors, and engineers, anyone with critical time measured functions to perform, find it useful. Pilots can use it to call out elapsed time every 10 seconds while they shoot approaches. Talking Time has both performance and functions unheard of even one year ago. The voice adds usefulness, which we think you'll agree, can only be described as Revolutionary.

Superb Chronometer That Can Keep Its Mouth Shut

Talking Time is a great chronometer in its own right. Its famous quartz accuracy is unsurpassed by chronometers costing $200 or more, but with half the functions. It has a large easy-to-read liquid crystal display (LCD) with 8 digits. A welcome relief for anyone who has poor eyesight. With Talking Time you hear as well as see the time. If you want silence, however, just tell it to keep its mouth shut. It won't let out a peep. You can still use it for its unsurpassed timekeeping accuracy.

Talking Time is easy to set in all modes, since it talks to you as you go along. You can keep track of what you're setting and how to do it. It may not be foolproof, but it's close.

Talking Time is worth its price as a superb clock alone. Its uses seem endless. It may add a whole new dimension to time. But don't take our word for it. Find out for yourself!

30 Day No-Risk Trial

To prove to you that Talking Time is everything we say, we offer it for 30 days at no risk. If during 30 days from receipt, you are unhappy with Talking Time for any reason, return it for a prompt refund of your purchase price. What have you got to lose?

One Full Year Warranty

Talking Time is all solid state. It should give you years of trouble free operation. If, in the unlikely event, anything should go wrong, it is warranted for one full year by Sharp Electronics. Talking Time comes complete with 2 AA batteries and handstrap. Under normal use, batteries should last for one year or longer.

To Order

Just fill in the order form and send it along with check or money order to our address. For even faster service credit card customers can call our Toll Free number listed below. Order yours today!

MEDIA MARKETING
10155 Plano Road • Dallas, TX 75238

___Yes, send me Talking Time with a 30 day no-risk return at $89.95 plus $3.00 delivery.

___**SAVE $10.00** Send me two Talking Times at $84.95 ea. plus $2.00 ea. delivery.

___**SAVE $30.00** Send me three Talking Times at $79.95 ea. plus $1.50 ea. delivery

Texas residents add 5% sales tax.

Name ____________________
Address ____________________
City ____________________
State__________Zip __________

☐ Check of M.O. Enclosed

For even faster service, charge to:

☐ MC/Visa — Expires__________

Account #____________________
Signature ____________________

Order Toll Free **1-800-527-7066.**
In Texas call COLLECT: **214-349-3120. sct-011**

English because I could not read the original Sanskrit."

"The sentence now before your eyes spent a month in Hungarian last year and was only recently translated back into English."

"If this sentence were in Chinese, it would say something else."

,werbeH ni erew ecnetnes siht fI"
".siht ekil ti gnidaer eb d'uoy

The last two sentences are examples of counterfactual conditionals. Such a sentence postulates in its first clause (the antecedent) some contrary-to-fact situation (sometimes called a "possible world") and extrapolates in its second clause (the consequent) some consequence of it. This type of sentence opens up a rich domain for self-reference. Some of the more intriguing self-referential counterfactual conditionals I have seen are the following:

"If this sentence didn't exist, somebody would have invented it."

"If I had finished this sentence...."

"If there were no counterfactuals, this sentence would not be paradoxical."

"If wishes were horses, the antecedent of this conditional would be true."

"If this sentence were false, beggars would ride."

"What would this sentence be like if it were not self-referential?"

"What would this sentence be like if π were 3?"

Let us ponder the last of these (invented by Scott Kim) for a moment. In a world where π actually did have the value 3, you wouldn't ask about how things would be "if π were 3." You might well say "if π were 2" or "if π *weren't* 3." So one's first answer to the question might be this:

"What would this sentence be like if π weren't 3?"

But there is a problem. The referent of "this sentence" has now changed identity. Is it fair to say, then, that the second sentence is an answer to the first? It is a little like a person who muses, "What would I be doing now if I had had different genes?" The problem is that he would not be himself; he would be someone else, perhaps the woman across the street mowing her lawn. The pronoun "I" cannot quite keep up with such strange hypothetical world shifts.

There is an even worse problem with the counterfactual above. Changing the value of π is, to put it mildly, a radical change in mathematics, and presumably you cannot change mathematics radically without also radically changing the fabric of the universe within which we live. So it is quite doubtful that any of the concepts in the sentence would make any sense if π were 3 (including the concepts of "π," "3" and so on).

Here are two more counterfactual conditionals to put in your pipe and smoke:

"If the subjunctive was no longer used

in English, this sentence would be grammatical."

"This sentence would be seven words long if it were six words shorter."

These two admirable examples, invented by Ann C. Trail (who is also responsible for others in this column), bring us around to sentences that comment on their own form. These sentences are quite distinct from ones that comment on their own content (such as the liar paradox, or the sentence that says "This sentence is not about itself but about whether it is about itself"). It is easy to make up a sentence that refers to its own form, but it is hard to make up an interesting one. Here are a few more good ones:

"because I didn't think of a good beginning for it."

"This sentence was in the past tense."

"This sentence has contains two verbs."

"a preposition. This sentence ends in"

"In the time it takes you to read this sentence, eighty-six letters could have been processed by your brain."

David Moser, a music student at Indiana University, is a specialist in self-reference of all kinds. (He has written a story in which every sentence is self-referential.) It might seem unlikely that in such a tiny domain individual styles could arise and flourish, but Moser has developed a self-referential style quite his own. As a friend (or was it Moser himself?) wittily observed, "If David Moser had thought of this sentence, it would have been funnier." Many Moser creations have been used above. Some further Moserian delights are these:

"This is not a complete. Sentence. This either."

"This sentence contains only one non-standard English flutzpah."

"This gubblick contains many nonsklarkish English flutzpahs, but the overall pluggandisp can be glorked from context."

"This sentence has sofa six words."

In my opinion it took quite a bit of flutzpah to just throw in a random word so that there *are* sofa six words in the sentence. That idea inspired the following: "This sentence has five (5) words." A few more miscellaneous Moser gems follow:

"This is to be or actually not two sentences to be, that is the question, combined."

"It feels so good to have your eyes run over my curves and serifs."

"This sentence is a !!!!! premature punctuator"

Sentences that talk about their own punctuation, as the last one does, can be quite amusing. Here are two more:

"This sentence, though not interrogative, nevertheless ends in a question mark?"

"This sentence has no punctuation semicolon the others do period"

Another ingenious inventor of self-referential sentences is Donald Byrd, several of whose sentences have been featured here. He too has a characteristic way of playing with self-reference. Two of his sentences follow:

"This here sentence don't know english to good."

"If you meet this sentence on the road, erase it."

The latter alludes, in its form, to the Buddhist saying, "If you meet the Buddha on the road, kill him."

Allusion through similarity of form is another rich vein of self-reference, but unfortunately I can give only two further examples. The first is "This sentence verbs good, like a sentence should." Its primary allusion is to the famous slogan "Winston tastes good, like a cigarette should," and its secondary allusion is to "This sentence no verb." The other example involves the following lovely self-referential remark, once made by the composer John Cage: "I have nothing to say, and I am saying it." This allows the following twist to be made: "I have nothing to allude to, and I am alluding to it."

Some of the best self-referential sentences are short but sweet. Here are five of my favorites, which seem to defy other types of categorization:

"Do you read me?"

"This point is well taken."

"You may quote me."

"I am going two-level with you."

"I have been sentenced to death."

Surely no column on self-reference would be complete without including a few good examples of self-fulfilling prophecy. Here are a few:

"This prophecy will come true."

"This sentence will end before you can say 'Jack Rob"

"Surely no column on self-reference would be complete without including a few good examples of self-fulfilling prophecy."

"Does this sentence remind you of Agatha Christie?"

This last sentence is intriguing. Clearly it has nothing to do with Agatha Christie, nor is it in her style, and so the answer ought to be no. Yet I'll be darned if I can read it without being reminded of Agatha Christie! (And what is even stranger is that I don't know the first thing about Agatha Christie.)

As I have indicated, the sentences featured in this column were invented by many people, not all of whose names I know. In addition to those already mentioned the inventors include Robert Filman, Margaret Minsky and me. Readers are invited to submit their own self-referential concoctions. Warning: The habit can become addictive!

In closing I cannot resist the plea of the following sentence: "Please publish me in your collection of self-referential sentences."

ISRAELI TECHNOLOGY TODAY

The Jewish state in the Middle East has had an impact on world markets out of all proportion to its size and circumstances thanks to the innovative ingenuity of its people. In this special report, Michael Knipe of The Times of London – recently the paper's Jerusalem Correspondent – examines the scientific and technological advances that are Israel's life-line to economic stability.

With its heavy defense burden, substantial welfare commitments and expensive import requirements, Israel faces formidable economic problems. It currently has the dubious distinction of having the world's highest annual rate of inflation – a massive 134 per cent.

And yet, through its scientific and technological expertise, the country displays a striking economic vitality.

In the first eight months of last year – a period of slowdown in world trade – export earnings increased by 26.7 per cent.

Israel's life-line to economic stability and progress is its remarkable capability for innovative research and development. In spite of the country's small size, its youth and its situation, far from the market places of the western world, it has established itself among the world leaders in several key areas of sophisticated exports. Particularly in the fields of electronics, agro-technology, computer graphics and software, and medical engineering, its scientists and engineers are at the forefront of applied knowledge. A number of companies have scored notable achievements internationally. Among them:

- Laser Industries Ltd., of Tel Aviv, a company that pioneered the development of laser beams for medical usage, has captured 60 per cent of the world market for CO_2 medical lasers.
- Sci-Tex Corporation of Herzylia, which designs and markets interactive fully computered prepress graphic design systems, has established itself in a series of market niches otherwise dominated by much larger American and European firms.
- Elscint, a Haifa based company that specializes in advanced medical imaging, has won 12 per cent of the world market in computerized axial tomography scanners competing against such multi-national conglomerates as General Electric, Siemens, Philips, Toshiba and Hitachi.

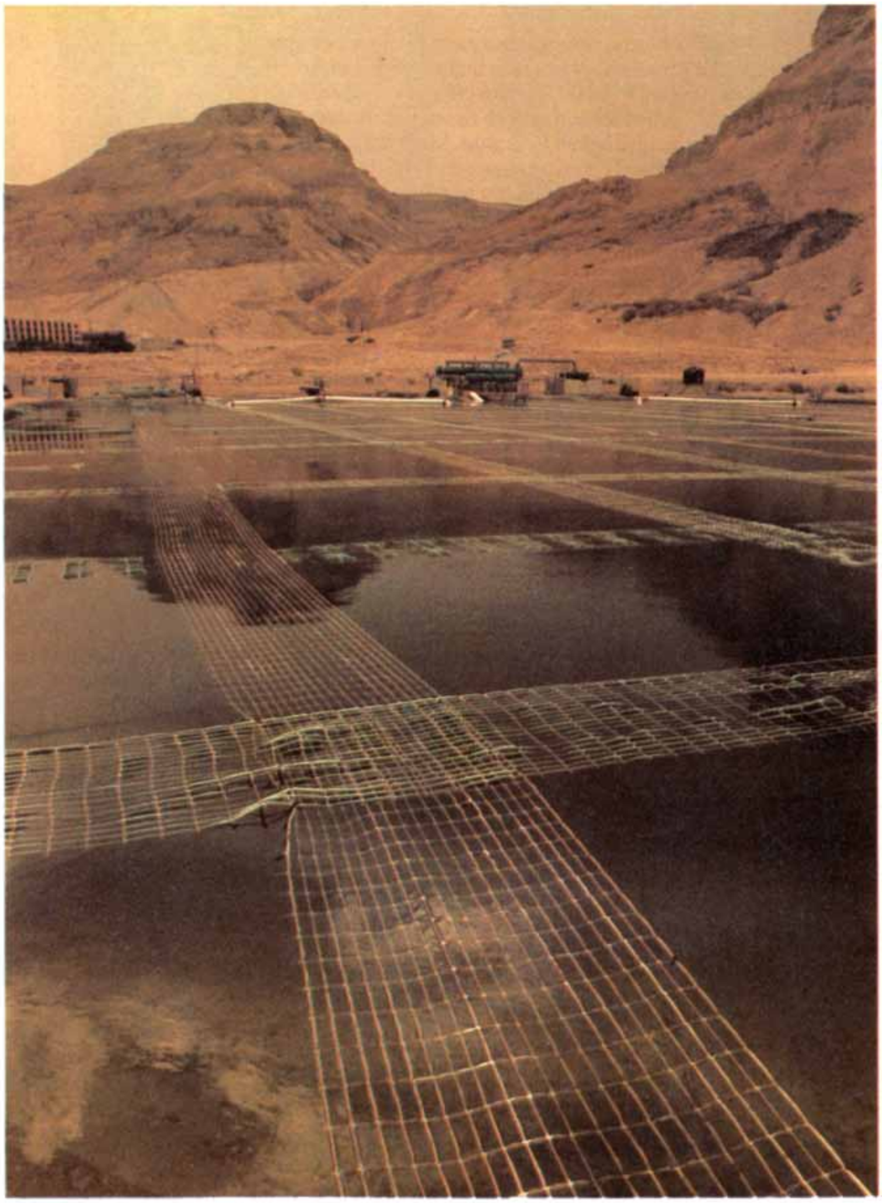
That such pre-eminence has been established by these concerns and many others, within the constraints of Israel's practical circumstances is impressive by any standards. In an arid climate, with hardly any raw materials and severely limited resources in terms of fertile land, water supplies and conventional energy sources, scientists and technologists have exhibited truly dramatic ingenuity.

They have developed life from the Dead Sea and are on the point of marketing structural components made out of desert sand. They lead the world in producing energy from the sun. They are even putting human foreskins – from eight day old babies – to good use as a means of mass producing the anti-infection and anti-cancer agent, interferon.

This year one third of the country's industrial exports – worth between $3-3.5 billion – comprise products developed in Israel. And there are at present more than 600 innovative products currently under development.

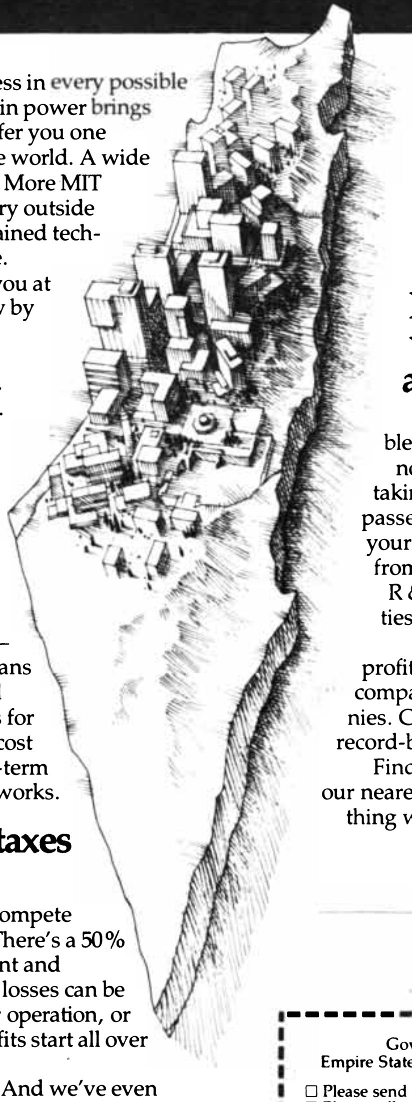
Mr. Gideon Patt, the Minister of Industry, Trade and Tourism, is justifiably proud of the special emphasis placed by the government on R&D and the fact that the country's sophisticated expertise is attracting significant foreign investment.

"Sometimes, when I meet with ministerial counterparts from other countries, they brag about their exports," said Mr. Patt. "But often they are talking about exports of raw materials. All they've had to do is dig them up



The solar pond at Ein Bokek, near the Dead Sea, provides the energy for the world's largest solar power station in operation. The netting is a floating wind break which minimizes the mixing of the convective and non-convective layers of the saline water. *(David Rubinger)*

THE SAME PEOPLE WHO BUILT A NATION CAN HELP YOU BUILD YOUR BUSINESS.



Israel's out to encourage business in every possible way, and it's no secret that sheer brain power brings a lot of investors to Israel. We can offer you one of the most skilled labor forces in the world. A wide knowledge of the English language. More MIT graduates per capita than any country outside the U.S. Some of the most highly trained technical and scientific people anywhere.

And they're ready to work for you at wage levels that are significantly low by Western European standards.

One of Israel's greatest natural resources is human resourcefulness. And every business could use some.

Israel wants your business enough to help pay for it.

To help you get your business going, we offer an extraordinary set of major financial incentives which significantly lowers costs and risks—and increases profits. Grants and loans that mount up to 75% of your initial capital investment. 50% cash grants for approved R & D projects. 90% low cost financing for working capital. Long-term leasing of industrial buildings. The works.

How to avoid paying taxes for years.

We've got a tax break that can compete with even the "tax free" countries. There's a 50% depreciation allowance on equipment and machinery, 20% on buildings—and losses can be carried forward. If you expand your operation, or reinvest in new facilities, your benefits start all over again.

Our corporate tax limit is 30%. And we've even got special concessions so you avoid taxes resulting from inflation.

Now you can sell to half a billion people duty free.

What you make in Israel gets into the common market duty free. And 2700 product categories get into the U.S. duty free. Plus you benefit from reduced tariffs in 8 other affluent markets. Including Japan.

You'll be right in the middle of the world market, at the crossroads of three continents. Israel offers international airports, seaports equipped to handle containers, highways—a modern infrastructure and an up-to-date communications network.

No other country in the world offers your business all these benefits.

When we say we're doing everything possible to encourage business, we mean it. There are no restrictions on bringing capital into Israel, or taking capital or dividends out in U.S. dollars. We passed the only law anywhere designed to protect your investment from taxation on profits resulting from devaluation. Plus we can offer you the superb R & D facilities of our world-renowned universities. And more.

It all adds up to one thing. Increased business profitability. And it's attracting giant multinational companies. Fast-growing high-technology companies. Companies like yours. Israel is experiencing a record-breaking boom in American investments.

Find out what Israel can do for your business. Call our nearest office, or send the coupon. We'll do everything we can.

And that's a promise.

Government of Israel Investment Authority
Empire State Building, 350 Fifth Avenue, New York, NY 10118

☐ Please send me more information about investing in Israel.
☐ Please call me.

Name ______
Company ______
Title ______ Phone ______
Address ______
City ______ State ______ Zip ______

ISRAEL
THE PROMISED LAND FOR BUSINESS.

Offices of the Israel Investment Authority:

New York, N.Y. 10118 (212) 560-0610 · Los Angeles, Ca. 90048 (213) 658-8721 · Chicago, Ill. 60601 (312) 332-2160 · Montreal, Que. H3A1B9 (514) 288-9276 · Toronto, Ont. M5S1M8 (416) 961-1242 · Boston, Mass. 02116 (617) 426-3636 · Philadelphia, Pa. 19102 (215) 546-4300 · Houston, Tx. 77046 (713) 840-0510 out of the ground and ship them! And often their economies rest on just a few exports. Israel's export revenues derive from thousands of different items."

Mr. Patt discounted the significance of the high inflation rate. "We are as good economists as any others. We know how to fight inflation. You create a small recession. You agree to live with some unemployment. And before you know it, inflation goes down to half its size.

"We don't want to do that for various reasons. We want Israel to continue to attract immigrants and people will not come to a country where there is unemployment. Secondly, unemployment could create a brain drain and before we knew it we might lose our major asset – the wonderful minds we have in this country."

Mr. Patt said a contributing factor to Israel's inflation was the increase in defense costs and energy prices in the world. In 1978, he said, investments from abroad increased over the previous year by 102 per cent and in 1979 there was a 10 per cent increase. In 1980, the authorities were more selective in approving investments so there would probably be no increase over the previous year.

The government was fully conscious of the fact that R&D was the key element in eventually closing the gap in the country's balance of payments, said Mr. Patt. Israel, he pointed out, was self-sufficient in terms of its domestic civilian economy. The negative balance of payments – an annual shortfall of between $3-3.5 billion – was the result of defense requirements and the escalating cost of oil supplies. Oil in 1980 cost $2.25 billion – almost three times the previous year's figure.

If the momentum of increasing exports could be maintained, Mr. Patt foresaw an export figure at the end of the decade of over $3 billion annually of products developed and manufactured in Israel. In 1948 when the state of Israel was reborn, total exports of goods and services had amounted to only $52 million, of which $27 million came from citrus products. "Now, 32 years later, we are speaking about $6 billion of exports. That is unique."

It is indeed striking that a country with the same population as Philadelphia should be capable of maintaining seven universities, an aerospace industry, its own army, navy and air-force and to compete with the world leaders in a multitude of technologies.

For all its disadvantages in terms of physical resources, Israel has several significant factors favoring its future development.

The most important of them is its relatively highly skilled workforce and the national tradition of academic learning. "Our main natural resource is brain power," said one academic, simply.

With 10,000 scientists and 20,000 engineers spearheading its efforts out of a total population of 3.8 million and the backing of seven academic institutions, Israel has certainly established a considerable reputation for advancing technological processes.

There were nearly 3,000 qualified personnel working on R&D in industry last year. This was an increase of about 80 per cent over the number five years ago. It is hoped to continue increasing the number by about 16 per cent a year to 4,670 by 1983.

In many respects Israel straddles the developing and developed worlds. It has first-hand experience of many of the problems facing the backward nations and is well suited to finding the right solutions.

The man primarily responsible for encouraging and coordinating Israel's industrial R&D is Professor Arie Lavie, the chief scientist at the Ministry of Industry, Trade and Tourism.

"Our short range goal," he says, "is to enable Israeli industries to reach an export of $1.3 billion of sophisticated, locally developed products by 1982.

"From 1976 to 1979 such exports rose from $280 million to $783 million. We want this trend of rapid increase to continue at an annual rate of 30 per cent for the next 5-10 years so that by 1986 we can reach a level of more than $2 billion of exports based on local R&D. To achieve such a degree of industrialization efficiently, we have to embark on a program of accelerated effort."

Professor Lavie perceives the strategy of achieving this target to involve four key elements – developing specific know-how in particular fields of sophisticated technology, exploiting Israel's local circumstances to full advantage, providing government incentives to encourage R&D and seeking international cooperation in funding such advances. Professor Lavie believes that Israel's highly trained scientists and engineers are the country's most important natural resource. Combined with geographic, defense and demographic elements they have given the state specific advantages in certain technical fields which, he maintains, should interest American and European industry in joint ventures.

More than $10 million worth of joint industrial R&D projects have been signed with American companies and investors in the past year. Professor Lavie emphasizes that all types of foreign involvement are encouraged, including the establishment of wholly-owned subsidiaries whose R&D projects can benefit from the 50 per cent support offered by the government. The only limitation is that the results of the projects should be commercialized in Israel wherever economically justified.

The extent to which this strategy is being pursued can be vividly indicated with examples of innovative R&D efforts right across the spectrum of industrial endeavor.

MEDICAL ENGINEERING

Israeli surgeons and engineers have been at the forefront of developing the use of laser beams for surgical purposes.

When Professor Isaac Kaplan, a Tel Aviv surgeon collaborated with Mr. Uzi Sharon, a local laser engineer to conceive and produce a clinical CO_2 appliance in 1972, they were breaking new grounds in terms of medical technology. Manufacturing and marketing a device under the trade name Sharplan began in 1973 after one year of experimental operations and clinical evaluations. Now, after eight years of pioneering effort, Laser Industries Ltd., which manufactures the Sharplan lasers, is dominating the world market. Of the 25 CO_2 lasers sold in the United States every month, 15 are produced by the Israeli company.

The surgical value of the CO_2 laser is its ability to destroy tissue selectively and precisely. Its pinprick beam is entirely absorbed by water and the biological tissue against which it is applied is 70-90 per cent water. This allows the beam to vaporize living tissue at its focal point while leaving adjacent tissue practically unaffected. By focusing the laser's energy on a small spot of tissue and moving that spot along a line, the surgeon can cut to a predetermined depth. And by sealing as it slices, the laser's action reduces blood loss, providing immense advantages in terms of clear and dry operating field, increased sterility through lack of contact and reduced pain and trauma for the patient. The areas of most obvious surgical application are gynecology, ear-nose-throat, neurosurgery, burns, plastic and reconstructive surgery and orthopaedic operations.

The Sharplan lasers have several unique features. They incorporate a particularly effective, patented articulated arm which is notable for its lightweight maneuvrability. And they are the only units that offer both free-hand and micro-surgery capability with rapid interchange possible. Three units are currently in production, the least expensive and most popular of which costs hospitals approximately $50,000.

A fourth which will be marketed this year is specifically aimed at private clinics rather than hospitals. It will be particularly compact and mobile and relatively inexpensive, costing approximately $25,000.

Mr. David Meridor, the president and chief executive officer of Laser Industries, believes that one of the significant advantages the company has over its competitors is the close and informal cooperation practised between the surgeons and the engineers involved in the development of the units, together with the pioneering spirit of the company shareholders. They invested something like $5 million into the laser project over an eight year period before a profit was recorded for the first time last year.

Currently the company is manufacturing about 300 units annually and expects later this year to be producing 50 units a month. Each one is composed of about 1,200 parts, of which 1,000 have been especially designed by Laser Industries and only 10 per cent are imported from overseas.

Last year the company recorded sales worth $2.2 million giving a first time profit of $200,000. Mr. Meridor says this year's sales are expected to be approximately $6 million and next year may be over $12 million.

Even with these figures, Laser Industries believe they are only skimming the surface of potential sales.

"Most laser surgery so far has been carried out by pioneering specialists," says Mr. Meridor. "There are still many usages which have not yet been researched properly."

Major cancer surgery had been carried out using laser units for several years. There had been some extremely exciting operations. But it would still be some time before the results would be widely publicized.

"The real potential for laser surgery lies with the general surgeon," he says. "Unfortunately there is no great rush among general surgeons to exchange a scalpel for a medical laser unit, even when the laser would be advantageous. But we believe it is only a question of time before the value – medical and economical – becomes more widely realized." In the meantime, Laser Industries is plowing funds into its R&D department where several major projects are underway. These include research into various types of endoscopes which enable operations to be performed without surgically opening the patient's body.

In a non-medical laser field Laser Industries is collaborating with two U.S. companies in a three year research program to develop a high energy industrial laser for welding and metal working.

Elscint Ltd., of Haifa, has established itself uniquely, as a specialist in medical imaging. The company develops and manufactures equipment in three main imaging fields – nuclear medicine, computerized tomography and ultrasound – and is entering the newly emerging field of digital radiography which it believes will revolutionize the conventional

X-ray industry in the near future.

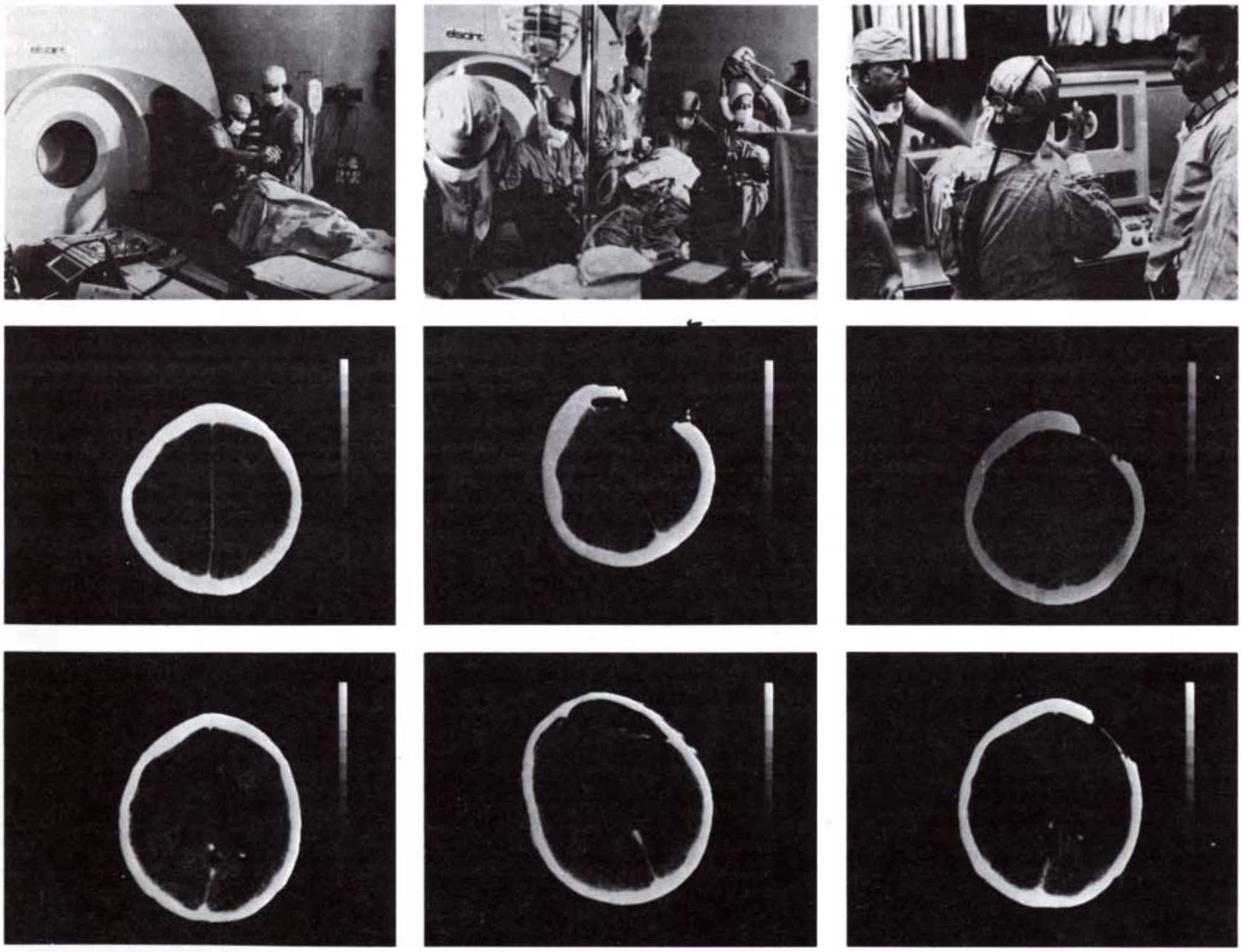
The company remains a small concern by comparison with the international conglomerates, General Electric and Siemens. It has captured 8 per cent of the nuclear medicine instrumentation business and 12 per cent of the computerized tomography and lies fourth internationally in these markets. Its target is to capture approximately 15 per cent of world sales in each field of activity – enough to be profitable without disturbing the giants – and it appears to be well on the way to doing so. Dr. Avraham Suhami, Elscint's president and chief executive, believes the way to achieve this is to concentrate on the increasing extent of common ground developing in the medical imaging business and the increasing know-how emanating from the company's particular degree of specialization in it.

It is not enough, he argues, to put together an activity of company X in nuclear medicine with company Y in ultrasound, to come up with effective imaging products, as some conglomerates have tried.

Computerized tomography (CT) is based on the principle that different types of tissue absorb different quantities of radiation. In CT, X-rays are measured by radiation detectors that, unlike the photographic film or phosphor screen systems used in conventional X-rays, capture virtually all X-rays and are able to transmit data to a computer for storage, subsequent processing and conversion to an image. The unique characteristic of the CT scanner is its ability to compile data from multiple viewing angles. The CT scanner provides clear images of planes within the body (including the interior of bones and organs) yielding information previously unobtainable with conventional X-ray equipment.

A CT unit scans the targeted 'slice' of the body from different angles with a beam of X-rays. Using a computer it then 'reconstructs' a density map of the cross section by mathematical analysis of the data obtained from the various angles. The resulting image is finally displayed on a monitor and may be readily transferred to hard copy. Important diagnostic advantages of CT include the possibility of rapid diagnosis of such problems as brain tumors, kidney stones, calcifications in arteries and liver lesions. CT virtually eliminates the problem of structural overlap and greatly improves resolution capability, compared with conventional two-dimensional X-ray views.

Elscint's CT scanner, the Exel 905, costs between $475,000 and $800,000 depending on the configuration and accessories. The company is now introducing a new model, the Exel 1002 which will utilize a substantially larger number of faster acting electronic detector devices. Users of the Exel 905 can upgrade their equipment to the Exel 1002 to obtain enhanced image quality at the five-second scan time of the 905 and comparable image quality at a scan time reduced to less than two seconds. The company is also introducing a low-cost CT unit, the Exel 820 based on the 905, for scanning only a patient's head. At present the company is filling orders for scanners at the rate of about five a month – more than 10 per cent of the world market.



INTRA-OPERATIVE COMPUTERIZED TOMOGRAPHY
A CT scan of a 57 year old male demonstrated a left frontal glioma (left column). To achieve a radical excision of the tumor the patient was operated on with intra-operative control by repeated CT examinations (middle column). The operation was performed while the patient was on the Elscint neuro-trolley, adjacent to the scanner, enabling precise repositioning of the scan plane. The operation was interrupted three times and the patient was rescanned to search for remaining tumor tissue. When no evidence of tumor residue could be detected in the CT images, ten biopsies were taken from the tumor bed. All, consequently, revealed no residue of tumor tissue. The patient was examined monthly by repeated CT scan. Right column demonstrates a scan taken recently, four months after operation; so far no signs of tumor recurrence could be detected. In the follow-up examinations, the clinical condition of the patient was found to be satisfactory.

Surgery was performed by Professor M. N. Shalit, M.D. at the CT Clinic operated by S. Matz, M.D. at Beilinson Hospital, Petah-Tikva, Israel.

For small hospitals lacking a CT specialist or a sufficient number of patients to justify having the entire system, the company has developed a cluster system in which one 'master' Exel 905 complete with reconstruction computer and full image processing capability can be supported by two or more 'slave' scanners which can acquire data in remote locations and convey it, electronically, to the 'master' system for image processing and evaluation. This system is also suitable for teaching hospitals and research institutions that want to separate daily routine CT examinations from tomography research activities.

Last year a U.S. company, Universal Inc. of Fort Lauderdale, Florida, purchased 10 Exel 905's for installation in air-conditioned trailers to provide CT facilities to several regional hospitals, clinics and private doctors.

Nuclear medical imaging provides cinematic pictorial information which gives detailed diagnostic information related to the function of organs rather than just their anatomy. Patients are injected with small quantities of radioactive isotope-tagged substances, known as radiopharmaceuticals, which diffuse into the organs or tissue targeted for study. The radiopharmaceuticals emit gamma rays which are detected by a gamma camera. The recorded image or sequence of images corresponds to the concentration and location of the radiopharmaceuticals within the organs or tissue as opposed to the density information provided by X-rays or computerized tomography. The maximum exposure of a patient to radiation occurs in nuclear cardiology but is less than one-tenth as great as the radiation incurred in the comparable conventional X-ray procedures. Images can be secured at the rate of 20-50 a second allowing physicians to follow the functions of the organ or tissue.

Measuring the cerebral blood flow with nuclear medicine techniques and correlating it with physiological parameters enables diagnosis of cerebral disorders. Through these techniques research is being done into the process of thinking. By tracking local changes in the blood supply and mapping the deviation from the normal flow, it is possible to obtain maps correlated with the process of thinking. Imaging the activity of the brain has shown, for example, that the map of 'anxiety' looks different from the map of 'fear.' The immense potential of these techniques for increasing understanding of the mind can hardly be overstated.

In Detroit last year, Elscint unveiled to the North American Nuclear Medicine Society, the world's first digital gamma camera, the Apex, which combines gamma ray detecting equipment and an image processing computer. It yields dynamic television images of any functioning organ selected for diagnosis with improved linearity, uniformity and spatial resolution, resulting in better images, and more quantitative information at higher count rates. Among ten or so variations on the Apex system, the Apex 215M is a manually driven, mobile, fully-integrated gamma camera-processor particularly suited for nuclear cardiology studies. Another, the 415M, is a large field mobile camera with the components mounted on a motor-driven vehicle suitable for serving several hospital departments or in cases where the patient cannot be moved. The price of the Apex models ranges from $135,000 to $245,000.

In ultrasound imaging, images are formed by focusing high frequency sound waves generated by a solid state transducer at the part of the body targeted for study. The returning echoes are converted into electrical signals which are processed and displayed. The advantages of ultrasound are that it is non-radioactive and low in cost, and thus useful in routine screening in obstetrics, gynecology, cardiology, study of the abdomen and ophthalmology.

Elscint is introducing this year a real-time ultrasound system for imaging the abdomen which will incorporate a modified version of the same computerized data processing system developed for the Apex nuclear medicine system.

Dr. Suhami estimates the world market for medical imaging to be in the region of $2.5 billion a year at present, with one-third of that in the U.S. Dividing it into the different modalities, he puts the nuclear medicine, ultrasound and computerized tomography each at around $300 million and the conventional X-ray market at between $1.2 billion and $1.5 billion.

"We view the field as one with a common know-how and, in many instances, common hardware and software," he says. "Accordingly, we have a well defined long term R&D program in order to be active in all the opportunity areas."

Dr. Suhami believes that as the diagnostic usefulness of imaging techniques becomes clearer to the practicing physicians and the medical community generally, their use will become indispensable and the costs of the equipment while not decreasing will be less of a problem. He suggests that the best method of affording the equipment will be on a monthly leasing arrangement in which the costs could be established per examination, payable either by insurance schemes or directly by the patients. As an example he suggests that people who suffer from chronic headaches would be willing to pay something like $200 to find out whether it was caused by a tumor.

Last year Elscint's sales increased by 44 per cent to above $31 million, primarily because of the success of the Exel 905, with orders worth more than $40 million.

To maintain its competitiveness, the company maintains an unusually large R&D group of 250 people and devotes more than five per cent of its revenues to their work, which is on a par with that of the multinationals.

Elscint's philosophy is that the most critical competitive factor is the quality of the image produced and it has demonstrated with its Exel and Apex range that it can provide the best resolution in the image at the lowest possible cost and with total reliability.

SOLAR ENERGY

In December 1979, Israel's Minister of Energy unveiled a 150 kW pilot solar-electric power station at Ein Bokek, a settlement on the shores of the Dead Sea, dramatically demonstrating, for the first time anywhere in the world, the feasibility of generating electricity from solar energy day and night, winter and summer.

Israelis have long been leaders in the manufacture and use of solar equipment for domestic water heating. Most apartment roofs carry a cylindrical tank heated by solar collector panels enabling a valuable saving of two per cent of the country's oil imports. New buildings are obliged to incorporate solar heating devices and Israeli companies have made deep penetrations in the still developing world market for such heaters.

The Ein Bokek operation was based on two original Israeli developments – a low temperature Organic Rankine Cycle turbine generator and a solar pond, produced respectively by Ormat Turbines Ltd and its associate company Solmat Systems Ltd, which Ormat established together with the Scientific Research Foundation of Israel to promote solar ponds.

The Ein Bokek plant is believed to be the largest operating solar power station, although some larger units of a different, power-tower type, are under construction. The pond, 1.8 acres in area, is somewhat smaller than that required to produce 150 kW on a 24 hour basis, but the fact that it provides 150 kW at will demonstrates the storage capability of solar ponds. This was further illustrated by the inauguration taking place at night!

The solar electric power generating station at Ein Bokek

"Since the inauguration we've managed to get 220 kW out of the plant so I think it is still the biggest solar power station in operation," says Mr. Lucian Bronicki, Ormat's president and technical director.

The plant uses a pool of salty water similar to the Dead Sea, where the density increases with depth. The energy from the sun penetrates the lighter upper layer and is retained as hot water at the bottom. The difference in salinity prevents heat loss through convection while at the same time the clear top layer acts as an insulator. The water, holding a temperature of about 80°C, is then used to power the low temperature turbine and supply electricity whenever it is wanted. The upper layer of water is used as a coolant.

Israel owes its pre-eminence in the field of solar energy to its first prime minister, David Ben Gurion, who, with characteristic far-sighted vision, recognized his country's need for cheap, self-sufficiency in energy and promoted research into solar energy sources. The program was curtailed when he left office, but hastily re-instituted when the oil crisis of 1973 made the need more obviously apparent. The development by Ormat of the low energy turbine generators took place when solar energy research had been abandoned.

"Engineers everywhere neglected turbines that would operate at low temperatures," explains Mr. Bronicki. "There was plenty of fuel and low temperature turbines were less efficient. You can't make a high temperature turbine work at low temperatures. But the low temperature turbine we developed will work with both solar energy and fossil fuels."

Using only one moving part, the turbine shaft itself, the generator is self-lubricating and requires no maintenance. It has proved ideal for remote locations such as light-houses, and micro-wave relay stations and in developing countries where there are short ages of skilled maintenance workers. Ormat has built and sold more than 3,000 of the units to 40 different countries, one of which, says Mr. Bronicki, has worked 12 years without stopping and without maintenance.

Currently the operation of the facility at Ein Bokek is being monitored and evaluated at varying system conditions. The next phase of the program, already underway, is to put into operation a five MW capacity plant, incorporating a 62-acre solar pond area by 1982.

The potential yields of energy from solar ponds are impressive because the ponds can be so large. Dr. Harry Tabor, director of the Scientific Research Foundation and one of the veteran solar-energy researchers, has calculated the yield of a 250 acre pond in a sunny climate such as Israel. For an annual insolation of 1700 kWh (thermal) per square yard and a pond collection efficiency of 20 per cent he says, the annual yield would be 400 million kWh – the equivalent of burning 43,000 tons of fuel oil at 80 per cent combustion efficiency. There would be enough built-in heat-storage to show almost no variation of output over a period of a week or so and even seasonal variations could be reduced by making the pond deeper. For ponds not less than 25 acres in area the latest estimate of cost is $11 per square yard including all plumbing but not the heat exchangers and assuming an expensive elastomer lining.

For sites where the ground is impervious, the lining could be omitted and the pond cost would drop by about 40 per cent. These costs assume free salt and no payment for the land.

A comprehensive study of the potential sites for solar ponds has not yet been completed. But, says Dr. Tabor, a study of 14 arbitrarily chosen sunny countries has shown that for power production alone – that is excluding heating applications – the potential to the end of the century is between 2,300 and 12,500 square miles. That is 30,000 MW to 160,000 MW total generating capacity.

Dr. Tabor believes that the non-convecting solar pond can be an effective method of harnessing solar energy in the multi-kilowatt to multi-megawatt range, thereby filling the gap between classical, small scale, solar collectors and the very sophisticated gigawatt range 'power-tower' and satellite concepts now being considered by some industrialized countries. This gap, he says, corresponds to the needs of many developing countries where the amounts of electricity required are less and where space may be available for pond sites.

As a source of low-temperature heat – up to 90°C – says Dr. Tabor, the estimates show calorie cost, for sunny climates, to be competitive with oil at $41 per ton. Electricity production in the 0.5-5 MW or larger ranges is possible – using low-temperature turbo-generators – at costs that are competitive with alternative supplies in areas not having a national electricity grid. Pond sites are limited roughly to plus or minus 40 degrees latitude and to places where salt or brine and water – not necessarily potable – is available. Thermal applications include heating and cooling of buildings and desalination.

By 1983 it is intended to complete a five MW power plant with a 60 acre pond on the shore of the Dead Sea. An additional experimental 25 acre solar pond is to be constructed and tested in parallel to this.

By 1984, an additional 20 MW power unit and an additional 250 acre solar lake is envisaged, with a further 1,000 acre solar lake area and a 50 MW power unit to be installed by 1985. In the following years, says Mr. Bronicki, a further series of 50 MW power units and accompanying solar lakes could be constructed in a manner best suited to Israel's national electricity grid needs for peaking, intermediate, and base load application. As the system expands towards the end of the century, a solar lake of about 195 square miles with a cluster of power plants of 50 to 1,000 MW could be in operation, reaching a total output of more than 2,000 MW. Israel currently uses some 1,500 MW and consumption is expected to rise in the mid 1980's to 2,250 MW.

According to Dr. Tabor, the five MW units will cost about $3,000/kW installed whilst the cost of 40 MW units will fall to $2,000/kW installed which, he says, would be comparable in price with hydro-electric and nuclear plant power.

COMPUTER GRAPHICS

Israel has made particularly striking advances in the developing spheres of computer software systems and electronics.

A prime example of how the national strength in this respect can be exploited is the Sci-Tex Corporation of Herzliya which designs, manufactures and markets interactive computer-aided design systems which use full color raster-format in pictorial image and graphic applications.

The company's systems automate the pre-press process in commercial color graphics reproduction. They can be used for printing magazines, catalogs, advertising material, flexible packaging materials, maps, textiles, and wall and floor coverings. The two principal units, the most sophisticated available, are the Response 200 and 300 systems which cost around $500,000 and $750,000 respectively.

Sci-Tex was responsible in 1971 for the earliest innovation of a computer-aided design system able to process color images. This was marketed in the knitting industry, greatly speeding up the process of converting designers' sketches into operating instructions for producing double-knit fabric. With it, designers were able to review and revise designs quickly and accurately and to estimate precisely the amounts of yarn required and the fabric costs. By the end of 1976, Sci-Tex had managed to sell more than 40 of the systems at prices ranging from $110,000 to $220,000.

The company's second generation system, the Response 200, was developed between 1973 and 1976. This used the technology of the initial system on a larger scale and with the addition of a laser-driven device to speed up the color separation process. Between 1974 and 1976 the company's sales nearly doubled from $4 million to $7.4 million.

In late 1976, Sci-Tex began to sell another new product, the Laser Plotter, priced at $125,000. The Plotter, an adaptation of the output unit of the Response 200, could be used for any purpose currently served by conventional computer output plotters and electronic drafting machines, such as the production of maps, printed circuit masters and diagrams of all sorts. It was suitable for engineering and architectural drawings, printed circuits and integrated circuits. It lent itself particularly to applications requiring large amounts of plotted data and had the unique advantage of large format, high resolution and data-independent speed.

The Response 300, a further refinement to the system, was marketed in 1979. The value of the system lies in the fact that it automates a series of processes involved in the preparation for printing of color pages which would take between 10 and 15 man-hours to complete by traditional manual techniques.

In this system the data concerning the color and intensity of each point in the picture is recorded in digital form, and is then displayed as arrays of points or pixels thus yielding the equivalent of a home television picture. This permits an operator sitting at a terminal to perform the tasks of image enhancement and page make-up much more quickly than previously possible.

By last year 50 Response 200 and eight Response 300 systems had been sold and revenue had increased to $14 million.

Mr. Ephraim Arazi, the dynamic founder and president of Sci-Tex, believes that in spite of the company's small size, it enjoys an unusually strong competitive edge in the markets it serves. This, he says, is based on technological expertise, particularly in software, which is constantly updated for Response owners, free of charge.

"We have a team here that is our greatest strength," he says. "Any of the international giants would be delighted to have our staff. Don't let anyone fool you. We did not invent new technology. We have used warmed-over military technology in industrial application and we have industrialized all sorts of life-line technology, the seeds from the sputnik and NASA space programs. There are two technologies where we lead the world. Lasers for output and interactive displays for super-quality graphic images, which is a growing field for such things as inter-active digital displays. We happen to excel in it. Of course, any of the giant companies have in their empire all the individual skills and technologies that we have. But to compete in our applications, they would have to put all those skills together and invest thousands more man-hours. And our R&D expenses are approximately $11 per hour while costs in the U.S. or Europe are about $40."

Sci-Tex is constantly adding and refining its system's capabilities. "We've just added word-processing and photo-type setting," says Mr. Arazi. "The results are stunning. Now the text elements in the pages rather than being scanned in can be brought in digitally by direct link, disc or facsimile transmission so that even though galleys never see the light of day until they are made into a plate or printing cylinder, they are seen by the computer screen."

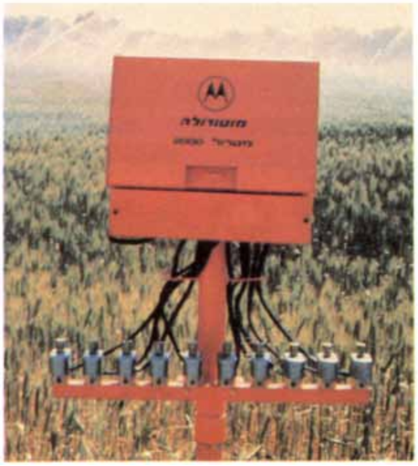
The company is now working on the fourth generation of the Response system. "It will," says Mr. Arazi, "be a more productive, a more advanced, a higher performance, and a higher priced unit."

Mr. Arazi believes Sci-Tex will remain a relative pygmy among the international giants of the computer-graphics world. "But the gaps we are filling are getting bigger all the time."

"I am an ex-military officer," he says, "and I never forget that targeting is the key element – matching the resources and anticipating the needs of industry."

AGRO-TECHNOLOGY

Agricultural research was a fundamental element in the Zionist philosophy and Israel is justly known for the extent to which the desert has been made to bloom. Since 1948 there has been a sixfold increase in the land under irrigation and optimum use is being made of the country's meager water resources. Research and development into irrigation techniques has produced an extensive list of sophisticated products with wide application. The use of either microprocessors or larger minicomputer systems, cleverly integrated with software and agro-technological know-how, is bringing computerization both to the smallest farming units and larger farms.



Motorola's computer-controlled irrigation system suitable for small farms

A unique computer-controlled irrigation system which processes operating data and implements accurate and reliable irrigation schedules for farms of under 250 acres has been developed by Motorola Israel Ltd. Called the Matarol 2,000, the system operates in a similar fashion to a pocket calculator. A compact keyboard and associated alpha-numeric display situated in the field enables easy operation which can be learned within an hour. The unit costs around $2,500 and the benefits include savings of up to 20 per cent in water, energy and manpower plus a higher yield of approximately 20 per cent.

Says Dr. Abraham Harel, Motorola's long-range planning and development manager, "The computerized irrigation system is a good example of how we in Israel can best exploit our advanced technological know-how in both the agrotechnology and electronic areas and there is immense potential in this area for further work."

Research is continuing into both how far computers can be used to sense out the presently unknown factors in farming and also the extent to which they can further simplify procedures for farmers lacking technological backgrounds.

CHEMICALS

A radical new technology for extracting calcium bromide from the Dead Sea is on the point of being put into commercial operation by Israel Chemicals Ltd. (ICL), the largest of the country's chemical conglomerates.

The Dead Sea provides a virtually inexhaustible supply of bromine as a result of which Israel is the world's second largest producer. And as the largest, the United States restricts its supplies to North America, Israel dominates the international market. With the uses of bromine becoming increasingly varied and extensive, ICL has invested heavily in bromine compound producing facilities and is expecting to reap significant returns. Production capacity now stands at 70,000 tons and is expected to expand within the next couple of years to 80,000 tons, with more than 90 per cent being exported.

The new extraction process, which bypasses the need for electrolysis and is thus a considerable energy-saver, is described by Dr. Uri Eisner, ICL's R&D director, as a novel and impressive technique which is likely to remain unique to Israel as it relies to a large degree on the specific brine composition of the Dead Sea. Another highly priced export resulting from an original Israeli exploitation of the Dead Sea's brine is a high-grade periclase for use in refractories, such as steel mill furnace brick linings. Magnesium chloride is converted to a particularly dense and pure form of magnesium oxide bringing in $25 million from the export of 50,000 tons a year.

World sales of ICL's products this year will be in the region of $600 million, all derived essentially from the country's few mineral resources – phosphate rock, a little low-grade copper in the Negev desert and the brines of the Dead Sea.

Potash exports stand at about 1.2 million tons a year and are scheduled to rise to 2 million tons by the end of this year.

At present the chemical industry contributes almost 50 per cent of Israel's industrial exports based on local R&D. Total chemical, plastics, rubber and oil products exports reached $726 million in 1979 an increase of 38 per cent on the previous year and by 1988 it is hoped to double them.

The industry's future is seen to lie in developing the most sophisticated chemicals – those selling for prices of between $5,000 and $50,000 a ton. The emphasis in development terms is thus being placed on flame

retardants, pesticides, pharmaceuticals, medical diagnostics, vaccines, and various other speciality chemicals.

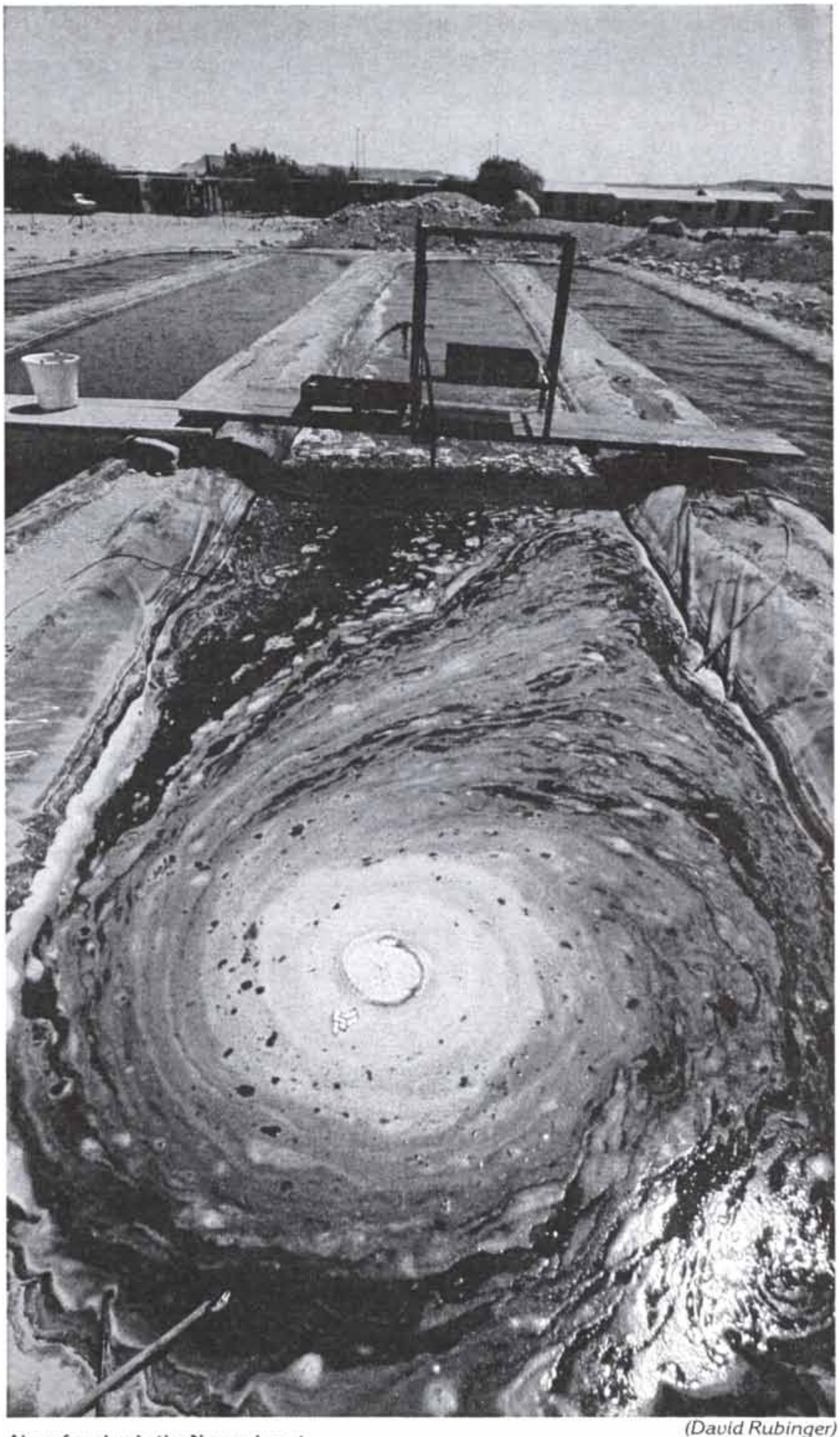
More than 100 different agricultural chemicals for export in ready-to-use formulations and 32 technical grade materials are produced by Makhteshim-Agan, the chief chemical producers within Koor Chemicals Ltd., the chemical division of Koor Industries Ltd., Israel's major industrial concern.

Makhteshim-Agan's R&D effort is directed mainly towards finding ultra-low-volume formulations and those providing superior ease in handling and storage stability. The company has managed to capture considerable chunks of the market in more than 50 countries, many of which have highly developed chemical industries of their own.

An imposing example of the development of the state's chemical industry is Ramat Hovav a $30 million industrial complex that has been erected in the Negev desert, seven miles outside Beersheba. The most modern technological facilities have been implanted in the sand with only the imperturbable bedouin with their goats and camels as neighbors. Four companies now operate there and the complex is rapidly becoming one of the country's most important chemical manufacturing centers. A rapidly growing labor force works three shifts around the clock, seven days a week, producing insecticides, bromine compounds and various chemical intermediates and pharmaceuticals. Makhteshim's plant at Ramat Hovav produces 4,000 tons a year of Ravyon, a particularly effective insecticide with low toxicity to mammals, and again, the Israel company is following the practice of taking a relatively small bite out of a massive market – a strategy calculated not to disturb a giant competitor.

In the sphere of pharmaceuticals, Israel is on the verge of moving forward from simply duplicating or improving existing drugs to marketing original products.

A landmark in this respect has been the introduction of the drug Alpha D3 which is regarded as a significant breakthrough in the treatment of renal bone disease and hypoparathyroid diseases in which there is an impairment of vitamin D metabolism. It is the first human pharmaceutical to be researched, developed and manufactured in Israel and is the result of collaboration between scientists at the Weizmann Institute and Teva Pharmaceutical Industries Ltd., Israel's largest pharmaceutical group. Teva's managing director, Mr. Eli Hurvitz, adds that Teva has recently launched a unique veterinary product based on the same active ingredient for the prevention of milk fever in cows. The product, Vetalpha, has been approved by the Israeli Ministry of Health. At this stage the company is actively negotiating with foreign companies for the international marketing of both products.



(David Rubinger)

Algae farming in the Negev desert.

BIO-TECHNOLOGY

Interferon, the agent that promises to aid significantly in the battle against cancer and to help the body resist viral disease, is being produced at only a dozen laboratories around the world. One of them is at the Weizmann Institute. Scientists there have adopted a human tissue culture system for making interferon for laboratory and clinical studies. One of three known sources of the glycoprotein, interferon, is the human skin, cells of which produce fibroblast interferon. Drs. D. Gurari-Rotman and T. Landau at the Weizmann Institute managed to isolate a particular self-propagating cell line, FS11, developed from a foreskin taken from a baby boy at circumcision. According to Professor Michel Reval, head of the Institute's department of virology, FS11 appears to be one of the best producers of interferon available. So the production facility is now being upscaled and a newly formed Israeli firm, Inter-Yeda, is shortly to begin commercial manufacture of the drug for expanded local clinical testing.

More than $1 million has been spent to build and equip a small commercial plant at the Kiryat Weizmann industrial plant near the Institute and it is planned to eventually produce some 20 billion units of inteferon from the supplies of foreskins available.

In another area of cancer research, a similarly ingenious technique is being used to obtain extracts from the thymus glands of calves that can be valuable in treating children whose own immune systems are not capable of preventing serious infection. A team of 50 scientists and technicians is currently operating a pilot plant to produce and test the extract and attempting to devise a method of producing it synthetically, with a view to commercial production.

A further application of bio-technology involves large-scale algae farming experiments which are being conducted at Israel's Red Sea coast by Koor Food Ltd. in collaboration with Weizmann Institute scientists.

A red single-cell algae organism, *Dunaliella bardawil*, which thrives in saline ponds, has been found to produce three eminently salable products: glycerol, an important component of many processed foods, cosmetics and pharmaceuticals; a high quality protein food; and beta-carotene, a widely used, natural food color from which the body makes its vitamin A.

At today's premium petroleum prices, plants which can also serve as a source of reduced carbon materials such as glycerol and beta-carotene are becoming increasingly attractive for commercial development. Koor has financed the construction of five acres of ponds at a site near Eilat to test growing conditions in open pool systems. Tests have established that from one acre of ponds, 30 tons of dry algae per year could be obtained yielding 60 per cent protein meal, 35 per cent glycerol and five per cent carotene. One estimate of the total cash value per acre has been put at nearly $61,000.

Koor's researchers have calculated that an area of about 33,000 acres of *Dunaliella* ponds could produce all the world needs to replace synthetic glycerol.

Investigations into farming *Dunaliella* algae began with a first cousin of the *Dunaliella bardawil* plant, *D. parva* a green plant which grows in the Dead Sea region and produces, in pilot ponds, about 40 per cent of its dry weight as glycerol.

In another project being conducted at the Technion, Israel's Institute of Technology at Haifa, algae is being grown in sewage as part of the anaeorbic digestion process absorbing protein from the waste. It can then be freeze-dried to form a supplement for feeds for fish grown in artificial ponds and for chicken and cattle, while the water, cleansed of the sewage, can be recycled for irrigation purposes. Another similar project on a more exploratory level is being conducted at Bar Ilan University.

DEFENSE

The pattern of Israel's national development has been considerably distorted by its security requirements. The hostility of the state's neighbors and the four major military conflicts that have marked the past 32 years have resulted in a third of the national budget being devoted to defense and the creation of an exceptionally self-sufficient and sophisticated R&D based defense establishment.

At the heart of this is Israel Aircraft Industries Ltd. (IAI), producing 350 military and civilian products and services, and employing 22,000 people. In the last financial year the budgeted turnover was $560 million and exports amounted to $325 million.

The company devotes 18 per cent of its turnover to R&D with more than one million direct man hours being financed from internal sources and a further one and a half million from external sources. Increasingly IAI's products are computer-based, with software accounting for as much as 60 per cent of many projects. "A modern aircraft," said one R&D executive, "is nothing more than a platform for computers."

Advances in computer software however, have been developed primarily by computer scientists as opposed to computer engineers and there is a worldwide shortage of the latter. "This is a new field of technology that is only advancing slowly," said IAI's computer systems manager. (Like the company's other R&D executives, he preferred to remain anonymous on security grounds.)

In the absence of any available software engineering technology, IAI is investing heavily in developing its own methodology. This includes the establishment of an educational center with its own laboratory and the development of what it believes is a unique 1,000 hour course to convert qualified engineers into software engineers. IAI's software system has been developed primarily by engineers who acquired their software experience on the job. It is considered unique in that it is highly interactive, practical and compact.

In the metallurgy sphere, IAI has developed two patented processes which promise considerable reductions in costs. One of these is a heat treatment process that prolongs the allowable life of nickel-base superalloys in jet engines. Turbine blades and discs are usually replaced prematurely as a precaution against creep, mechanical fatigue, thermal fatigue and gaseous corrosion. Creep – the gradual change of shape under stress – is the most limiting of these and IAI's engineering division developed a rejuvenation process. This is based on the hypothesis that secondary creep damage is due to concentration, at the superalloy's grain boundaries, of atomic defects that can be dispersed by heat-treatment. After their dispersal, the atomic defects can be 'pinned' to their new location so that the steady-state creep life of parts so treated can be extended by up to 300 per cent. The process is repeatable which means that the life of the turbine blades and discs can be extended to the end of the life of the aircraft. The other process improves the stress corrosion resistance of high strength aluminum. A low temperature heat treatment cycle raises the threshold stress for stress corrosion while maintaining the minimum strength properties providing weight savings of 15 per cent.

IAI's aircraft manufacturing division produces the Kfir C-2 multi-mission combat aircraft, the Arava STOL passenger and transport plane, the Westwind executive business jet, the Sea-Scan maritime surveillance plane and the Scout mini-remotely-piloted-vehicle.

Based around these planes, the company markets a combined and integrated weapon system for air, land and sea operations consisting of 16 sophisticated items. Apart from the aircraft these include radar systems, coast patrol vessels, cannons, armored vehicles, refuelling units, ground to air missiles and electronic fences.

The company is particularly proud of the success on the export market of the Westwind jet. "The sale of military aircraft and weapons system may be dictated at least partially by political considerations as well as technological standards," said one senior executive, "but we are well satisfied that with the Westwind, we can compete on the free market. Its success is true evidence of our advanced technological capabilities." Westwind jets are currently being manufactured and sold at the rate of four or five a month.

Israel's largest enterprise integrating optical, electronic and mechanical design and manufacture is the Israel Electro-Optical Industry Ltd. (El-Op). It established itself initially in the defense field and is still oriented chiefly in this direction; but it is now striving to achieve a 50 per cent ratio between civilian and defense markets.

The company is primarily involved in the field of night vision devices and fire control

FOR YOUR NEXT FOREIGN VENTURE – think of ISRAEL!

Industrial R&D makes Israel a world center for innovation

Despite its small size, Israel is unquestionably a world power in technological industry. In agrotechnology, electronics, medical technology, special machinery, heavy chemicals, fine chemicals, food technology and solar energy, Israel is a leader, both in research and production. More than 600 industrial R&D projects are started each year, each based on a new idea for sophisticated product development. More than 350 Israeli companies have their own laboratories and scientists aided by the well known Israeli universities and research institutes. Most of the companies are small, flexible and quick to adapt to changing market needs.

Israel's great strides in industrialization through innovation have been achieved through its own sophisticated manpower, and through cooperation with foreign industry. More than 150 foreign firms are involved in Israeli industry, having entered into profitable ventures ranging from wholly-owned subsidiaries to sponsored research projects. Both the Israeli economy and the foreign companies have benefited from such cooperation. Now, after 33 years of development, Israeli scientists and technologists generate more profitable ideas than the country can commercialize by itself. There is still room for your company to come and explore opportunities for cooperation. Not only can you become a partner in developing an Israeli idea, but Israeli expertise can provide the ingenuity to put your ideas into production.

INNOVATION is Israel's byword. New products developed locally brought in $1 billion in exports during 1980. Research and Development projects succeed more than twice as often as in other advanced countries.

STABILITY of Israeli industry, economy and society has been proved over the past 33 years.

RESEARCH has long been Israel's strong point – starting with agriculture, and growing dramatically in industrial research and development. Israel makes maximum use of its most important natural resource: its highly trained scientists and engineers.

ADAPTABILITY A new idea can be marketed quickly by an Israeli high technology firm without going through the labyrinth of corporate approvals typical of other industrialized countries.

EXPORT is the goal of all industrial R&D in Israel, since only high value sophisticated products can help Israel balance its payments. Already exports amount to 40% of the gross national product, and this figure will grow. Your firm can join the growing number of foreign companies participating in export-oriented Israeli R&D programs.

LOW COST is yet another advantage of doing industrial R&D and production in Israel. Salaries are lower than for similarly trained personnel in Europe or America, and the high success rate of innovative projects makes them even more cost effective. Add to this the 50% Israeli government participation in approved R&D projects, and the cost to a foreign firm to enter a new field through R&D becomes less than anywhere else.

How can one join the 150 foreign-based firms already profiting from Israeli operations?

- Write or telex to the Ministry of Industry, Trade and Tourism, Office of Chief Scientist, Jerusalem, Israel, Telex 25211 or visit the nearest Israeli Consulate or Embassy or Investment Authority office.
- Arrange for complete literature to be sent to your firm, and if you wish, a representative of the Ministry or Investment Authority will visit.
- Review the industrial R&D opportunities in your company's field and weigh the advantages of looking to Israel for your next overseas venture.
- Send a delegation of management and technical personnel to see for themselves what Israeli industrial R&D can do for your company. Ministry officials will be happy to plan the tour and accompany your personnel on a 1-2 week visit through laboratories, factories, institutes and offices.
- Decide on one or more of these ways to profit from the innovative atmosphere in Israel:-
  - ★ Establishment of a wholly or partly owned subsidiary.
  - ★ Starting a joint venture with an existing Israeli high-technology firm.
  - ★ Subcontracting projects to Israeli companies or institutes.
  - ★ Acquisition of marketing rights for Israeli innovative products.
  - ★ Starting a joint R&D project with an Israeli company with 50% support from the Government or the BIRD Foundation.
  - ★ Investment in a promising R&D project or company.

STATE OF ISRAEL

The Westwind II – Israel Aircraft Industries' executive jet

systems. It produces a laser range finder that looks like a pair of binoculars and is used in the same way but which gives an individual soldier the capability of measuring distances of up to 12.5 miles with an accuracy of plus/minus 11 yards.

El-Op has exploited a gap in the international defense markets by developing the facility to rejuvenate old tanks by installing laser range finders. "We could not compete with the major defense industries by producing a modern tank," says El-Op's manager of strategic planning, Mr. Samuel Neumann. "But we can rejuvenate a 20 year old tank, giving it the fighting capabilities of the 1980's at a fraction of the cost."

El-Op's activity is centered in four main areas of defense products – sights and cameras, including periscopes, gun sights and gun-sight cameras; active night vision devices utilizing high intensity infra-red light, driverscopes, goggles and filters; passive night vision equipment involving viewing systems utilizing and intensifying electronically natural star or moon light; and a weapon delivery system for aircraft and tanks.

The company's R&D staff is at the forefront internationally in the design of innovative optical systems and manufacturing methods.

"In one field – high vacuum deposition technology – we are perhaps the leaders," says Mr. Neumann. "We have a group of scientists working in the theoretical development and practical application of the methods developed in the manufacturing of reflection reducing, reflection enhancing beam splitters and dichroic filters and all related subjects for optical instruments in the ultra-violet visible and infra-red parts of the spectrum."

There are often instances where the technology developed for the defense market can be applied to civilian products. However, there are, he says, critical differences between marketing civilian and defense products. "In the military market you have to prove your device is better than what the competitor has to offer, but once you convince the right person you can sell hundreds. In the civilian market salesmanship is much more important, and must be directed to a much wider and less centralized customer base."

Mr. Judah Lando, the company's long term planning executive, emphasized the considerable advantage the Israeli defense industry has in the fact that virtually everyone is involved in military reserve duty. "This means that guys are often out in the field with equipment they helped to manufacture. In extreme cases they go into combat with it. A field officer can raise points that can lead to valuable product refinements and even new products."

INVESTMENT AND EXPANSION

Israel's fundamental dilemma has been whether to concentrate its R&D resources on selected areas, adopting the patterns of international specialization or minimally maintaining a broad spectrum of scientific fields, if only to create an indigenous capability to absorb science and technology originating from other countries. Two partial solutions have been attempted. First the development of specialized centers of excellence in science and technology and second, the extension of scientific resources through growth and international cooperation.

About two per cent of the gross national product (GNP) is invested in civilian and military R&D (1.2 per cent for civilian R&D). This places Israel close to some of the most developed countries like the United States, Japan, West Germany, The Netherlands and Great Britain which allocate between two and 2.3 per cent of their GNP to R&D.

However, in Israel most of the funds for civilian R&D go to institutions of higher learning while less than a third of such funds are spent in the industrial sector – the reverse of most developed countries. This situation is gradually changing, but for the moment investments in basic research are relatively high – 45 per cent of the civilian R&D budget compared to between 13 per cent and 24 per cent in other countries.

The government is the largest sponsor of R&D in Israel and its share in this financing is the highest among western countries reflecting the appreciation of the importance of R&D.

Professor Lavie believes that one of the most important tools he has to promote development is a research fund from which he can provide additional financial resources. Money is channelled to below critical sized firms lacking in risk capital; to potentially inventive types of industry short of capital; and to newly begun enterprises showing initiative and signs of ingenuity. Funds are allocated to the applying company as a grant, matching its own investment in specific R&D projects. The project is appraised, technically and economically. If approved, the R&D can be conducted in the firm's own facilities or in any recognized research or academic institute. Thus even a company lacking the needed facilities can benefit from the fund. The Office of the Chief Scientist (OCS) claims no ownership of the know-how, technology or patents resulting from the grant, provided the developer makes use of the results to the benefit of the country's

a finger on the pulse

Of the nation's requirements. Of the world's markets. Of the future's demands. For chemicals, electronics, telecommunications, consumer goods, food, metals and non-ferrous metals.

As Israel's largest industrial complex Koor keeps its finger on the pulse. Through research and development projects carried out. Through know-how agreements with high technology companies. Through joint projects in Israel and abroad with international corporations.

Koor keeps its finger on the pulse of world trade. By continuously diversifying the product mix of its manufacturing groups. Through an investment program geared to technologically advanced production techniques. And by steadily expanding the worldwide activities of its trade company on six continents.

Koor's finger is on the pulse. Of today's modern industrial society. Of tomorrow's technological challenges.

koor industries ltd

P.O.B. 33333, Tel Aviv, Israel

SOLCOOR INCORPORATED
Two Park Avenue
New York, N.Y. 10016 USA

economy.

The R&D fund, through its contributions, promoted R&D to the extent of $200,000 in the first year of its operation in 1968. By 1973 the figure reached $9 million; in 1975 it was $25 million and in 1979 approximately $70 million of which the fund's share was about 50 per cent.

The need to accelerate industrial development still faster caused the evolution of 'national projects' which are those with potentially great contribution to the national economy but on which industry is not ready to embark because of financial risks, shortage of funds and the length of time before fruition. If, after thorough examinaton, the chief scientist is convinced of the project's potential he may classify it as a 'Project of National Importance.' If approved, the participation of the government may be as high as 80 per cent.

The chief scientist also administers research institutes which serve as links between basic and industrial research, for both know-how and manpower, in the fields of fibers, physics, ceramics, rubber, paint, fermentation and plastics. Industry is encouraged to take advantage of the institutes' services, and the institutes are expected to generate up to 50 per cent of their budget from projects with industry.

The chief scientist's research fund has had a strong effect on the academic community in shifting interest and activity to industrial R&D. Universities have increasingly encouraged faculty members as well as graduate students to serve industry, and most academic institutes have established departments to serve as a liaison between their trained manpower and industry to promote this cooperation. The chief scientist's office has also contributed incentives to industry to utilize the universities' R&D potential by increasing the share of the grant used for university sub-contracting. Among similar incentives to promote applied research, is a scheme to pay for infrastructure projects designed to add know-how to serve a whole sector in industry. In some cases complete laboratories have been set up to serve groups of industrial firms.

Says Professor Lavie, "The effect and benefit of such encouragement to industry-university cooperation has been tremendous – to both parties. We regard it as a first priority to assist academic people to become industry-minded, to encourage them to work with and for industry and to inspire the students to turn to industry at an early stage. My department regards the professional manpower concentrated at the universities as the most valuable form of resources available to the country and so does know-how thirsty industry."

While Israel has made great advances by its own effort in establishing a high technology export industry, the OCS has also taken forceful initiatives in promoting the advantages of cooperation with other advanced countries. Although Israeli industry develops and produces a wide variety of sophisticated items which are aimed at high added-value export markets, in many instances companies cannot be fully aware of various critical factors which could affect their export potential. For this reason the Government Investment Authority, a department of the Finance Ministry, promotes cooperation agreements with foreign companies during the R&D stage of product development to develop full knowledge of international requirements.

The foreign companies benefit through their utilization of Israeli expertise particularly in such fields as communications, agrotechnology medical instrumentation and solar energy.

There are numerous ways in which foreign companies can participate in Israel's industrial R&D efforts, and benefit from its innovative atmosphere and government incentives: establishment of subsidiaries to develop and produce sophisticated products; joint ventures in establishing new companies with high technology input; investment in industrial R&D projects in Israel with Israeli government loans to augment the investments; acquisition of marketing rights for Israeli innovations; and technology transfer in both directions.

In these programs the foreign company can take advantage of the incentives offered by the Israeli government to their Israeli partners in the form of R&D grants which are usually 50 per cent of the project costs plus loans and grants for science based industry of up to 50 per cent of establishment.

Rephael Benvenisti, managing director of the Investment Authority describes his role as the art of attracting and encouraging a few quality investors like past blue-chip successes GTE&E and National Semiconductor. He cites Motorola as an ideal investor for Israel, a company that not only brought technology in the first instance, but developed unique research capabilities over the years.

A network of agreements has been organized to foster R&D activity. There are working arrangements with two countries so far – the United States and the United Kingdom. Parallel arrangements include a Binational Industrial R&D Foundation (BIRD F) with the U.S. government yielding about $2.5 million a year for joint projects and a new company in the U.K. for promoting technological cooperation.

The BIRD F, an innovation in itself, was established in 1977 specifically in recognition of the fact that Israeli companies, with their relatively high concentration of scientific and technological capabilities could be natural partners for U.S. concerns with compatible interests.

A typical example of its efforts concerns an Israeli company called Telrad which manufactures telephones and switch gear for essentially internal usage. Under the auspices of BIRD F, Telrad has established an association with Pentacom Inc., of Yonkers, New York, which sells small imported telephone systems to the American market.

Explains Dr. A. I. Mlavsky, executive director of BIRD F, "Pentacom have developed a good understanding of what certain sectors of the market in America really need, and have put forward the specifications. What is required for smaller businesses is an economic telephone system with the same rapid flexible and efficient communications facilities that characterize large modern, operator controlled systems. With its know-how, Telrad, the Israeli company, is developing the Key BX system in which a microprocessor together with sophisticated software permits a multi-function desk unit to operate with only two or three line pairs. The software is such that functions can be added, deleted or modified with little or no change in the basic hardware."

BIRD F is supporting about a dozen similar projects with sums, which are expected to cover about half the costs, of up to $250,000 per year for projects lasting a maximum of three years. The financing is done in the form of 'conditional grants' whereby the project partners undertake to make repayments to BIRD F from whatever commercial revenues eventually transpire.

The three major banks in Israel, Hapoalim, Discount Bank and Bank Leumi, also play an important role in launching technology oriented industry and in nurturing it through its early years. Each of the three has a subsidiary specifically for financing industry.

Bank Leumi, originally founded in the 1920's as the financial arm of the Zionist movement was the first to finance industry. Its subsidiary, Otsar Lataasyia (loosely translated as Industrial Treasures) financed one of the state's oldest and most successful companies, Dead Sea Works Ltd., now part of Israel Chemicals Ltd. Another early project was the first commercial raising of carp for gefilte fish at one of the original kibbutzim.

Bank Hapoalim's subsidiary, Ampal, and the Discount Bank Investment Corporation function similarly as specialized agencies for financing of technology intensive industry, which may require years of support before they become profitable. Sci-Tex and Elscint are typical examples.

Vigorous efforts to extend the list of foreign companies with which there are cooperative projects are being conducted by Professor Lavie of the Ministry of Industry, Trade and Tourism.

With the establishment of peace with Egypt and, hopefully, eventually throughout the Middle East, it will be possible, he believes, to begin industrial R&D cooperation with neighboring countries on projects of mutual economic benefit.

There is little doubt that the initial, spectacular development of the two Israeli sectors using R&D intensively – agriculture and defense – has had a tremendous effect on Israeli civilian industry. Advanced, intensive agriculture is based on novel industrial products such as fertilizers, insecticides, machinery for cultivating, harvesting and packaging, and irrigation equipment and controls. The close cooperation and location of developer, producer and user and the reputation of Israeli agriculture has made such products popular in the world markets. It has also resulted in the growth of related industries such as food, fine chemicals and veterinary products.

In the case of defense R&D and production, the world markets for such products – marine, aviation, electronic transportation and communication – whether for military or civilian uses, helped Israeli industrial firms to grow by leaps and bounds. In particular, the resulting trained manpower, scientists, engineers and technicians, has provided the backbone of new industries.

Expenditure on industrial R&D has increased dramatically in the past 14 years, from 11 per cent in 1966 to 60 per cent last year.

In addition to the 600 R&D projects being supported in some 300 Israeli companies, about 200 long range projects are being backed in universities and research institutes. The government's financial commitment in 1980 was $40 million. In real terms investment in R&D has doubled in the past three years.

Today an increasing emphasis is being put on intermediate and long range planning and on the benefits to be derived from international cooperation. Professor Lavie predicts that under the right circumstances, Israel's industrial exports based on local R&D will reach more than $2 million (in present value) by 1986.

"Mining" the sun is just one of Israel's accomplishments.

The sun: a safe, clean inexhaustible source of energy.

Israel learned how to "mine" it; pioneering the use of solar energy converters. And is exporting solar energy systems and know-how to other countries.

For 59 years, Bank Hapoalim has helped in the development of Israel's potential. Resulting in an impressive growth rate both at home and abroad.

Talk to the banking professionals at Bank Hapoalim in Israel and other major business centers around the globe.

New York • Los Angeles
Chicago • Philadelphia
Miami • Boston • Toronto
Montreal • London
Manchester • Paris
Zurich • Luxembourg
Buenos Aires • Sao Paulo
Caracas • Montevideo
Punta del Este
Cayman Islands

And over 320 branches of the group in Israel.

COME. GROW WITH US.

ARIELY

Bank Hapoalim B.M.

HEAD OFFICE: 50 Rothschild Blvd., Tel Aviv, Israel.

BOOKS

Amber, diamond, a new whole earth catalog and a pictorial record of the rise of science

by Philip Morrison

AMBER: THE GOLDEN GEM OF THE AGES, by Patty C. Rice. Van Nostrand Reinhold Company ($26.95). DIAMONDS: MYTH, MAGIC, AND REALITY, editor-in-chief, Robert Maillard. Crown Publishers, Inc. ($50). The Vistula River of eastern Europe is linked to golden amber, as the Orange River of southern Africa is to unyielding diamond. Both of these rare natural substances are found widely around the world, but now as in the past they are harvested mostly in the two classic places. Each year about two-thirds of the world's production of some 500 tons of the fossil tree resin is won by open-pit mining methods from a blue earth, a clay-rich layer a few meters thick under some 40 meters of overburden. The important pits lie close together on the peninsula of Sambia, along the Baltic coast just west of Immanuel Kant's old city, then Königsberg and now Kaliningrad, part of the Russian Soviet Federated Socialist Republic. Gem diamonds are found in "blue ground": kimberlite, a dense rock out of the depths. The world yields only about two tons of gem diamonds per year. It of course also produces five or six tons of natural industrial diamonds, but their total value is less than a fourth of that of the gems.

Amber too has its industrial uses, but that is mostly a thing of the past. Once upon a time the smaller and lower-grade pieces were heated to drive off volatiles; the dark brown residue, called colophony, was widely used in solution as a high-quality wood varnish, the favorite of violinmakers. Today synthetic varnishes dominate, and even the inexpensive tropical tree resins called copal are in little demand, except as a surrogate for the true fossil amber.

The Baltic deposits appear to promise a total of perhaps 100,000 tons of the product of the Eocene evergreens; South Africa seems likely to give up only hundreds of tons of diamonds. The amber workings surrender a handful of glowing pieces per ton of material removed; the South African diamond mines have an average yield of less than one part in 50 million, say a rough gem crystal the size of a small peppercorn in a ton of rock.

Amber is the complex isoprenoid hydrocarbons of soft tree resin, strongly cross-linked, polymerized and hardened by time. It has been oxidized somewhat, an oxygen atom for every dozen carbons. Pliny recognized that amber originated with tree gum. The amber-bearing forests thrived for five million years from the North Sea to the Caspian. The blue earth is not, however, a former forest floor; it abounds in fossil oyster shells. It was laid down in the encroaching shallow seas of the Eocene epoch by streams that washed light water-resistant lumps out of the forest soils. Amber is both mobile and durable in water; to this day the shores of the Baltic yield it as flotsam. Worldwide analogues of the process have been found in every epoch from the Carboniferous to the Recent. It was the amber given up by the sea that was worked by the Neolithic dwellers of the Baltic coast and that diffused throughout Europe in a clearly marked trade from the earliest Bronze Age times. After the Baltic deposit the most plentiful source of amber is one first reported in print by Columbus in the mountains along the northern shore of the Dominican Republic.

Diamonds too disdain the attack of water. Their surface is not easily wetted, a property exploited to recover diamonds after a long process of gravity enrichment with respect to the lighter grains of the crushed ore. Since the days of the Golconda miners it has been known that diamonds adhere to greasy surfaces and wetter materials do not. Greased belts now collect the diamonds from the concentrate; the diamonds are released from the scraped-off grease by melting it. The weathered kimberlite of some mines resists the process, and since the discovery in 1958 by Russian research workers that X rays induce a strong blue fluorescence in most diamonds photoelectric sorters are widely employed.

Zircons and garnets often accompany diamonds all the way to the sorting table. There a keen eye and a quick hand finish the job. At the mouth of the Orange, along the desert beaches of Namibia, heavy earthmovers (scheduled by computer) work the gravels right down to the old bedrock ("the jagged paleo-coast"). These seaside deposits are richest in gemstones; the better, rounder single crystals preferentially survive the long weathering in the watery journey from the kimberlite pipes of the high interior down to the sea.

Amber easily reveals itself: it is too light to be a stone. The prized fires of the diamond, however, lie hidden within the uncut stone. The most interesting essay in *Diamonds: Myth, Magic, and Reality* makes a strong case for human admiration of diamond's fire having emerged slowly and late. The elaborate craft of diamond-cut-diamond that facets the gem is a modern one. Diamonds were known only from stream deposits in India and in Borneo until they were found in Brazil in the 18th century. The Sanskrit writers, whose lore somehow informed the Roman admirers of diamond, seemed to hold it precious because of its natural octahedral form, vaguely and metaphysically the source of its indestructible nature. Neither brilliant nor colorful, these uncut diamonds were seen by the Roman Manilius as the "point of a stone more precious than gold."

In the book we see a remarkable painting of Henry IV of England, done from life in about 1400, his golden sleeves "adorned with two spendid octahedrons" painted rather larger than is plausible. It is pretty certain that it was the metaphorical and not the visible virtues of these first of all painted diamonds that made them kingly. Indeed, it seems that the recognition of diamond's hardness itself emerged slowly out of confusion. The old authors talk of the red diamonds so much prized by the Parsees, jewels the color of sacred fire. In nature, however, there are very few such gems. What even the wealthy Parsees collected must have been the commoner octahedral red spinel.

Amber: The Golden Gem of the Ages is the satisfying work of a single collector and investigator. Dr. Rice gives the general reader a useful and up-to-date account. She has avoided the error of many collector-authors, who devote most of their attention to the issues of quality assessment and the marketplace. She has not forgotten to tell you how to make sure that what you buy is amber, not the reddish faceted bakelite of the 1920's or the beads of lightly worked modern plastic rods now pouring out of East Africa. She has also laid out the whole long story, from the room with amber walls and amber furnishings given to Peter the Great (now lost and being re-created by the People's Master

Artists of Applied Art of the Latvian S.S.R.) to the infrared spectroscopy that can reveal the botanical affinities of each amber deposit. Her professionalism enables her to avoid a second pitfall of collectors: a lack of documentation and skepticism. Thirty years ago the amber story was told by that wonderful writer Willy Ley, whose central theme was the science of the insects trapped in the ancient gums: "dragons in amber." The zoological theme is not salient in Dr. Rice's book; her work is all the same a worthy successor to Ley's 1951 classic for the general reader.

Diamonds: Myth, Magic, and Reality is another kind of book altogether. It is the bulky and lavish celebration of a rich industry, its editors and contributors an international panel of leaders of the trade, the craft and the scholarship of diamond. The photography is brilliant. The color-filled lens has been taken everywhere, to look at the slender ankle of a Bombay model ornamented with an ankle ring of elegant Jaipur workmanship set with diamonds of antique cut, at the microinclusions of foreign crystals deep inside diamonds, at the mines, the cutters and the gems themselves around the world and through history. There are even two black-and-white photographs from the workings of the Russian diamond industry in the permafrost of Yakutsk, a mere glimpse of the world's second-largest producer. In 1972 the purchasing monopoly long held over rough diamonds by the great firm of De Beers was ended by negotiation; those Russian diamonds are now sold in the biggest market, Antwerp, by a joint Russian-Belgian company with the name of Russalmaz.

Gems are epiphenomena in the needs of humanity, although they seem to be universal to our species. Amber and diamond also share a tragic element of the human experience. The amber of the Baltic is caught up, more or less by the accident of geography, in one of the oldest conflicts of European life, grown more virulent in our century. The names on the map, the borders, the workmen and the flow of trade change; a study of amber is in some ways an oblique look at the high tragedies of two world wars: Tannenberg and Leningrad. Diamond, on the other hand, seems close not to war but to slavery. The Brazilian experience is vivid in an old print, black men at work raking gravel in a long line of water-fed pits. Before each miner sits a uniformed white inspector, alert for the find of gems. The lucky man who raked up a big enough diamond would be garlanded, manumitted and given a job at a salary. The diamond industry still depends on hard labor, both at the rocky source of the diamonds and at the demanding and better-paid work of the cutter's bench. That kind of economics is not dealt with in this market-centered chronicle, but it appears to a reflective reader that producers who can somehow set favorable terms for their miners' pay have a social resource as valuable as a pipe of rich blue ground.

THE NEXT WHOLE EARTH CATALOG: ACCESS TO TOOLS, edited by Stewart Brand. Distributed in the U.S. by Random House, Inc. ($14). The first avatar of this work, in a way both substance and symbol of its times, was reviewed here in 1969. It was then about a tenth this size, yet sold for five less puffy dollars. On that minimum appraisal alone the new book (with the same spirit and editor) is a bargain. Quality again follows quantity; it is a prize. The inside front cover offers a world biogeographical map on the big, floppy page of the whopper paperback (a publishing style the first version pioneered), and the inside back cover is a computer map of a million galaxies. The front and back covers both show the whole earth, in one case the view at lunar earthrise. In between the 600 pages offer coherent reviews of good books old and new and keen judgments of specialized magazines, all of them shrewd guides to sources, both of knowledge in print and of the tools of almost all trades. There is one broad limitation: the volume is aimed at people who will read, think and act by themselves and with those with whom they share both the result and the labor.

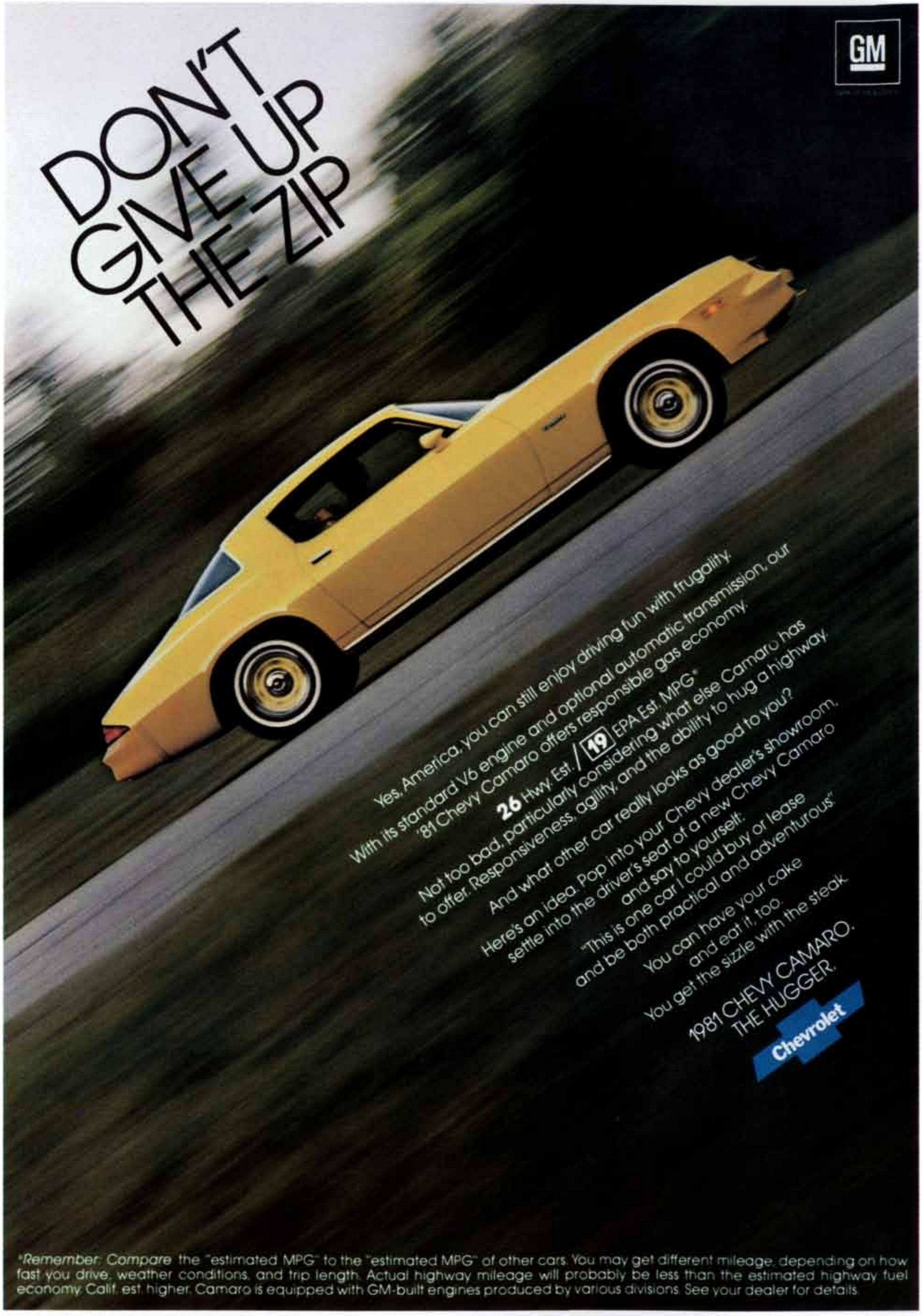
Stewart Brand is as acute a writer as the offset press has known. He explains that the edition of 1971 (1969 too, it seems sure) was aimed at the young counterculturists of the late 1960's and early 1970's, and that the new edition is meant to be of use to their now older contemporaries ("one of us governs California"), to new college agers ("with [their] unusual ability . . . and . . . lack of confidence") and to the "vast citizenry" who are learning that when you build or fix your own car, house, body or community, you get your time's worth, untaxed and uninflated.

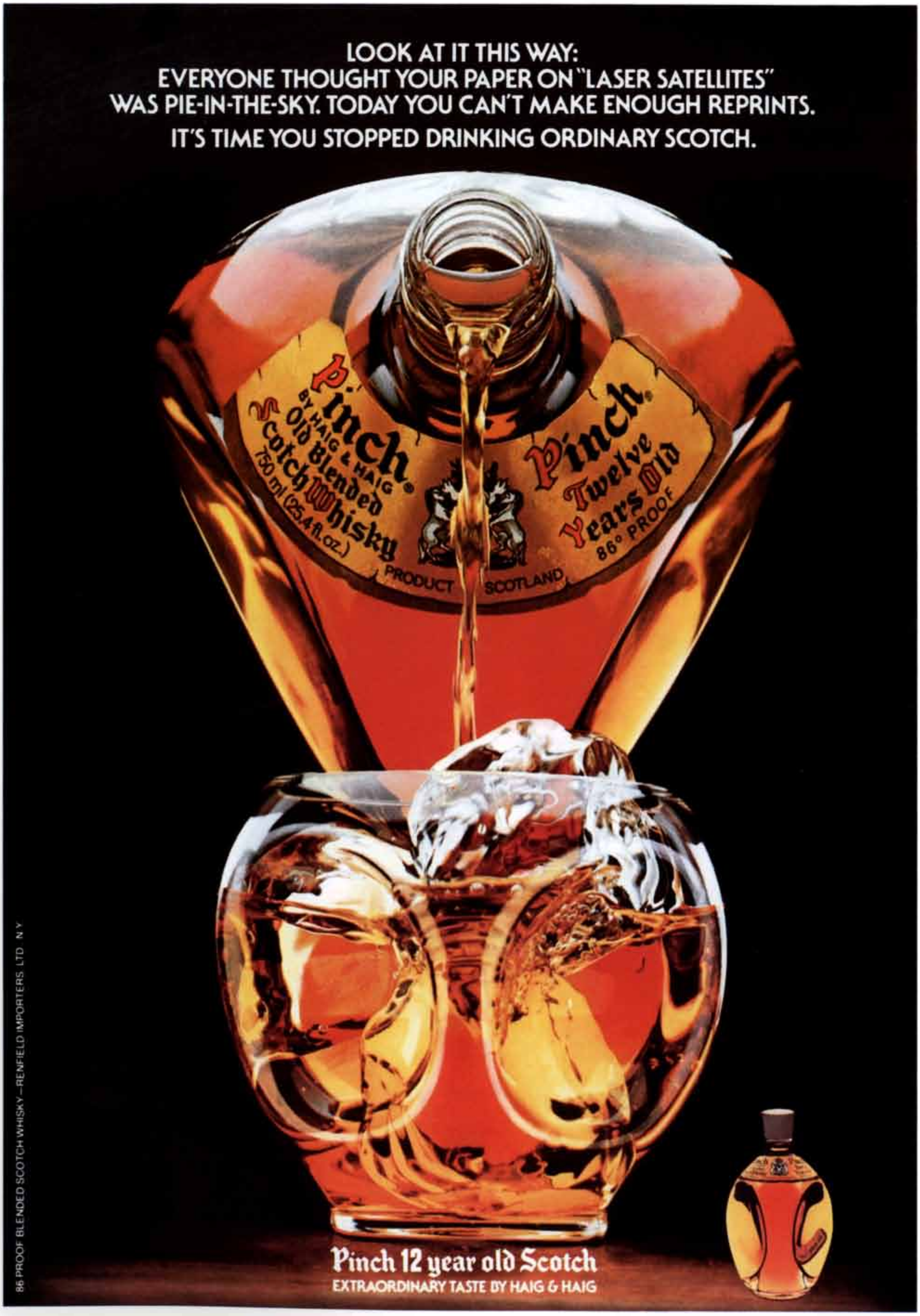
Just about every page outlines a first-rate book, journal or catalogue (access is always postal) with comments by candid reviewers. The material is topically arranged, some 600 topics from terminal self-care to cockroaches, from beer to computer networking. Opinion is free but always documented; the items are generally the best to be found, at least along the self-taught routes the book travels. Technology is not soft out of preconception but out of performance and capital cost; old diesel-powered fishing boats, plans for space colonies and the Onyx business computer are admired. The books occupy the spectrum from *The Atlas of Sexual Pleasures, The Diamond Sutra and the Sutra of Hui Neng, Home-Satellite TV Reception Handbook, The Merck Manual* and *Bartlett's Familiar Quotations* to Joseph Needham's great *Science and Civilization in China.* Both drop-in experts and staff members have contributed.

Consider three of the treatments, each a few pages that alone would repay anyone who needed them the cost of the book. In the first treatment maps, atlases and photographs from the air and from orbit are appraised and described with workable details of how to get them, including deserved praise of the U.S. Geological Survey as "the friendliest and easiest big Government office to work with." Topographic quadrangles, Landsat images, CIA atlases, state atlases, the geological highway maps put out from Tulsa, the marvelous hand-drawn landform cartography of Erwin Raisz (the best U.S. physical map to be found, still available at about a dollar)—all are made accessible not only by title or address but also with the ring of actual experience through the mails.

The second treatment is one on children's books about how things work. The complete corpus of three splendid author-illustrators is listed and described, with much more. The third treatment, particularly persuasive to this inexperienced reader, is a page on coppiced woodlots. A well-run coppice (the technique has been widespread in Europe for millenniums) offers cooking fuel in a steady state for a family of four from a tenth of an acre. Deciduous trees are cut to stumps, from which new stems sprout repeatedly and grow in a few years to firewood diameter. French and English references are compared, and suppliers of seed and advice are listed.

Most of the material is seen in clear light by the enthusiasts who have written about it. The mystic is here, and there is a page on the cultivation of cannabis, exactly as objective as one on self-publishing. Trendier entrapments get little space, such as UFO's and Carlos Casteneda (who receives a sensible if friendly discussion, discriminating his early fictions of some content from his shameless potboiling). The indispensability of the work for libraries is probably limited at the younger end by its explicit treatment of sexual behavior: hetero-, homo- and auto-. There is some first-rate writing, both by the editor and by various anecdotal and aphoristic contributors. Hear Brand on the uniqueness of the space environment: "Space is not like a continent, or the Pacific Ocean, or anything else we've experienced except possibly death and rebirth. It's more like a Buddhist chant: No-air-no-gravity-no-night-no-day-no-up-no-down-no-motion-no-past-no-stand-


GM
DON'T GIVE UP THE ZIP
Yes, America, you can still enjoy driving fun with frugality.
With its standard V6 engine and optional automatic transmission, our '81 Chevy Camaro offers responsible gas economy.
26 Hwy. Est. / 19 EPA Est. MPG*
Not too bad, particularly considering what else Camaro has to offer. Responsiveness, agility, and the ability to hug a highway.
And what other car really looks as good to you?
Here's an idea. Pop into your Chevy dealer's showroom, settle into the driver's seat of a new Chevy Camaro and say to yourself:
"This is one car I could buy or lease and be both practical and adventurous."
You can have your cake and eat it, too.
You get the sizzle with the steak.
1981 CHEVY CAMARO. THE HUGGER.
Chevrolet
*Remember: Compare the "estimated MPG" to the "estimated MPG" of other cars. You may get different mileage, depending on how fast you drive, weather conditions, and trip length. Actual highway mileage will probably be less than the estimated highway fuel economy. Calif. est. higher. Camaro is equipped with GM-built engines produced by various divisions. See your dealer for details.


LOOK AT IT THIS WAY:
EVERYONE THOUGHT YOUR PAPER ON "LASER SATELLITES"
WAS PIE-IN-THE-SKY. TODAY YOU CAN'T MAKE ENOUGH REPRINTS.
IT'S TIME YOU STOPPED DRINKING ORDINARY SCOTCH.
Pinch
BY HAIG & HAIG
Old Blended
Scotch Whisky
750 ml (25.4 fl. oz.)
Pinch
Twelve
Years Old
86° PROOF
PRODUCT
SCOTLAND
Pinch 12 year old Scotch
EXTRAORDINARY TASTE BY HAIG & HAIG
86 PROOF BLENDED SCOTCH WHISKY—RENFIELD IMPORTERS LTD N Y

ing-still-no-life-also-no-death-no-thing-only-waves-of-star-star-star-star."

A glow of goodwill, courage, reason and profound engagement suffuses this remarkable volume, a reference compendium for hardware and software, plumbing and philosophy, at once traditional and iconoclastic. It celebrates the hand and the reading eye for the self-helping American as consumer, small-business entrepreneur, pioneer, sportsman, parent, neighbor, cooperator. It is free of prejudice, although not of zealous and sometimes even cockeyed conviction. The producers of the camera-ready copy are here in picture and word—the entire crew, from the editor to the proofreaders and typesetters.

The *Catalog* is prepared on the San Francisco Bay waterfront for the bigger world of its New York publisher and Indiana printers. It remains consistent; there is no other enterprise that offers such a clear and frank page on its own accounting costs, wages, hopes and income. Of your money the nonprofit foundation that employs the people you have met gets $2. There is even a final accounting of the implied tree budget; all writers today communicate by the immolation of forest. *The Next Whole Earth Catalog* took less than 15 acres of watershed; it promises a net tree gain if one reader in 20 will plant a tree. Perhaps this is too narrow a circle of concern; you can never change only one thing. *The Next Whole Earth Catalog* can reasonably hope to change a good many minds and actions over a period of time.

EARLY RADIO WAVE DETECTORS, by Vivian J. Phillips. Peter Peregrinus Ltd., IEEE Service Center, 445 Hoes Lane, Piscataway, N.J. 08854 ($42.75). Ernest Rutherford and Oliver J. Lodge, C. V. Boys and of course Guglielmo Marconi, with many other physicists and engineer-inventors, walk through this careful and affectionate history of an important side path along the high-road to contemporary science and technology. There was a clear task before these motley pioneers: devices to display the presence of radio waves. Detection here does not mean only the rectification process of the current technical radio vocabulary; rather, it holds the more general meaning of rendering radio-frequency waves perceptible to the senses. The obvious technological need and the variety of processes that could be employed served to recruit a swelling company of the empirically ingenious during the three decades between Heinrich Hertz and the victory of the crystal and the tube, the selfsame epoch as the rise of modern physics, with the high vacuum, the electron, the atom, the nucleus and relativity.

Hertz first of all looked for a tiny spark, and the obvious developments followed, to Thomas Edison with the gap held in a little dark box and Lodge with adjustment screws. But the prototype detector everyone knows about is the Branly coherer. The effect was 50 years old when Édouard Branly worked on it in the 1890's. A very poor contact between conductors can become much better after the application of a few tenths of a volt, even at high frequency. There were many realizations. Marconi used a glass tube full of a mixture of nickel and silver filings between silver-amalgam contacts, with the tube evacuated for stability. A relay system caused a bell hammer to tap the tube after each coherence, thus reestablishing the high resistance.

A parade of patents modified this scheme, although by 1910 or so it was obsolescent. The standard detector was by then electrolytic. There were various forms of electrolytic detectors, falling in two general classes. One, the work of Lee De Forest, was an anticoherer. It arranged for the growth of minute tin or silver crystals that delicately bridged a gap until the high-frequency current broke down the invisible structure—perhaps. Another, dear to W. H. Eccles and Reginald A. Fessenden, depended on a fine-point metal electrode in a fluid, with the circuit polarized to high resistance by a film of gas. The signal somehow depolarized the film, in an obscure process no longer of interest to electrochemists. Indeed, the action of the coherer is not obvious even with today's solid-state wisdom. The author of this delightful out-of-the-way monograph has himself experimented with coherers to show that the formation of chains of particles by their mutual electrostatic attraction, the breakdown of the thin oxide surface layer on the particles and localized heating at contact points all play a role in the overall effect.

Other techniques came on the scene. The most unexpected portent of the future is mechanical oscillators, usually in the form of fast-spinning multicontact wheels, to generate a local current oscillation that could match and beat against the incoming signal. It was a crude mechanical forerunner of the heterodyne idea, "one of the most important innovations in the whole field of radio." And a hint of the signal-to-noise concept is found in an entirely practical context way back in 1905.

Here we see a table of comparisons, published in 1911, that lists and assigns figures of merit to the chief detectors of the day. It comments on vacuum diode "valves," known for five years or more from the experiments of J. A. Fleming: "suitable for general work, but no advantage in using in preference to electrolytic or thermal." The electronic tube, now with grids included, came to full dominance in about 1925, to dim out in our semiconductor era. Some quantitative data were long known. The coherer needed a couple of ergs of signal energy if it was to detect a Morse-code dot. The electrolytic detector was almost 1,000 times more energy-sensitive, and a good modern receiver, nothing exceptional, has improved on that performance by six orders of magnitude. (The best technique of modern radio astronomy lies another 10,000-fold beyond that.)

A Dr. Lefeuvre, professor of physiology at Rennes in Brittany, used the sciatic nerve and muscle of a frog's leg to amplify the weak output from an electrolytic detector. The smoked-drum-surface record of the twitches is reproduced here, the Paris dots and dashes quite clear 200 miles away. It was, after all, a similar system set up to display a nearby lightning flash that had long before led Luigi Galvani to the discovery of electrochemical currents. This steady improvement in frog-leg detectors has somehow never been followed up.

RAINBOWS, HALOS, AND GLORIES, by Robert Greenler. Cambridge University Press ($24.95). The Saint Petersburg sun was bright on the morning of June 18, 1790, and in the blue sky there stood for hours a complex set of glowing white halos and spots, circles and tangent arcs, a marvelous display. We have a meticulous drawing of it, made by the court apothecary Tobias Lovits (or Lowitz), a research chemist of originality in the years of Lavoisier and a scrupulous observer to boot. In two diagrams of this handsome and absorbing book we see the test of Lovits' claim to accurate reporting, a piece of sky archaeology. On one page is an engraving of what he drew for publication in *Nova Acta Academiae Scientiarum Imperialis Petropolitanae;* on the facing page is a modern computer simulation, produced as one outcome of a dozen years of work by Professor Greenler and his friends.

Greenler is a surface physicist at the University of Wisconsin at Milwaukee who has been caught up in the optical phenomena of bright skies. Nowadays there are keen observers everywhere, and a good many of them are hard-core sky watchers, camera ready, well versed in the arts of fish-eye or garden-globe wide-angle pictures and scattered to the ends of the earth. Indeed, there are quite a few color photographs here from the South Pole, some of them the work of the author.

The complicated displays are the consequences of the sun's rays being refracted and reflected, internally and externally, from small airborne crystals of the mineral ice. For most of the effects only two forms of ice-crystal grain need be invoked; these forms are common enough. One photograph shows the two: little hexagonal prisms, like tiny unsharpened pencils, and flat hexagonal plates, like slices of the same pencil, collected as they fell to the surface at the South Pole station. In warmer climates (such as Peter's city) they are often confined to higher altitudes aboveground.

An excellent qualitative physical account always goes along with the detailed simulations in the book. The old descriptions are made clearer and a few important new results emerge. A fine table relates the crystal type and orientation to the name of the effect and to its simulation. The simulations are striking. What one sees is a sky picture formed of dots. Each dot marks the direction of a ray of light, computer-traced from the assumed sun position through the crystal cloud to the observer's eye, with all projection and optical effects included. Intensity is represented by the clever device of using chance: only those lucky dots are plotted that have won a computer lottery in which the odds are proportional to the calculated intensity. Certain arcs, for which we have only very few representations (such as Lovits'), may arise from plate crystals spinning on a horizontal axis through points of the hexagon. That result is not yet secure; it does seem that the old drawing must be, reasonably enough, not a snapshot picture of the great 1790 display but a composite from sketches made during the entire wonderful morning.

Another fine multiple-halo event was reported by an Australian, J. R. Blake, as he bounced across the South Polar ice cap in a Weasel 20 years ago; his report too matches very well the 100,000 dots of the simulation. The sun pillars that often mark the sun, the moon or even streetlights in more ordinary climes appear as simple shafts of light vertically across the source. They turn out much more closely to match the model of reflections from the sides of horizontal pencil crystals than to support the older view: that pillars originate with reflection from the flat surfaces of horizontal plates. Both effects can occur, and here are some clues for discrimination in the field.

Greenler's text includes, and his interest encompasses, much more than the tangled challenge of the complex computer-studied events. It is he who first showed that the familiar rainbow is invariably extended to an infrared rainbow, clear to photographic plate and filter but invisible to the highly evolved eye. The rainbow is of course the circular consequence of clear refracting spheres, a form natural to the world of liquid water but unknown to that of ice. Greenler sets a puzzle: a rainbow photographed in an open icy landscape. What made the rainbow? These columns will not disclose the convincing special circumstances.

Here are remarkable mirages and glories, iridescent clouds and the dark night sky revealed just as blue as the day. Here are sky illusions of perspective, green flashes, sundogs and rainbows, rainbows, rainbows, in both image and explanation. Not as wide-ranging as the simpler little classic *The Nature of Light and Colour in the Open Air,* by Marcel Minnaert, but deeper in many places, the book is a varied and insightful pleasure. It is bound to be an indispensable learned resource for all who would grow in their understanding of "the dull catalogue of common things."

ALBUM OF SCIENCE: FROM LEONARDO TO LAVOISIER, by I. Bernard Cohen. Charles Scribner's Sons ($37.50). Historians of science are tireless construers of the word. This fine collection is exceptional; its Harvard compiler, a teacher and scholar well known for his authoritative and detailed studies of the work and times of Newton and of Franklin and for his lively book for the general reader on the rise of dynamics, has taken us to the stacks of rare books. But there we enjoy not texts but images, symbolic materials closer to perceived reality. Ever since books were first printed, multiple copies of illustrations have celebrated, enriched and (in fields such as botany) authenticated knowledge and its accurate transmission. In a thoughtful blend of the indispensable and the fresh ("I believe that there are no readers who will be familiar with all of the illustrations in this book") Cohen reproduces a wealth of engravings, etchings, woodcuts and drawings.

Most of the 368 illustrations are indeed from printed books; some excellent ones enter from manuscript or gallery wall, and there are a few photographs of artifacts. This grand library tour is systematically organized, and each picture is flanked by a paragraph to set it reflectively in the context of the time of the rise of modern science. The library without shelves is arranged in a couple of dozen sections. They owe much to Harvard's own treasury of holdings but are indebted as well to Cambridge, Florence, Bologna, Leningrad, Leiden, London and many other places where the archives of the scientific revolution are stored. Each section is headed by a page or two of comment by Professor Cohen, and the book ends with a useful annotated bibliography for the reader who would go further.

If Cohen has finally chosen 368 out of "more than a thousand absolutely first-rate illustrations," here we can describe only half a dozen. In 1612 Ludovico Cardi da Cigoli painted within the dome of the papal chapel in Santa Maria Maggiore in Rome a panel of the Assumption of the Virgin. The crescent moon is often depicted at the Virgin's feet; Cigoli is true to that tradition, but his moon is cratered; it is the telescope's moon first seen just three years earlier by Galileo, the artist's friend. We can admire the observatory at Danzig (now Gdańsk), until Greenwich and Paris the finest in Europe. It was private, and we see its keen owner-observers, Johannes and Catherina Elisabetha Koopman Hevelius, jointly at work on the big sextant.

A remarkable manuscript drawing, found in a letter sent to Kepler in 1624, unknown until it turned up in Leningrad after World War II, shows the first mechanical digital computer, a scheme with setting dials and mechanized Napierian multiplying rods. The principle was extended to a line of successful calculators before World War I, but the original machine was destroyed by fire before Kepler could use it. Kepler's inventive friend at Tübingen was Wilhelm Schickard. Then there is a photograph of the elaborate silver microscope made in 1761 by George Adams for King George III. One can hardly reach the tube for the sculptured figures; usefulness is sacrificed to ornament.

A drawing made for eventual engraving shows Galvani's frog-leg preparation, which twitched each time a spark was drawn from the electrostatic machine nearby. A page of 64 small drawings is the pictorial specification of the proper portable chemical laboratory of 1689. The Frankfurt chemist Johann Becher had an inventory of apparatus quite like a position at any beginning chemistry bench of a few decades ago, without a Bunsen burner but including a glass-cased analytical balance.

The volume ends with the Enlightenment and with science institutionalized. Diderot's *Encyclopédie* acclaims faith in reason, and the hope for a social order as rational as that observed in the natural and technological world. We no longer see progress as a short-term certainty. Yet here too is the crowded collector's cabinet of the Copenhagen physician Ole Worm in 1655. The room contains a bizarre assortment, but there is a fine collection of worked stone implements, then considered to be only a kind of thunderbolt, not human artifacts at all. The history of mankind is slower and longer than they thought, even in the day of Condorcet and Franklin. Maybe there will still be time.

The reproductions are excellent; a few suffer from the necessary reduction in size. The volume is dedicated to Charles and Ray Eames, masters of the images of science, who joined meaning and form in our fast-flashing day.

Take this 4-volume Basic Astronomy Library for $2.95 (a $60.80 value)

if you will join the **Astronomy Book Club** now for a trial period and agree to buy 3 more books—at handsome discounts—over the next 12 months.

Black holes, quasars, pulsars, supernovae, white dwarfs—explore the fantastic discoveries being made about the universe with this outstanding selection of four critically acclaimed astronomy books. Together they form a marvelously comprehensive Basic Astronomy Library—a library which will allow you to tour our cosmos—from the Red Planet to the Seyfert galaxies . . . from the Earth's core to a black holes horizon . . . from how the universe was born to how, when, and where it will die.

Here is what your library will include:

WHITNEY'S STAR FINDER. *Charles A. Whitney.* Top-notch field guide to the heavens. Enables you to identify every major star in the sky over North America. Comes with a Star Finder-Locater Wheel. Publisher's price $6.95.

BLACK HOLES: The Edge of Space, The End of Time. *Walter Sullivan.* Bottomless pits into which whole stars can disappear . . . *Black Holes* takes you on an incredible journey that will change your picture of the universe. Hardbound. Publisher's price $17.95.

CATALOGUE OF THE UNIVERSE. *Paul Murdin et al.* Three-volumes-in-one: an atlas, encyclopedia and book of descriptive astronomy. Surveys over 220 astronomical objects. More than 300 photographs. Publisher's price $17.95.

THE ILLUSTRATED ENCYCLOPEDIA OF ASTRONOMY AND SPACE. Revised edition. *Edited by Ian Ridpath.* Your window to the universe. Over 1,000 entries and 150 photographs and illustrations. Hardbound. Publisher's price $17.95.

. . . quite a trip for only $2.95—and that's just the beginning!

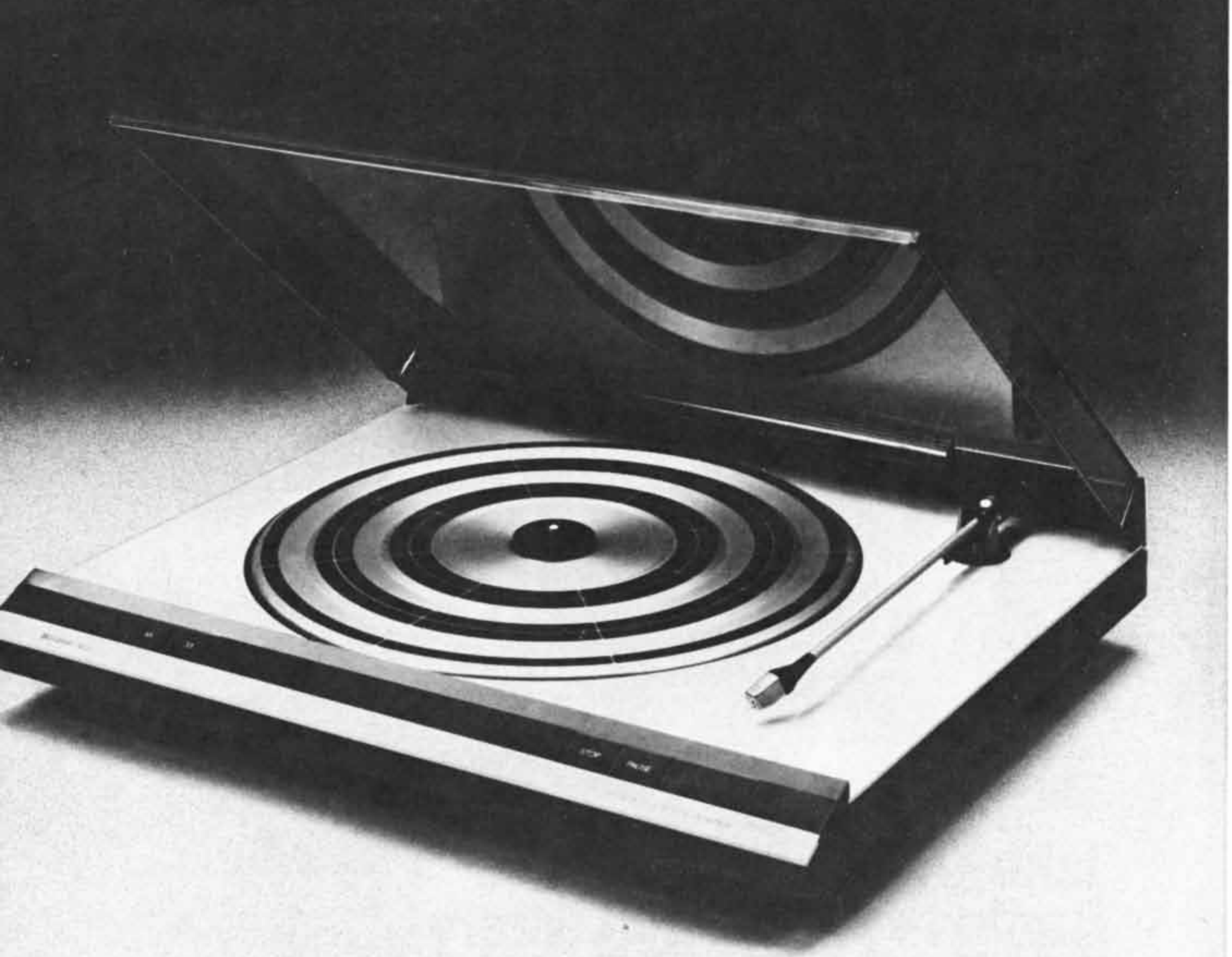
What is the Astronomy Book Club?

Astronomy Book Club is a very special club—a book club operated exclusively for astronomy enthusiasts. It's devoted to making the finest astronomy books available conveniently and inexpensively. Recent offerings include titles by such eminent authorities as Carl Sagan, Robert Jastrow, John Gribbin, Paolo Maffei, and Isaac Asimov.

As an **Astronomy Book Club** member, you'll receive books of your choice at savings up to *30% off publishers' prices.* In addition, you'll immediately become eligible for our Bonus Book Plan, with savings of at least 70% off the publisher's price on any book we offer.

Discover the phenomena of space—join **Astronomy Book Club** today.

If reply card has been removed, please write to: **Astronomy Book Club**, Dept L-AH6, Riverside, N.J. 08075 to obtain membership information and an application.



ITS BEAUTY IS MORE THAN SKIN DEEP.

Meet the Beogram® 3404 ... an intelligent, thoughtful and very well spoken turntable.

Like all Bang & Olufsen turntables, the 3404 effectively combines superior sound reproduction with simplified—not complicated—operation. That's the real beauty of the 3404.

That's why the controls are outside of the dust cover where they're easily accessible.

And that's why just a light touch of the START button does everything. It determines if a record is on the platter, and if so, quickly sets the correct speed and lowers the stylus to the record.

Slam! Pound the shelf next to the 3404. Not a skip, not a jump ... thanks to the rock-steady patented suspension system. Audible acoustic feedback disappears as a result.

A low inertia tonearm with a knife-edge bearing system eliminates audible distortion. The Bang & Olufsen MMC cartridge works in flawless harmony with the tonearm. The unique self-correcting electronic servo-drive is so precise that the need for a conventional strobe device is eliminated.

The Beogram 3404 even responds to remote control commands when used with the Beomaster 2400 receiver.

At Bang & Olufsen, good looks and brains run in the family. So discover the inner beauty of our full line of fine turntables at your local Bang & Olufsen dealer or write us for complete informative literature.

Bang&Olufsen

Bang & Olufsen of America, Inc.
515 Busse Road
Elk Grove Village, Illinois 60007
Attention: Sonya Anderson

Established 1845 SCIENTIFIC AMERICAN January 1981 Volume 244 Number 1

Toward a Rational Strategy for Oil Exploration

How much oil does the U.S. have left? Clues are yielded by computer simulation of the results of drilling for oil at random, but decisions on energy policy call for a thoroughgoing inventory of the resource

by H. William Menard

The U.S. is having difficulty arriving at a consensus on energy policy, partly because policy objectives rest on widely differing assumptions. Some people assume that the nation's oil pools are running dry and that incentives should be offered to speed development for reasons of economic health or national defense. Others believe that oil is abundant and that either competition should be strengthened or the profits of the oil companies should be limited. Still others believe conservation or other sources of energy offer alternatives to the development of oil and gas, particularly in environmentally sensitive areas such as the continental shelf near recreational seashores.

The facts about the nation's dependency on oil are familiar and quickly stated. Oil provides nearly 45 percent of the nation's energy, and oil and gas together supply 73 percent. In the year just ended it is estimated that the total U.S. demand for oil was 6.2 billion barrels, of which 41 percent had to be imported. (This marked a substantial decline, however, from the 1979 demand of 6.8 billion barrels, 44 percent of it imported.) The nation's proved reserves of oil are about 30 billion barrels, not counting reserves on the North Slope of Alaska, which may add another 10 billion to 15 billion barrels. Since the beginning of the oil industry in 1859 roughly 100 billion barrels of oil have been pumped from U.S. wells, most of them lying within the 48 contiguous states or only a short distance offshore. One would like to know what fraction of the nation's initial geological endowment of oil has been exhausted. Is there more oil remaining undiscovered than has yet been extracted, as some geologists contend? How much of it is held in giant fields and how much in fields too small for their oil to be recovered economically? As we shall see, the historical record can be used to test various hypotheses about the size and character of the remaining oil reserve.

One of the reasons for the uncertainty in national energy policy is that makers of policy lack the information needed to choose rational objectives. A case in point is the domestic response to international increases in the price of oil. Domestic drilling was predictably galvanized in 1974. According to the annual review of the American Association of Petroleum Geologists, the total length of the holes drilled in 1974 increased by 10.2 percent (to 29,000 miles) over the amount drilled in 1973, and the number of new wells increased by 19.2 percent (to 32,900). Most of the holes served to develop known fields, but new-field wildcat wells increased by 13.3 percent to 5,562, and 805 new fields of oil or gas (or both) were discovered. This was the largest number of new fields since 1956. The success ratio (the number of new fields divided by the number of new-field wildcat wells) was 14.5 percent, the highest in history. The incentive of a large price increase had every expected result but one: there was no net addition to the nation's reserves. The amount of oil and gas in the new fields was about the lowest on record. Clearly the average size of the new fields was much smaller than it had been in the past. Does this portend that future discoveries will also be small fields, which are expensive to develop, or are there giant fields like the great producers of the past still to be discovered?

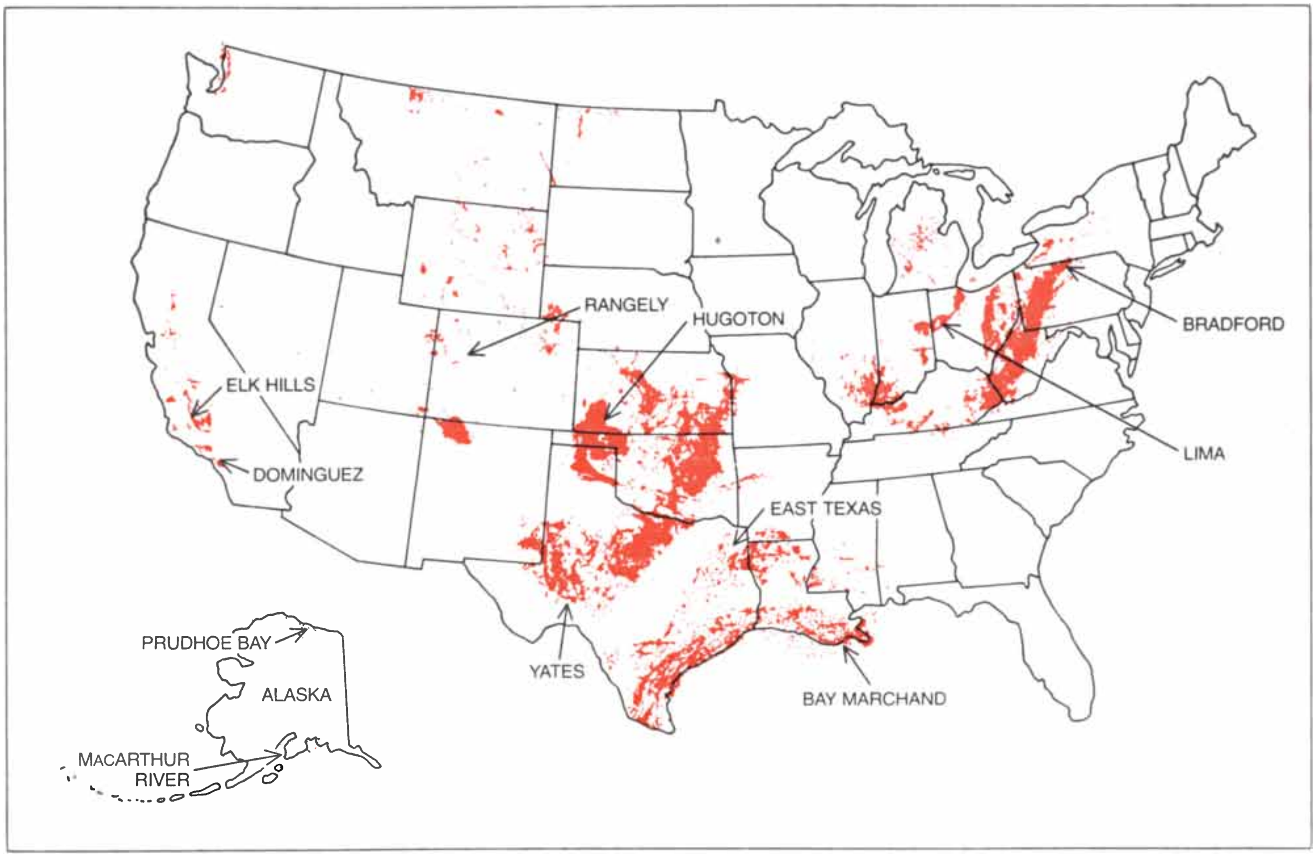
Knowledge of the total amount of recoverable oil and gas and the size distribution of the fields are important factors that must be considered in determining national petroleum policy. Here I shall focus on the oil in a finite area, the long-explored land area of the 48 contiguous states, thus excluding the relatively unknown areas of Alaska and the continental shelf, and in addition the overthrust belt of the Rocky Mountains, in which interest has just begun to develop. By concentrating on long-explored areas one may hope to clarify the history of petroleum exploration in the U.S., where petroleum geology and petroleum-exploration techniques have reached a level unrivaled anywhere else. The history of oil discovery has often been obscured by changing the areas under analysis and by aggregating oil with gas. I shall describe briefly how perceptions of past petroleum discoveries and predictions of future discoveries have both influenced national policy. I shall stress the desirability of an early national inventory of petroleum resources, regardless of policy aims.

Oil is a finite resource, depleted by the exploitation of each discovery. The rate of discovery per unit effort of exploration depends on many factors, including chance. Discoveries are reevaluated after three years and again after six years, so that the latest complete set of data is the one for 1974. In that year the chance of finding a new oil or gas field by drilling an exploratory hole or a new-field wildcat well in the U.S. was about one in seven. For the three-year period 1973–75 the chance of finding a giant field, one holding 100 million barrels of oil or more, was about one in 2,800. (No giant field was found in 1974 but four were discovered in 1973 and two in 1975.)

The element of chance can be evaluated by considering the probability of finding a field by purely random drilling. For a given field that probability is merely the area of the field divided by the area being searched. For example, the East Texas oil field, with an area of some 570 square kilometers, is the largest in the 48 contiguous states. Although the total land area of the 48 states is 7.8 million square kilometers, the area of sedimentary, potentially oil-bearing rock is only 4.7 million square kilometers. The probability of finding the East Texas field and its 5.1 billion barrels of oil by drilling holes at random in likely areas is therefore 570 divided by 4.7 million, or one in 8,250. In Kuwait, which has estimated oil reserves of some 65 billion barrels, the probability of finding one giant field per random hole would be about one in 15.

For a group of oil fields the probability of discovery is the sum of their areas divided by the area of search. As each field is found the probability of finding the remaining fields decreases, and the decline is exponential. This probability model of oil exploration can be compared with the historical record, provided the record can be expressed in terms of area discovered per exploration effort expended. A useful measure of exploration effort is 100 million feet of exploration drilling, or a Hubbert unit (named for M. King Hubbert, who pioneered in the application of mathematics to the analysis of oil and gas reserves). Within the 48 states the first Hubbert unit was completed in 1902, the second in 1929, the third in 1937 and so on. The 14th Hubbert unit, equivalent to 33 holes all the way through the earth, was completed in 1965. Over most of the past 50 years a Hubbert unit has been roughly equal to 20,000 holes each 5,000 feet deep; earlier a unit consisted of more holes of less depth.

A few years ago, when I was working at the Scripps Institution of Oceanography, George Sharman and I developed a random-drilling model of oil exploration and compared it with a historical record of giant fields compiled by the oil consultant Michel T. Halbouty. The model made the prediction that over the period from 1937 to 1957 (the third Hubbert unit through the 10th) the area discovered per Hubbert unit declined exponentially at the same rate as the historical record. The exponential decline has apparently continued up to the present. The rate of discovery is now so low, however, having declined from several thousand square kilometers per Hubbert unit to less than 100 square kilometers, that the points on a graph tend to scatter and obscure the trend [*see bottom illustration on opposite page*]. The agreement between the random-drilling mod-



OIL AND GAS FIELDS in the 48 contiguous states are concentrated in five major regions where hydrocarbons derived from organisms that lived as much as 600 million years ago have been trapped in sedimentary rocks. The youngest oil-bearing formations in the U.S. are in California and on the Gulf Coast. The biggest oil discoveries in the 48 states have been in California and in East Texas. The map identifies 11 giant fields, most of which contained more than 500 million barrels of recoverable oil (or gas equivalent of oil). The Prudhoe Bay field on Alaska's North Slope, the largest on the continent, holds from 10 billion to 15 billion barrels. Since 1859, when oil was discovered at Titusville, Pa., roughly 100 billion barrels of oil have been pumped from wells in the 48 contiguous states and in areas immediately offshore. The U.S. rate of production is currently about 3.7 billion barrels per year, or 10.1 million barrels per day, of which a little more than 15 percent now comes from Alaska. Of the total land area in the 48 contiguous states, 60 percent, or 4.7 million square kilometers, consists of sedimentary basins with oil-bearing potential. The U.S. Geological Survey estimates that about 45 percent of the basins have been fairly well explored. The author evaluates the possibility of discovering oil in fields of various sizes in unexplored areas.

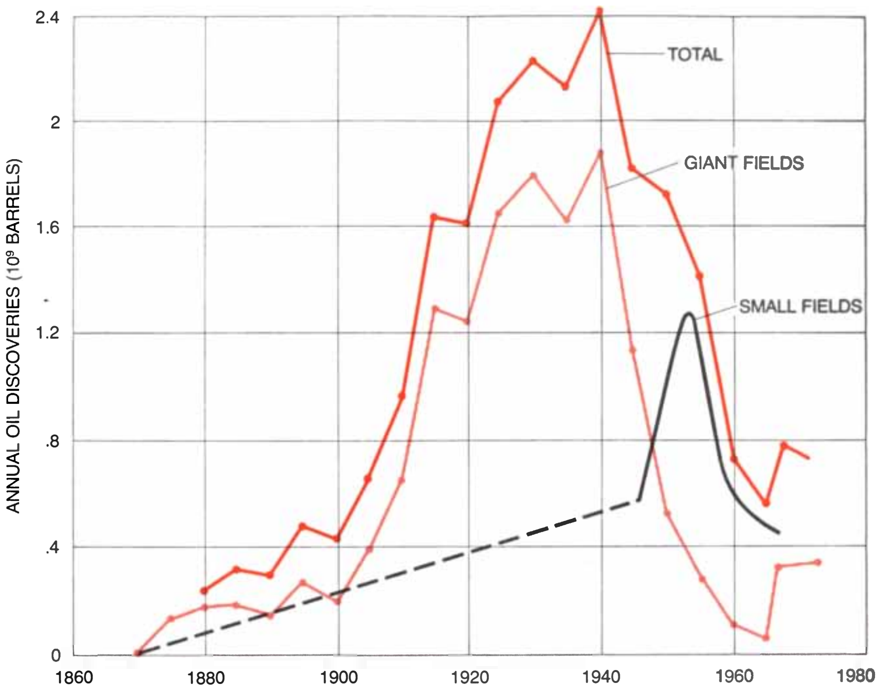
el and the historical record does not in itself imply that exploration was either more or less successful than random drilling. It does indicate that influences other than chance and the natural decline of a finite resource were fairly constant or tended to cancel one another.

Although the oil found per unit of drilling effort declined exponentially, the effort per year and per decade increased rapidly over the period from 1880 to the 1960's. As a result the quantity of oil discovered per year in giant fields increased until it peaked at about 1.8 billion barrels per year in the 1930's. After 1940 the quantity of oil that was discovered in giant fields declined precipitously.

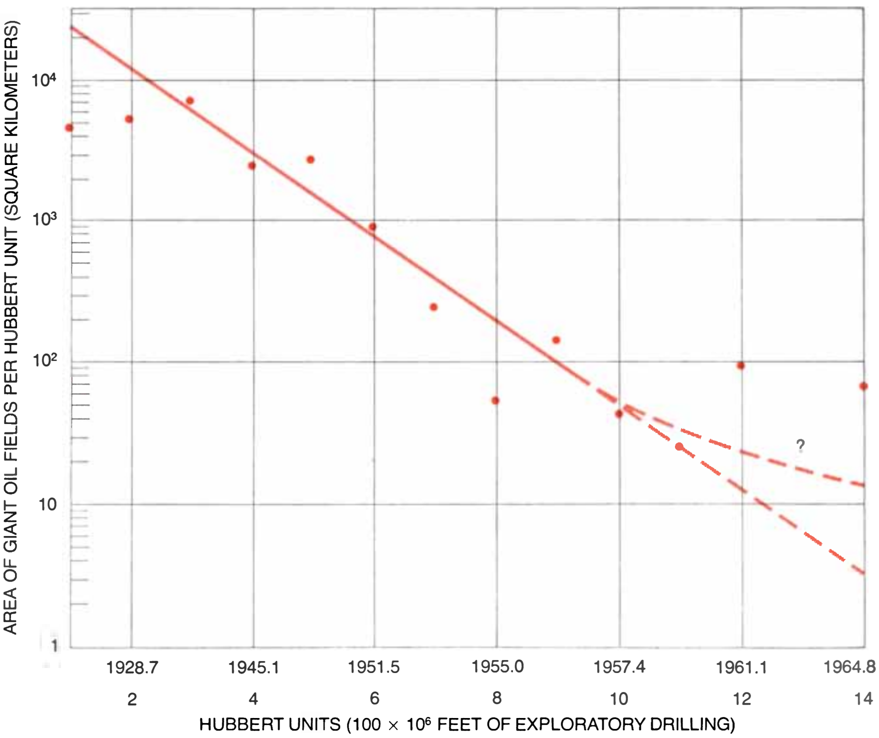
The early history of discovery of smaller fields is obscure, but since 1945 the discovery rate has been compiled annually by the American Association of Petroleum Geologists according to field size. The size classes have average volumes in millions of barrels as follows: class A, 100; class B, 37; class C, 17; class D, 3; class E, .3. For the purposes of this discussion it is also useful to distinguish two subclasses: E1, ranging downward from one million to 100,000 barrels, with an average size of 500,000 barrels, and E2, ranging downward from 100,000 to 10,000 barrels, with an average size of 50,000 barrels. Even shallow fields in these two subclasses were not regarded as being economically significant until the recent price increases.

Since 1945 the number of oil fields in the intermediate size classes B, C and D discovered per unit of search effort has been declining exponentially, just as it has for the class A giant fields. (The small class E fields appear to represent an exception to this rule since about 1955, a matter to which I shall return.) Moreover, the rate of decline in the discovery of intermediate-size fields is a direct function of field size as predicted by the random-drilling model. The fraction of larger fields among the annual discoveries decreases constantly. The rate of exploration, however, increased during most of the period, so that up to the early 1950's the intensified effort overshadowed the exponential decline in the effectiveness of the exploration. The peak discovery rate of 1.3 billion barrels per year in smaller fields was reached in the period between 1950 and 1953, and it has since declined by two-thirds in spite of accelerating exploration.

If one assumes that the discovery of smaller fields before 1945 was roughly proportional to the search effort, one can compile a complete history of discovery for the land area of the 48 contiguous states. The rate of discovery rose rapidly between 1900 and 1920, peaked in the 1930's and plunged in the next two



ANNUAL DISCOVERIES OF OIL in the 48 contiguous states have declined precipitously from a peak reached in about 1940, as is shown in curves averaged over five-year periods. Giant fields, which hold at least 100 million barrels of oil, were once the predominant source. Records of small-field discoveries have been kept only since 1945, when they began to predominate. Roughly 60 percent of all the oil discovered in the U.S. has been in some 300 giant fields.



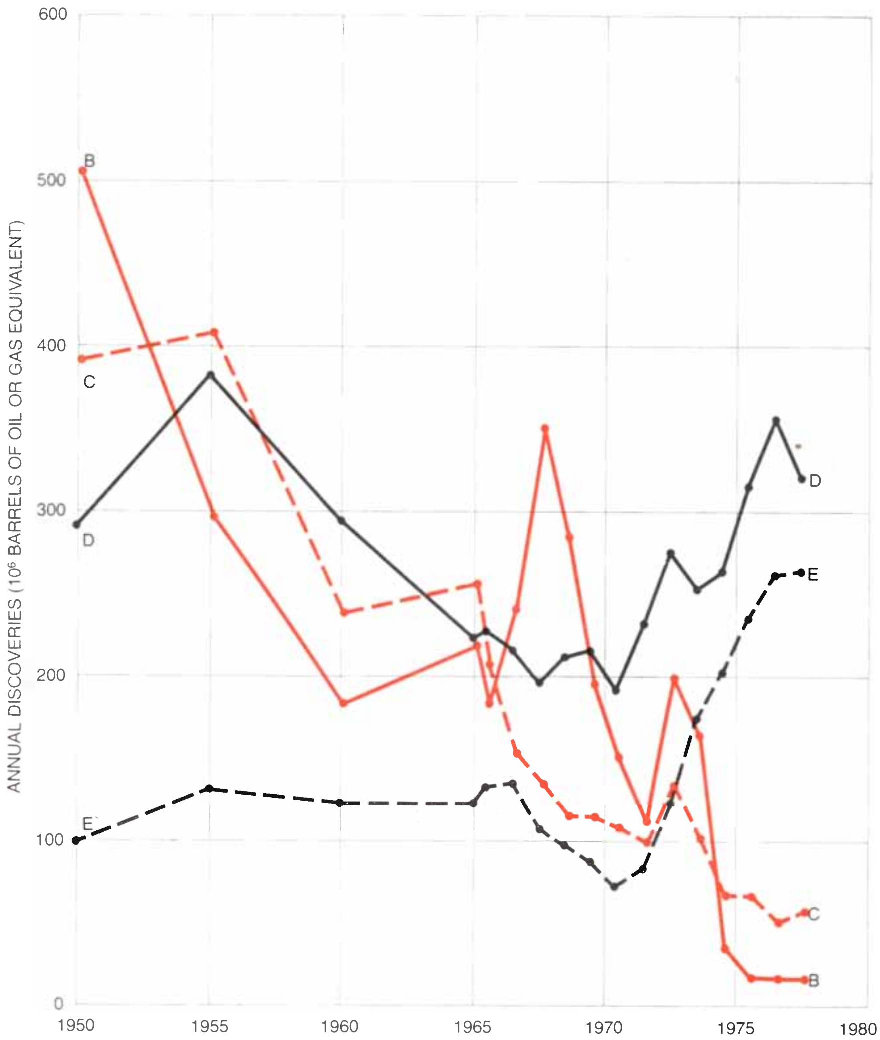
DIMINISHING RETURN per unit of exploratory effort is clearly seen in the history of discoveries of giant oil fields in the 48 contiguous states. One measure of exploratory effort is the Hubbert unit: 100 million feet of drilling in rock where oil has not previously been found. Discoveries of giant fields are recorded here in terms of area per Hubbert unit. The minimum area of a giant field is three square kilometers; a typical area is 15 square kilometers. Through the 14th Hubbert unit the area of giant fields discovered exceeded 20,000 square kilometers. In recent years discovery rate of giant fields has been so low (none at all in the 13th unit) that the trend line is uncertain. Some evidence suggests the upper broken curve is the more probable.

decades. Now it declines more slowly [*see top illustration on preceding page*]. The value for the reserves at any time can be obtained by subtracting the cumulative production from the cumulative discoveries. Reserves within the land area of the 48 states peaked in the middle 1940's, just as World War II was ending. Since then the decline in reserves has followed a curve that is symmetrical with the earlier increase.

Another important variable in the history of oil discovery is the size distribution of oil fields, which influences the cost of discovery and development. Until about 1940 most of the oil discovered was in giant fields. As late as the early 1940's such fields accounted for some three-fourths of the annual discoveries. By the early 1950's the situation was reversed, with most of the oil being found in smaller fields. At present discoveries are by volume about equally divided between giant fields and smaller ones. If the current exponential decline continues, however, a radical change will soon occur. By 1985 nearly all new oil will be from small fields. Within another decade oil may be obtainable mainly from fields that are smaller than the size considered economic for development today. Let me emphasize, however, that these generalizations do not apply to Alaska, the Atlantic and Pacific continental shelves adjacent to the 48 states or the newly discovered oil provinces of the overthrust belts of the Rockies. The sizes of fields that may lie hidden there are largely unknown. I am discussing only the familiar oil basins that have been drilled repeatedly.

It is possible that the fraction of oil discovered annually in small fields will increase even faster than is suggested by the general decline curves. This possibility can be seen by considering the rate of discovery of both oil and gas fields in all parts of the U.S. including Alaska and the continental shelf. The data are not as consistent as they are for the land area of 48 contiguous states but the trends nonetheless seem clear. Often when a new-field wildcat well is being drilled, there is no way of knowing in advance whether a field will be an oil field or a gas field, and many fields turn out to contain both oil and gas. Therefore an appraisal of the search effort must properly consider the sum of the oil and gas fields discovered.

It turns out that the rate of discovery of class B and class C oil and gas fields per search effort has declined exponentially, just as it has for oil fields alone. For the smaller class D and class E fields, however, the pattern is different. Since 1965 the number of class D oil and gas fields discovered per 10 million feet of exploratory drilling has if anything been climbing slightly. Since 1970 the number of class E oil and gas fields discovered for the same amount of drilling has virtually tripled, from 75 to 220. Data for the past six years are subject to reevaluation, but for more than a decade the changes have been small.



AVERAGE CLASS	OIL OR GAS EQUIVALENT
B	37 × 10^6 BARRELS
C	17 × 10^6 BARRELS
D	3 × 10^6 BARRELS
E	.3 × 10^6 BARRELS

PATTERN of recent discoveries of oil (or gas equivalent of oil) in fields of intermediate and small size shows that the volume of oil or gas discovered annually has increased in the smallest fields, which make up classes D and E, and has decreased in the somewhat larger fields of classes B and C, which rank below class A fields, or giant fields. The data for the past six years have not yet been given the reviews to which earlier data have been subjected and are therefore less reliable. Until the past few years class E fields (from 10,000 barrels to one million barrels) were considered too small to be of any economic significance.

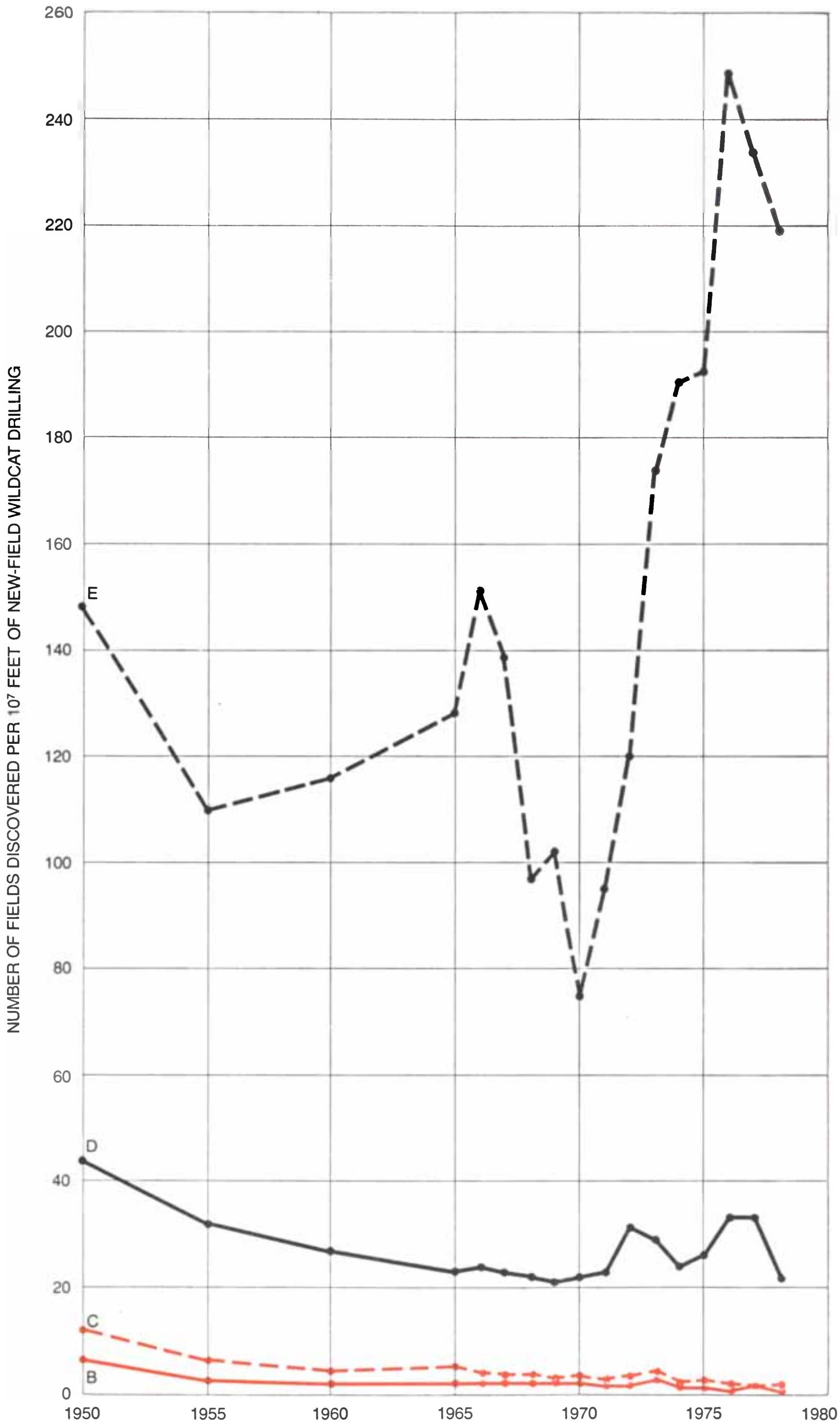
It is evident that the oil and gas industry has become increasingly adept at finding small fields even though the number of fields is finite and therefore the probability of finding one by random drilling is decreasing exponentially. There are two possible explanations, which are not mutually exclusive, for the industry's recent experience in terms of the random-drilling model of exploration. Perhaps the exploration system, meaning the combination of theory and technology, used to seek small fields is becoming more efficient. And perhaps oil fields that were regarded as being too small to be economic were formerly avoided and are now being actively sought.

Until a means is developed to evaluate these possibilities and thereby to calibrate the effectiveness of exploration, the methods I have described cannot be applied to predict the total number of small fields, even though oil-field discoveries appear to be declining. This raises the possibility that there are a great many small fields and makes it pertinent to estimate their maximum number and the amount of oil they may hold in reserve.

A possible maximum number can be derived from consideration of the observed size distribution of fields and the current chance of discovery. Detailed studies of the field-size distribution in the Denver-Julesburg sedimentary basin of Colorado have been made by the oil geologists J. J. Arps and Thomas G. Roberts. The distribution is what statisticians call log-normal, that is, when fields are plotted according to size and number on a logarithmic scale, the result is a symmetrical parabolic curve, reflecting the fact that the most abundant fields are of intermediate size. (On a linear, or arithmetical, scale the distribution curve would be a highly distorted bell.) The origin of the observed distribution is complex. The abundance of both large and small fields is limited by geological factors. The number of small fields is also limited by economics, by luck in exploration and by the definition of an oil field.

The size-frequency distribution of oil and gas fields in other regions is also log-normal, and it is reasonable to assume that with the same economic controls the log-normal distribution applies to the entire U.S. Class E fields are the most abundant, and the evidence suggests that there are roughly equal numbers of fields in the subclasses E1 and E2 (with a respective average size of 500,000 and 50,000 barrels). Fields in larger classes are much less plentiful, and presumably so are those in smaller classes.

The largest reasonable area of undiscovered fields of various sizes can be estimated from the present success ratio of one in seven for new-field wildcat drilling. Practically all the fields that are now being discovered and will be discovered in the future are of class E. Assume for convenience that all will be. The area of established oil basins in the 48 contiguous states is 2.6 million square kilometers. (The remaining 2.1 million square kilometers of sedimentary rock is too thin or of the wrong kind to contain oil.) If the industry's wildcat success rate of one in seven is no higher than what could be achieved by random drilling, the undiscovered area of class E fields should come to 370,000 square kilometers (2.6 million divided by seven). The area of a median-size E1 field is .9 square kilometer and that of an E2 field is .136 square kilometer. Therefore the maximum number of fields is roughly 370,000 in each class, corresponding to an indicated maximum oil reserve of about 200 billion barrels, or nearly sev-

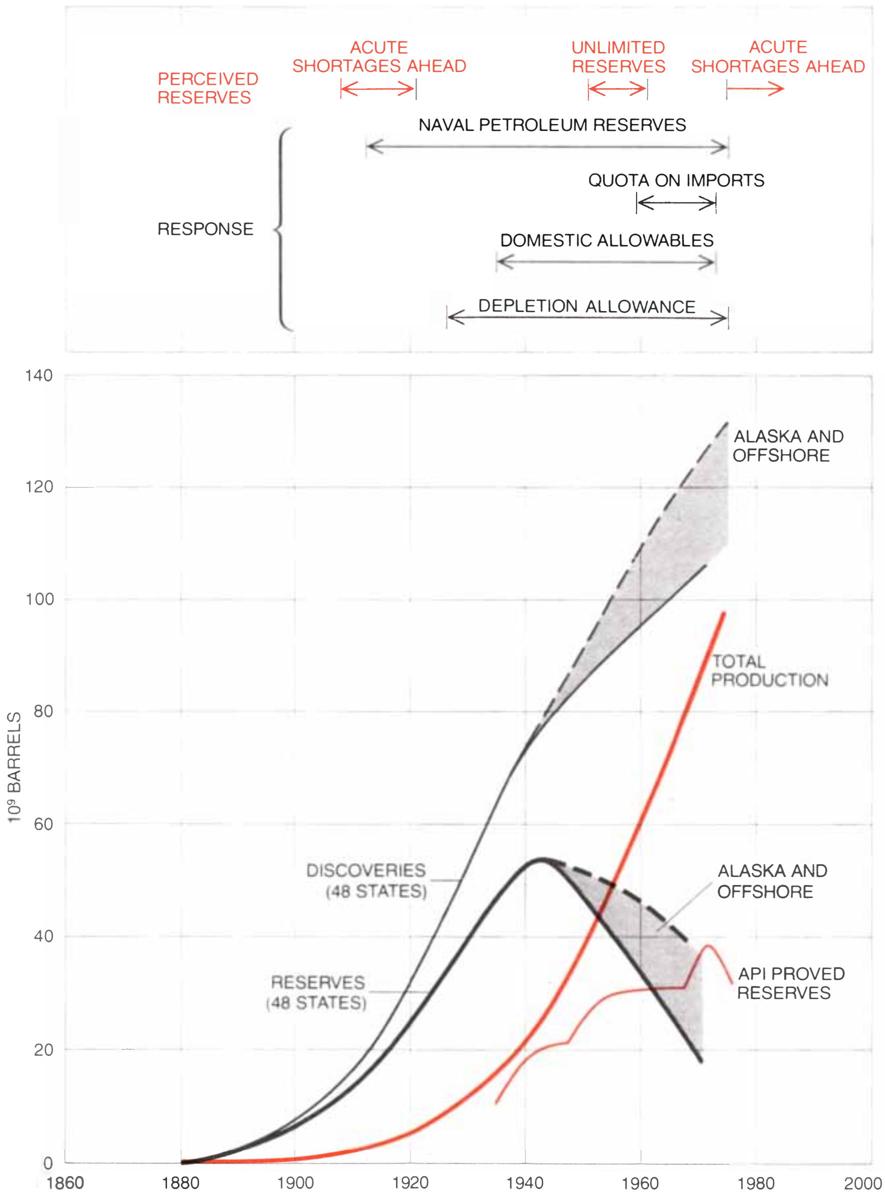


AVERAGE CLASS	OIL OR GAS EQUIVALENT
B	37×10^6 BARRELS
C	17×10^6 BARRELS
D	3×10^6 BARRELS
E	$.3 \times 10^6$ BARRELS

EFFECTIVENESS of oil and gas exploration has declined for fields of classes B, C and D as it has for giant fields. Here the unit of search effectiveness is taken to be 10 million feet of new-field wildcat drilling, which is more narrowly confined to the established oil-bearing basins than the Hubbert unit. The rise since 1970 in the discovery of class E fields per unit of drilling effort may represent an increase in the industry's ability to find small fields, an increase in the economic interest of small fields or both. The curves do not provide a foundation for predicting the number of small fields ultimately discovered.

en times the present U.S. proved reserves in known fields of all sizes.

If the industry's present success rate of one in seven is actually better than what could be achieved by random drilling, the reserve in class E fields must be proportionately less than 200 billion barrels. In other words, the success rate of random drilling is not known. It might be as low as, say, one in 70, in which case the undiscovered area of class E fields would be reduced by a factor of 10 from 370,000 square kilometers to 37,000 (2.6 million divided by 70 rather than by seven). The maximum oil reserve in class E fields would then be only 20 billion barrels instead of 200 billion.



PERCEPTIONS OF SHORTAGE AND GLUT have had a major influence on national oil policies, which have usually turned out to be counterproductive. The policies adopted in response to the nation's perceived reserves are indicated by the labeled arrows at the top of the illustration. The curves in the lower graph show the actual oil reserves in the 48 contiguous states (*heavy black curve*), calculated by crediting all the oil that will ultimately be produced in a field to the year of discovery and subtracting production thereafter. The "Proved reserves" identified by the American Petroleum Institute (A.P.I.) are the reserves actually identified year by year by that organization since 1925. The perceived reserves led the Federal Government to stockpile oil for the Navy and to give tax encouragement to exploration (in the form of a depletion allowance) just when the actual reserves were building up to a glut. The glut led the major oil-producing states to set quotas on production ("Domestic allowables"), not abandoned until 1973. A perception of abundance arose just as the actual reserves began to plummet, leading the Government to limit imports to encourage a healthy domestic industry for reasons of national defense. Even though the perception and the reality are now in closer agreement ("Acute shortages ahead"), the Naval Petroleum Reserves are being opened to development and the tax incentives for exploration have been eliminated for major oil companies.

Estimates between 20 billion and 200 billion barrels probably bracket the amount of oil remaining in class E fields within the boundaries of the 48 contiguous states. Curves representing these outside values yield plausible log-normal distributions when they are combined with the known size and abundance of fields in classes A, B, C and D. If the lower bound, 20 billion barrels, is roughly correct, one can conclude that most oil was initially in giant fields and that the remaining reserve is evenly divided among undiscovered fields of all sizes. If the upper bound, 200 billion barrels, is more nearly right, the total geological reserve was initially divided fairly evenly among fields of all sizes but by now is almost entirely in very small fields.

We have been considering the discovery history as it is now reconstructed by crediting all the oil in a field to the year of discovery. This would have been the perceived history if the maintenance of a systematic inventory of reserves had been part of national policy. Since it was not, the perceived history was quite different from the retrospective history.

In the decades before World War I the U.S. produced between 50 and 90 percent of the world's oil and dominated the world petroleum trade more than the Middle East does now. The cost of oil rarely reached $1 per barrel and waste was prodigious. By 1908, however, there were predictions of shortages and discussion of a reduction in exports. President Taft withdrew three million acres (12,000 square kilometers) of public lands from petroleum leasing in 1909, and in 1912 he designated the withdrawn acreage as the Naval Petroleum Reserves to conserve oil for national defense. Soon thereafter the keels were laid for the first two oil-burning battleships. The belief in impending oil shortages lasted through World War I, and in 1926 a depletion allowance was established to encourage exploration. (Such an allowance reduces a well owner's property taxes in proportion to the amount of oil withdrawn on the theory that his capital asset is being depleted.)

In 1925 the American Petroleum Institute (A.P.I.), representing the oil industry, prepared an outstanding report predicting the supply and demand for oil in the U.S. It identified proved reserves of 5.3 billion barrels of flowing and pumpable oil and a much larger amount that could be recovered by the secondary processes already known at the time. The A.P.I. also surmised that there were enormous, although unpredictable, amounts of oil in undiscovered fields. Its report allowed for expected conservation, for improved gasoline-engine efficiency and for new technology to exploit oil shale. The report accurately forecast "no imminent danger of the exhaustion of the petroleum reserves of the United States."

Within a few years new discoveries released a flood of oil. In 1930, when the discovery of the East Texas field triggered economic chaos, the governor of the state called on the National Guard to stop production. Many states instituted "conservation" programs to limit production to some economically sensible level.

Up to the 1960's there was little perceived incentive for the Federal Government to assume responsibility for predicting oil reserves. The industry, through the A.P.I., continued to issue its own annual reports on proved reserves, which consistently indicated about a 15-year supply at the rate of consumption then prevailing. After 30 years of such reports it appeared that new domestic discoveries would meet future needs indefinitely. The problem in the 1950's was that the enormous discoveries in the Middle East threatened to stifle both domestic production and the incentive to explore, since the imported oil would be cheaper than the domestic. The Government accepted the thesis that a healthy domestic industry constantly finding new oil was essential for national security. In 1959 quotas limiting imports were instituted.

Within the past few years the national perception of the oil situation has changed dramatically. Imports have risen sharply as quotas have been lifted, limits on domestic production have been dropped, the depletion allowance has been revoked for large companies (but not for small ones) and the Naval Petroleum Reserves are being developed. After half a century of complacency nearly all the proposals for increasing conservation, raising gasoline-engine efficiency and extracting oil from shale have reappeared.

The perceptions that justified the various policies bearing on oil tend to seem curious when they are compared with the real history that has now been reconstructed. The Naval Petroleum Reserves were established just as oil reserves became abundant, and they are going to be drained instead of being saved for the Navy just as other reserves are being depleted. The depletion allowance to encourage exploration was also instituted as oil began to be abundant. It was retained throughout the period of maximum discovery of reserves. Then as reserves became short and field sizes declined the depletion allowance was revoked for large companies. Import quotas were established only when reserves were declining rapidly, thereby accelerating the process. It is difficult to find fault with anyone, or with any institution, for acting on the perceived information. It is quite clear, however, that a rational national policy for the development of petroleum resources or other finite resources calls for more information than has been available in the past.

The ignorance of the past with regard to domestic oil resources can be largely explained by the fact that the extractive industries are traditionally secretive and Government officials saw no reason either to demand information or to encourage its acquisition. Indeed, Government policies virtually ensured ignorance of oil reserves: oil, once discovered, was usually subject to property taxes. In addition the industry could absorb only limited amounts of new oil, which meant that investment in exploration amd development could not be recovered rapidly. Hence an oil company that discovered what appeared to be a small, shallow field and had the surrounding area under lease had every reason to expand development outward and downward as slowly as contracts allowed.

Even so, the "self-control" exercised by the owners in discovering the successively deeper producing horizons, or oil-bearing strata, in the Dominguez oil field in California was remarkable. The "1 st Callender" horizon was discovered at a depth of 3,950 feet in 1923, just when three giant fields in California were overwhelming the market. In California more than 1,400 wells were shut down that year. In 1924 the "2nd Callender" horizon was discovered at 4,250 feet, and so it went with increasing depth: successive horizons were discovered at 4,530 feet in 1925, 4,830 feet in 1927, 5,300 feet in 1931, 6,360 feet in 1933 and 7,050 feet in 1936. It took 12 years to explore a vertical range of less than 3,000 feet.

To be sure, oil fields are often complex three-dimensional puzzles that take time to solve. The accurate estimation of reserves is not simple and is always subject to revision. All the same, much of the ignorance about reserves in past decades may have owed less to geological factors than to institutional and economic ones. Such factors are susceptible to change if it is deemed in the national interest to base oil policy on a broader foundation of information than was available in the past.

Any discussion of undiscovered petroleum reserves must deal with three important variables: the total amount of oil potentially discoverable, the size distribution of fields and the maximum rate of extraction. Most estimates of reserves have dealt only with the total amount. Even so, the estimates are so variable that they elicit skepticism about their validity. Estimates by different organizations or even by the same organization at different times can differ by an order of magnitude. Attempts to conceal uncertainties under a layer of statistics are exposed when one year's minimum estimate is greater than the next year's maximum estimate. The estimates in this article also incorporate uncertainties, but they have the virtue of predicting all three important aspects of reserves and taking into account factors that have not been incorporated in most estimates.

For a closely explored region such as the land area of the 48 contiguous states the reserve estimates have some implications about exploration success that have not been widely appreciated. The most favorable interpretation of the industry's ability to find oil is that after some two billion feet of exploratory drilling most of the oil has been found. Accordingly any hypothesis about large undiscovered reserves leads to a less favorable interpretation. By the same token the most favorable interpretation of the industry's ability to find giant fields is that almost all of them have been found. This follows from the very low rate of discovery extending back more than two decades.

A useful way to make quantitative comparisons of the implications and therefore of the credibility of estimates of undiscovered reserves is to analyze what would have happened if a region holding such reserves had been explored by random drilling. My colleague Sharman and I made the most favorable interpretation of industry's ability to find giant fields as of 1964, that is, we assumed that Halbouty's inventory of known giant fields was exhaustive. This was only a convenient simplification, of course, because a giant oil field is still discovered in the U.S. once every few years. It seemed justified, however, by the low rate of discovery and the desirability of having a consistent body of data.

We programmed a computer to sample randomly for giant oil fields in U.S. sedimentary basins. When the computer program simulated the discovery of giant fields over the first Hubbert unit (1880 to 1902), when all the fields known today were targets for search, the results of the simulation could be compared with the actual discoveries in the period. In order to continue the comparison the fields actually discovered during the first Hubbert unit were eliminated. The computer then searched randomly for the remaining giant fields during the second Hubbert unit, the third and so on.

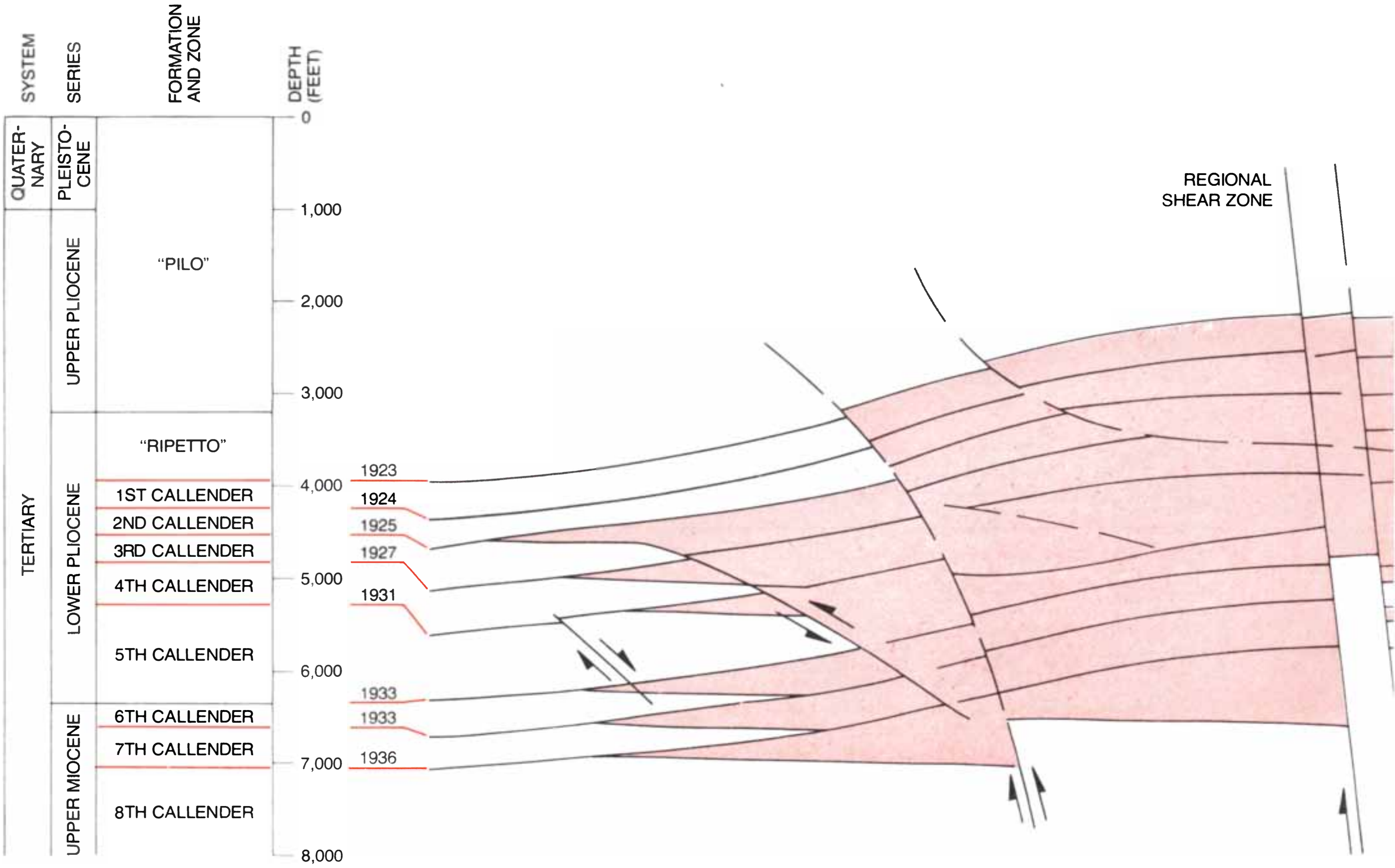
The actual discoveries of the industry over the first Hubbert unit were some 4.6 billion barrels of oil in fields with an area of 4,100 square kilometers. The average discovery in 10 computer simulations of random drilling in the same Hubbert unit was 22.7 billion barrels in an area of 13,780 square kilometers. By the time of the third Hubbert unit the historical rate of discovery was within the range of the 10 simulations. Thereafter, in the course of the remaining 11 Hubbert units, actual and simulated discoveries declined together [*see illustration on page 64*].

It is necessary to distinguish between success in finding fields as defined by area and in finding fields as defined by volume. The computer simulation more closely approximates the historical record in finding area than in finding volume, and it is volume that counts. There is also some question whether the definition of a Hubbert unit given the computer simulation may not include holes that actually were drilled in exploring the limits of known fields rather than in seeking new ones. Other factors, however, may have a balancing effect. For example, when the simulated search found the same field twice in the same Hubbert unit, it was credited with only one discovery and was not given a free turn to drill again. Moreover, most of the time the simulation drilled deeper than the actual historical average; hence it wasted drilling feet (since the simulation could not strike oil at depths greater than those of the actual fields in Halbouty's inventory). Finally we may find in due course that the computer "discovered" fields still awaiting discovery.

Our conclusion is that even after giving the oil industry the benefit of various assumptions, the discovery of giant oil fields in the land area of the 48 contiguous states has been about what would have been achieved if the sedimentary basins had been drilled at random. We then have the paradox that earth scientists unquestionably can find favorable geological structures, such as anticlines, in ways far more efficient than throwing darts at a map, and that 78 percent of the giant fields found up to 1964 are in structural traps such as anticlines.

One possible explanation for the unexpected success of the computer simulation is that the ratio of the area of the oil in structural traps to the total area of the traps is the same as the ratio of the area of the oil in all sedimentary basins to the total area of the basins. That is equivalent to saying there is no advantage in being able to find structural traps. This possibility may seem to smack of a denial of the usefulness of the scientific approach, in view of the fact that most new-field wildcat wells are drilled over geological structures. Nevertheless, such eminent petroleum geologists as Hubbert and Halbouty have stated within the past decade that oil may exist in traps where it is not ordinarily sought, a prediction borne out by the recent discoveries of an increasing number of fields in the overthrust belt of the Rockies near the Utah-Wyoming border.

A second explanation for the success of the computer simulation assumes that the oil industry actually enjoys a substantial scientific advantage over random drilling in exploration, and that any advantages resulting from geological and empirical knowledge are exactly neutralized by industrial and governmental practices and restrictions. I might emphasize that the computer simulation assumed ignorance of all geology except that oil is found in sedimentary rocks. If the simulation had included information about the distribution of oil seepage or the thickness of sediment, it would have been much more successful.



RESTRAINT IN EXPLORATORY DRILLING is encouraged by the common practice of levying property taxes on the volume of declared oil reserves. A notable example is the restraint exercised by industry in drilling the almost continuous oil sands of the giant Dominguez oil field in California. Following the discovery of oil at the "1st Callender" horizon at a depth of 3,950 feet in 1923, it was 13 years before the drilling was extended a further 3,000 feet. In that period additional oil was "discovered" at seven lower horizons. The

And if it had simply drilled on a grid, it would have found all fields of more than a certain size.

I prefer to assume that the standoff between drilling with knowledge and drilling without it is coincidental rather than expression of some kind of Parkinson's Law to the effect that the sum of all human efforts is often zero. Indeed, our computer exercise suggests there was a time when the sum was not zero: the time before 1902, when the first Hubbert unit was being drilled. In the early days the level of science in petroleum prospecting ranged from low to negligible. The oil industry then found only about a third of the area of giant fields that would have been found by random drilling. One must conclude that science had to increase the success rate threefold to raise it to the level of the random-drilling model.

We have been considering the hypothesis that all the giant oil fields within the 48 contiguous states have been found. This is the most favorable possible evaluation of the oil industry's exploration success. An example of an implicitly less favorable evaluation follows from a hypothesis advanced in 1968 by the oil geologist J. D. Moody and his colleagues. They proposed that from 25 to 28 giant fields, each holding more than 500 million barrels of oil, remain undiscovered in North America, including Alaska, Canada and the continental shelf.

Let us assume that the undiscovered fields are comparable in size to known fields and that they are evenly distributed in the regions where the known fields have been found. On this assumption there should be 22 giant fields with a total area of 3,700 square kilometers in the land area of the 48 contiguous states, the remaining three to six being in Alaska, Canada or offshore. We reran our computer simulation, adding this area to the target sought by the simulation after each Hubbert unit. At first it made little difference, but as the remaining fields in the historical inventory were discovered and were eliminated from the search the hypothesized 22 fields loomed ever larger. By 1955 a random search by computer simulation would, by finding more and more of the hypothesized fields, have found roughly 50 times more oil than was actually discovered up to that point, which coincided with the completion of the eighth Hubbert unit of exploratory drilling. Therefore if one accepts the Moody hypothesis, one must believe the oil industry would have done 50 times better by drilling random holes than by following its geological knowledge. A reasonable conclusion is that the Moody hypothesis of there being more than 20 500-million-barrel fields remaining to be discovered in the land area of the 48 contiguous states is improbable. (This conclusion does not apply, of course, to Alaska or to the continental shelf.)

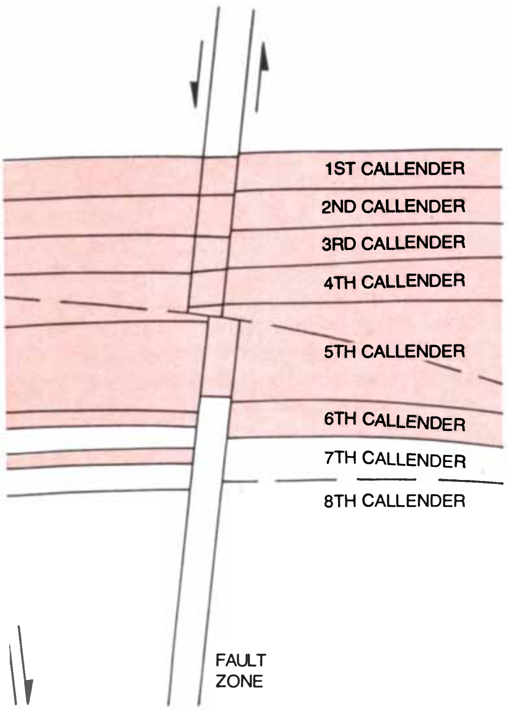
Other predictions of undiscovered oil are more difficult to evaluate because they do not give the assumed size distribution of the undiscovered fields. Presumably the predictors have assumed that the size distribution of the undiscovered fields is the same as the initial size distribution of all fields, or that it is the same as the average size distribution of all the discovered fields. It might appear that either assumption implies the oil industry has been drilling randomly; actually the assumptions are far less complimentary. The industry finds a steadily decreasing proportion of large oil fields per unit of search effort. Any hypothesis assuming that the fraction of giant fields is as large among undiscovered fields as it is among discovered ones implies that in its search for giant fields the industry has been drilling systematically in the wrong places.

Consider the hypothesis of Alfred D. Zapp of the U.S. Geological Survey that the chance of finding oil in undrilled sediment is as great as the chance of finding it in sediment that has been drilled. Zapp assumed that less than half of the nation's total reserve of oil has been discovered. If one assumes that the undiscovered oil fields have the initial size distribution, random drilling of the 48 contiguous states would initially have discovered twice as much oil as we found in running our basic computer simulation. At present the Zapp reserves would be the same as the initial reserves in the simulation. Half of the giant fields would still be virgin. A random search today would find, after one Hubbert unit of drilling, what the basic simulation found in the first Hubbert unit, namely giant fields with an aggregate area of 13,780 square kilometers holding 22.7 billion barrels of oil. Because the oil industry is now discovering only about 100 square kilometers of oil in giant fields per Hubbert unit the Zapp hypothesis implies that the industry would be 130 times more successful if it drilled holes at random.

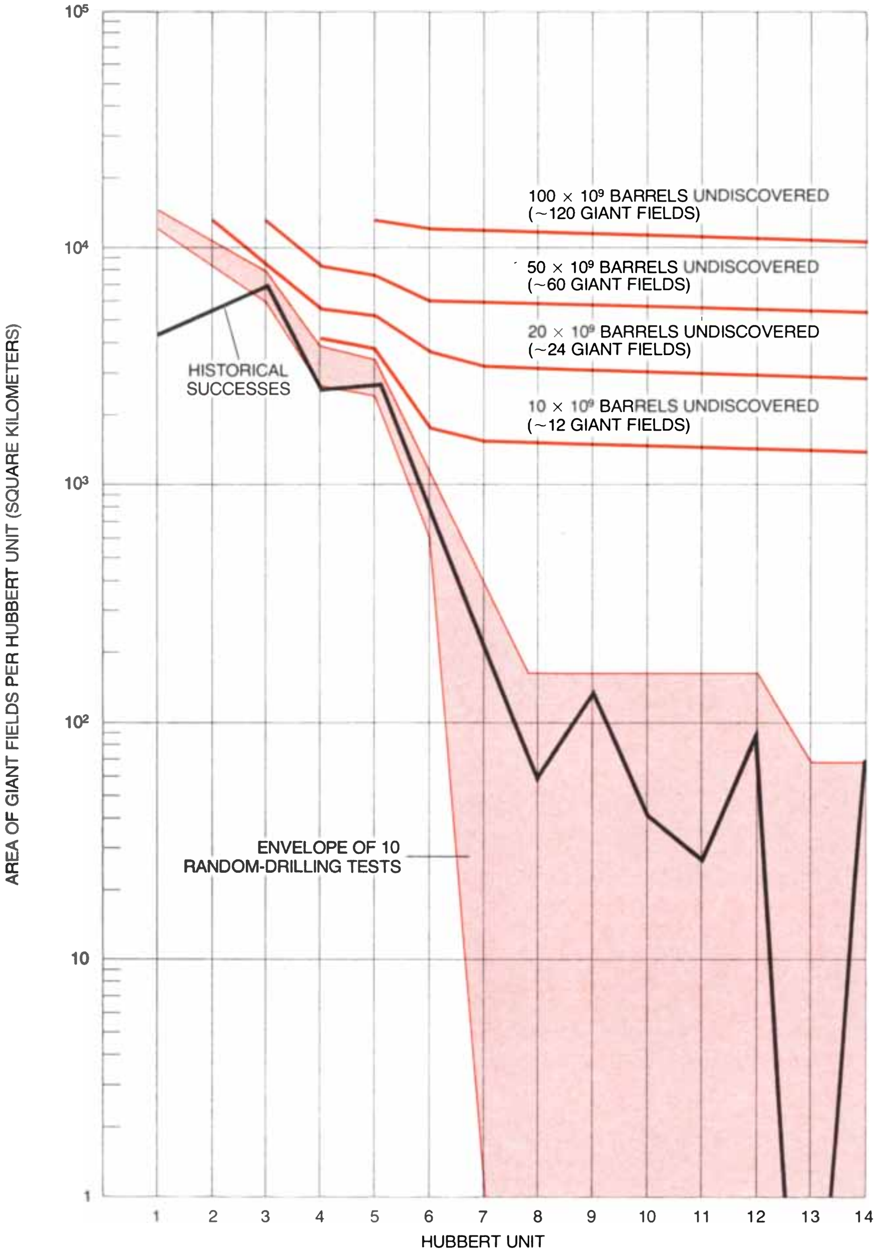
What if the Zapp hypothesis, or others that predict large reserves of undiscovered oil, incorporated the prediction implied by the exponential decline curves, which is that almost all future discoveries will be in small fields? As we have seen, any estimate up to about 200 billion barrels of undiscovered oil is allowed by the historical data, provided that by far the largest part of the oil resides in class E fields. It may seem that the possibility of large reserves in tiny fields is hardly compatible with the exponential decline in oil discovered per Hubbert unit. Such is not the case. The overall decline is merely the sum of the different exponential decline rates for fields of various sizes. As the larger fields are exhausted the overall rate of exploration effectiveness becomes that of the smallest fields. And, as we have also seen, this rate has been increasing. Lest it seem that a new reservoir of oil has been uncovered, I should repeat that class E fields have been less than the minimum size considered to be of economic significance. Until very recently they have not been desired targets.

The other thing to remember about the class E fields is that in recent years only about 30 million feet of new-field wildcat holes have been drilled per year. At 5,000 feet per hole and a success rate of one in seven this comes to about 860 new class E fields per year. On the assumption of equal numbers of fields of the subclasses E1 and E2, and the average volumes given above, one arrives at a total volume of roughly 235 million barrels, or barely two weeks of domestic consumption at current rates. Consequently even if the reserves in tiny fields are very large, they offer little hope of solving the country's short-term energy problems.

We have been considering the discov-



recoverable volume of the field is now placed at 265 million barrels, of which 243 million barrels had been extracted by 1967. The field has an area of some seven square kilometers.



RANDOM-DRILLING MODEL of oil exploration, constructed with the aid of a computer, was compared with historical discoveries of giant oil fields by the author and George Sharman at the Scripps Institution of Oceanography. The computer was programmed to "drill" wildcat holes 5,000 feet deep at random in the 4.7 million square kilometers of potentially oil-bearing sedimentary rock within the land area of the 48 contiguous states. If the computer simulation "struck oil" in a known giant field, it was credited with the area of that field. The jagged black curve represents industry's actual success in finding giant oil fields in successive Hubbert units of drilling. (The curve connects the points in the bottom illustration on page 57.) The area in light color represents the envelope of 10 tests run by the computer. For the first two Hubbert units the computer simulation widely surpassed the historical record. In the first Hubbert unit alone the random-drilling method "found" 13,780 square kilometers of giant fields holding 22.7 billion barrels of oil, equal to a third of all the oil discovered in giant fields. In the first Hubbert unit of actual drilling the industry found 4.6 billion barrels of oil in 4,100 square kilometers. From the third Hubbert unit onward the historical discoveries fell within the envelope of the 10 computer simulations. The upper family of four curves shows what the computer simulation would discover on the average if it were programmed to search not only for known giant fields but also for hypothetical giant fields presumed to hold 60 percent of the undiscovered resources with an aggregate volume of from 10 billion to 100 billion barrels, as has been proposed by some geologists. If each giant field held 500 million barrels of oil, there would be from 12 to 120 giant fields in all. In the reprogrammed model they are distributed randomly in areas not yet drilled. Under these new conditions the random search is far more successful than the actual search. Beginning at about the seventh Hubbert unit the industry would have been 10 to 100 times more successful in finding giant fields if it had drilled at random. Since this seems unlikely, one assumes not many giant fields remain to be discovered.

ery of oil by existing exploration systems, and we find that the prospects of discovery in the land area of the 48 contiguous states are bleak. Except for tiny fields, the success rate continues to decline exponentially just as the random-drilling model predicts. An improvement in the exploration system merely accelerates the decline of a finite resource. Must we therefore seek elsewhere for the oil and gas needed to meet our short-term energy demands? Apparently so, unless there are unknown or obscure types of oil fields that are not being targeted by the exploration systems now in use.

The big oil fields in structural traps in the land area of the 48 contiguous states are becoming scarcer, and the little fields cannot be emptied very fast. The nation's only potential domestic source of flowing oil and gas fields are therefore either anomalous types of oil fields or the structural traps on the public lands of Alaska and the continental shelf. The development of these areas is subject to the evaluation of potentially harmful effects on the environment and on the economic and social structure of nearby communities. The difficulty in making such evaluations is the paucity of knowledge about the agent that is supposed to cause the harmful effects.

The procedures for leasing tracts on the outer continental shelf amount to a double-blind auction. The Government brings a crate out of a warehouse, and the oil companies bid for the contents—the numbered tracts—without being allowed to inspect them. The law provides that companies can get permission to drill on promising structures on the outer continental shelf without committing themselves to development. Curiously, the industry has not so far availed itself of that opportunity. It continues to drill "stratigraphic tests," which are holes in places where everyone agrees there is no possibility of finding an oil field. Such test sites are not, for that reason, included among the desirable sites being auctioned. Therefore the participants in the blind auction are not totally without information. They are able to look into certain other "boxes," just not the ones being auctioned.

A common misconception is that the oil companies know how much oil is under the continental shelf but the Government does not. I believe this is mistaken: both sides in the auction are equally ignorant. On the Government side data have been compiled on the amount of oil per tract estimated by the U.S. Geological Survey compared with the amount actually found after development. There is some correlation between the two sets of figures, but the prelease estimates are often 10 times too large or too small. Comparable figures from the industry are not available in the public record. In their absence one can only analyze the dollar bids for tracts, which include many factors besides estimates of resources. Consideration of such bids for one large group of offshore tracts shows that the number of successful bids by any one company is roughly proportional to the total number of bids by that company. This suggests that an element of randomness enters into the leasing. If an oil company makes 10 percent of the bids, it requires little science to obtain 10 percent of the tracts.

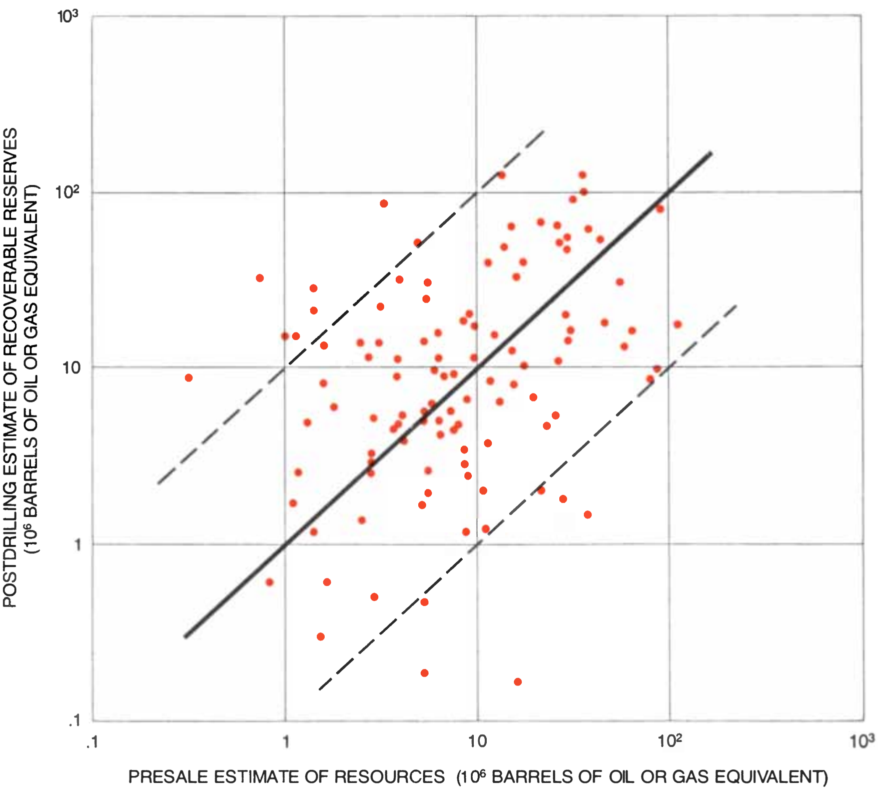
I should explain that companies are not the only "bidders" at such auctions; the Geological Survey also takes part. Its purpose is not to make high or winning bids but only to set a fair value for each tract to ensure that the Government does not accept a low value without cause. (For the tracts considered above the Geological Survey was a conservative bidder, with few high bids.) It is this difference in objectives that sometimes causes a public outcry about the disparity between Geological Survey bids and winning bids for certain tracts. It is said that once again the oil companies swindled the Government. As a matter of fact, for the tracts considered here all the bids made by everyone were too high; the data are taken from leases that were subsequently abandoned as being worthless. Such abandoned leases are currently the only ones for which bids can be compared with ultimate values. Exploration for oil and gas is still a gamble. It is only after promising structures are extensively drilled that the resources can be known and the environmental costs assessed.

What of estimates of oil reserves in and off Alaska and on the outer continental shelf? Is there enough oil there to provide energy independence for the U.S. over a significant period, or just more empty structures? A refined energy policy depends in part on getting answers to such questions. Are existing national policies suitable for getting reasonably quick answers?

The policies are based on the general assumption that there are important reserves of oil in Alaska and under the continental shelf, that there are no alternative sources of fluid energy in the short run and that the local environmental costs of development will be more than justified by the general economic benefits. The policies are designed to facilitate the development of regions sequentially in an order determined by some combination of the supposed reserves and the supposed environmental costs. The implementation of the policies has been frustrated by lawsuits and delaying tactics by groups entertaining doubts about the assumptions.

Many of the elements of the controversy arise from the lack of information and would be minimized if a higher priority were given to the speedy acquisition of a national inventory of oil and gas reserves. What kind of inventory would suit the present purposes? A complete inventory would require identification of the amount of oil in all the undiscovered fields. Since the physical resources for drilling are limited, a new-found field cannot be intensively inventoried much in advance of development without losing urgently needed current development. What seems to be a dilemma can be solved in the short run, however, because of the size distribution of oil fields. Most of the oil that can be extracted quickly is in giant fields. The only way for the U.S. to meet most of or all its requirements for oil over the next two decades is to discover a substantial number of giant fields, no fewer, say, than 100. Intensive geophysical exploration has already identified many potential giant oil-bearing structures on Federal lands. The essence of the necessary inventory can be achieved by speeding the exploration of the 50 to 100 leading prospects for giant new fields.

Possible ways to speed the inventory of giant new fields include the following that have been advocated at various times. I merely list them. The Government could do more of the exploration itself, as it has been doing in the National Petroleum Reserves in Alaska. The oil industry could be required or induced to drill structures prior to leasing. The Government could allow single companies to lease scattered tracts instead of encouraging many companies to bid on tracts within compact areas. Tract sizes could also be increased from the current standard nine square miles to areas 100 to 1,000 times larger, as is the practice in other countries. Finally, the time between leasing and the beginning of exploration could be shortened. The choice among these and other possible legal and institutional arrangements for compiling the inventory is largely dependent on political and economic philosophy, which lies beyond the scope of this discussion. Regardless of philosophy, however, it should be recognized that without a rapid inventory of its oil and gas resources the nation risks repeating the mistake of basing policy on illusion rather than on information.



DIFFICULTY OF ESTIMATING OIL RESOURCES in advance of drilling is demonstrated by comparing the U.S. Geological Survey estimates of reserves of oil (or gas equivalent of oil) in 108 offshore tracts with the amount actually found. Points that fall on the 45-degree line passing through the origin of the graph identify tracts where the estimate and the actual deposit agree exactly. The broken lines running parallel to the central line indicate where the predrilling estimates were either too high (*lower line*) or too low (*upper line*) by a factor of 10.

Viroids

They are the smallest known agents of infectious disease: short strands of RNA. They cause several plant diseases and possibly are implicated in enigmatic diseases of man and other animals

by T. O. Diener

There are certain infectious diseases of plants that cannot be associated with any of the usual causative agents such as fungi, microorganisms or viruses. The only notable thing about a plant afflicted with one of these diseases seems to be that small molecules of an unusual form of the genetic material ribonucleic acid (RNA) can be isolated from its tissues. Such molecules cannot be detected in a healthy plant of the same species; if they are introduced into a healthy plant, they proliferate and give rise to the characteristic symptoms of the disease. In other words, these RNA molecules, which are called viroids, are the causative agent of the disease in question. They are the smallest known agents of infectious disease.

Viroids are much smaller than viruses and much simpler. Whereas a typical virus consists of genetic material (either RNA or DNA) surrounded by a protein coat, a viroid is nothing more than a very short strand of RNA. To date viroids have been identified with fewer than a dozen specific diseases, all of which affect higher plants, but there are indications that they may also cause animal diseases, perhaps including some rare nerve diseases that affect human beings.

In certain respects the discovery of viroids recalls the discovery of viruses some 75 years earlier. The conclusive demonstration by Louis Pasteur, Robert Koch and others that microorganisms, notably bacteria, were responsible for a number of infectious diseases gave rise by the 1890's to a general assumption that all infectious diseases must be caused by microorganisms. Investigations by the Russian microbiologist Dmitri Ivanovski and the Dutch botanist Martinus Beijerinck of a plant disease, tobacco mosaic, disproved that generalization by showing that the causative agent of the disease would pass through filters with pores small enough to retain bacteria. Their concept of a "filterable virus" finally led to the identification of virus particles and an understanding of their role as pathogens.

A new generalization took hold: all infectious diseases of plants and animals were widely held to be caused either by microorganisms or by viruses. Until quite recently any transmissible disease with which no microorganism or other agent could be associated was automatically assumed to be caused by a virus; if the virus could not be isolated, it was assumed to be simply too elusive to be detected by the available techniques. It was the investigation of a plant disease that contravened this generalization too and led to the first isolation of a viroid.

The spindle-tuber disease of potatoes, which gives rise to gnarled, elongated tubers that are sometimes dissected by deep cracks in the surface, was first noted in the northeastern U.S. in the 1920's. Plant pathologists soon established that the disease was transmissible and that no bacteria or other microorganisms were consistently associated with it, and so it was assumed to be caused by a virus. Over the years several attempts were made to isolate the putative virus, always without success. That did not seem surprising; many viruses are quite hard to isolate and purify.

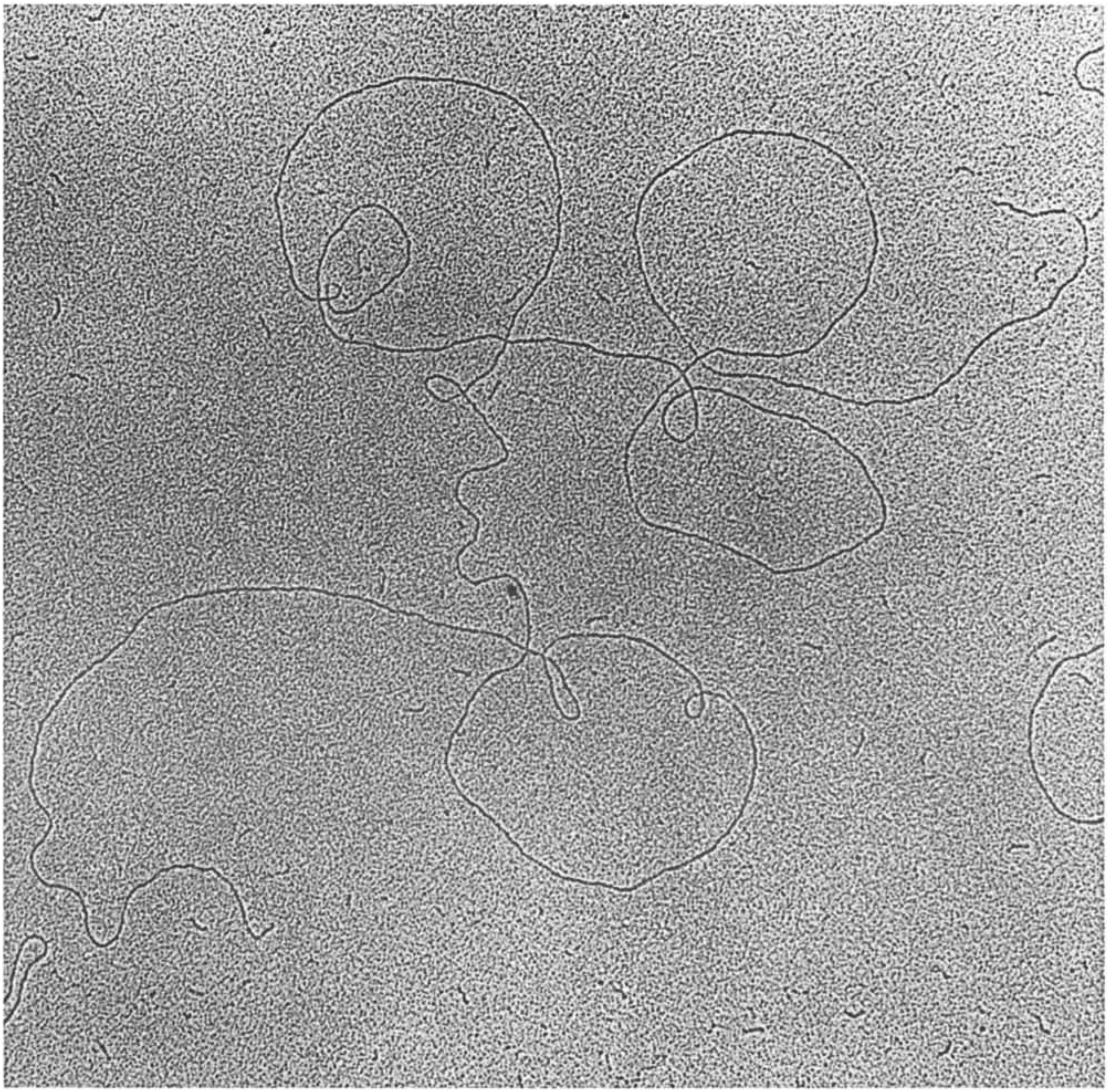
In 1962 William B. Raymer and Muriel J. O'Brien, working at the Agricultural Research Center of the U.S. Department of Agriculture in Beltsville, Md., found that the spindle-tuber agent could be transmitted from potato plants to young tomato plants, in which it multiplied and caused characteristic symptoms (stunt and twisted leaves) within about two weeks—much faster than in potatoes, where the disease usually becomes manifest only after the tubers have developed. That observation made it possible to do systematic experiments far more expeditiously. Raymer soon found he could prepare highly infectious extracts by grinding infected tomato leaves in a phosphate solution. It seemed that purification of the virus should not present any great difficulty; one or two cycles of differential centrifugation should do the trick.

In that procedure an infectious extract is exposed alternately to moderate and very high centrifugal forces by spinning in a centrifuge. Centrifugation at about 10,000 times gravity tends to sediment impurities, whereas any virus particles remain suspended in the supernatant liquid. Exposing the supernatant to a force of 100,000 *g* or more in an ultracentrifuge sediments most virus particles into a pellet. The pellet is resuspended; the suspension is again centrifuged to remove more impurities and again ultracentrifuged to pelletize the virus. When Raymer subjected his infectious tomato-leaf extracts to this procedure, he found that most of the infectious material remained in the supernatant liquid even after exposure to 100,000 *g* for four hours. That is, the supernatant remained highly infectious and the material in the pellet was only slightly so. There was something unusual about such a result, and when Raymer brought it to my attention, we decided to investigate the problem together.

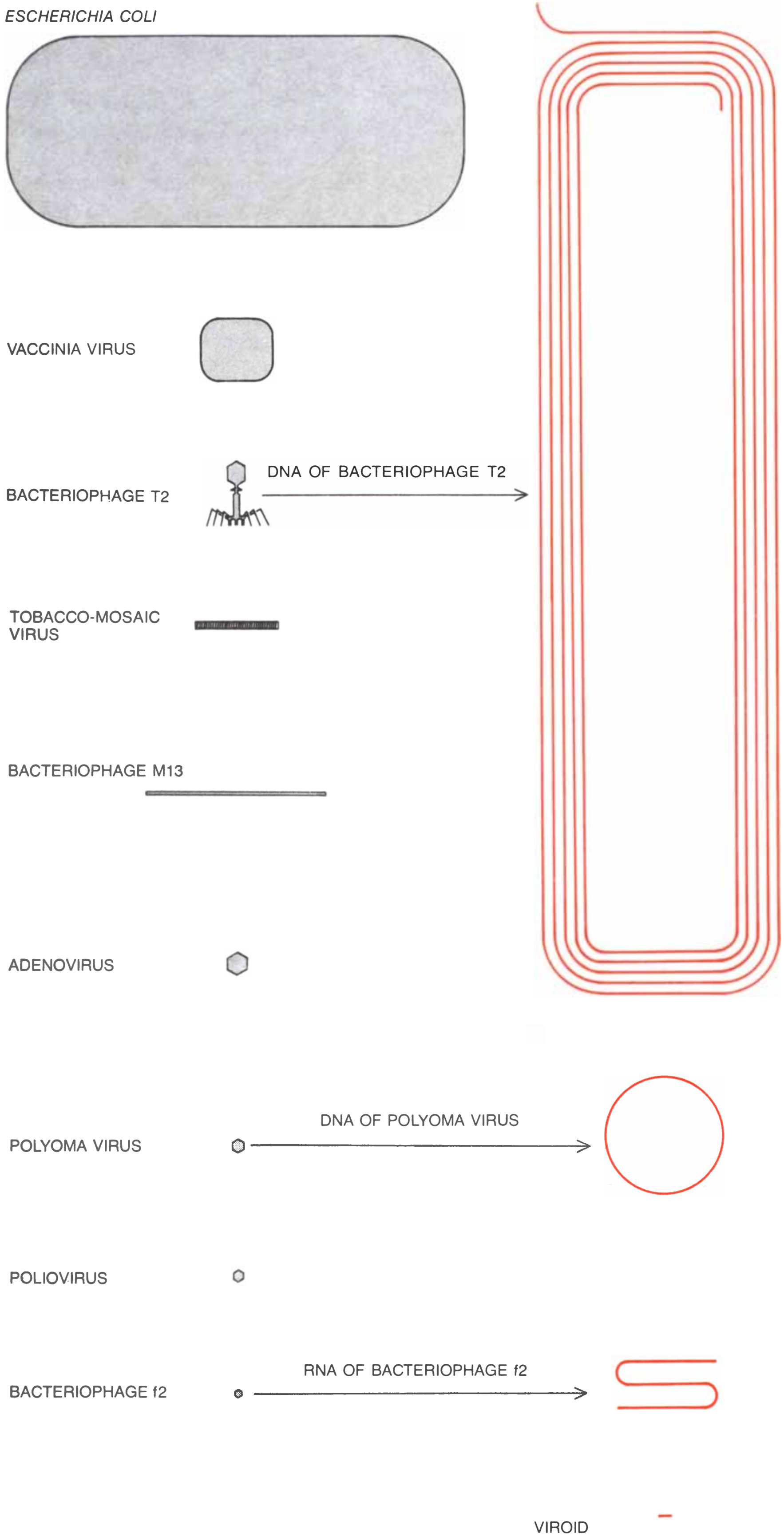
One possible explanation—that the agent was a virus containing so much lipid (fat) that its density was particularly low—had already been ruled out: lipid solvents failed to inactivate the agent. The only other plausible explanation was that the infectious agent was remarkably small. To get a better estimate of its sedimentation properties we turned to centrifugation in a sucrose density gradient, a powerful technique developed by Myron K. Brakke of the University of Nebraska. We were surprised to find that the infectious agent sedimented not only at a lower rate than most virus particles we tested but also more slowly than the nucleic acid component of such particles. We had no choice but to suspect that the agent might consist of nothing but DNA or RNA. Treating the infectious material with the RNA-digesting enzyme ribonuclease inactivated the agent. Treatment with enzymes that break down DNA or proteins, on the other hand, had no effect on either the infectivity or the sedimentation behavior of the agent. This meant the essential element in infection had to be RNA and protein was probably not involved. Many other experiments confirmed these conclusions and convinced us that in living plants as well as in the laboratory the infectious RNA was not encapsulated, as it is in viruses, in a protein coat. In 1967 we proposed that the disease agent must be a free RNA.

Just how small was the infectious RNA molecule? It was hard to tell because the very low concentration of the RNA in infected tissue meant it could be detected only indirectly, by observing its biological activity: its ability to give rise to the characteristic symptoms when it was rubbed on the leaves of tomato plants. Even knowing the sedimentation rate of the RNA was not enough. Nucleic acid molecules of the same size can sediment at quite different rates depending on whether they are single-strand or double-strand molecules and whether they have a compact or a loose conformation. We were far from knowing such structural characteristics of the spindle-tuber agent.

It was only by combining data from the density-gradient sedimentation and from gel electrophoresis that I was able, with the help of Dennis R. Smith, to estimate the size of the spindle-tuber RNA. Gel electrophoresis exploits the fact that nucleic acid molecules, which are negatively charged, move toward the positive pole of an electric field. The preparation to be analyzed is placed in a well at one end of a polyacrylamide gel. When an electric current is applied, the various molecules are separated by size as they migrate through the gel at a rate



POTATO SPINDLE-TUBER VIROIDS are the scattered, short rods, each about 3.5 millimeters long, visible in this electron micrograph made by José M. Sogo and Theo Koller of the Swiss Federal Institute of Technology. They are short single strands of RNA, but in their "native" form, as here, they have a collapsed-circle or hairpin conformation and hence appear to be double-strand molecules comparable in width to the long double-strand DNA of the bacteriophage (bacterial virus) T7, included to provide a size standard. The T7 DNA is about 14 micrometers (thousandths of a millimeter) long; the native viroid is only about 50 nanometers (.05 micrometer) long.



SMALL SIZE OF A VIROID can be appreciated when the bacterium *Escherichia coli*, a number of viruses and the nucleic acid genomes of some of the viruses are compared with the native form of the potato spindle-tuber viroid (PSTV), all of them enlarged some 40,000 diameters in these drawings. The RNA of the bacterial virus f2, one of the smallest genomes that is large enough to direct independent replication of its virus, is much longer than viroid RNA.

inversely proportional to the logarithm of their molecular weight; eventually they form a series of bands each of which represents a collection of molecules of the same size. The bands can ordinarily be visualized by staining, and then the molecular weight of the material in each band is estimated by measuring its distance from the origin or from bands of identifiable molecules whose size is known. The small amount of spindle-tuber agent in our samples was not visible on staining, and so we depended again on its biological activity to determine its position in the gels. We cut each gel into thin slices and tested each slice for infectivity by rubbing it on the leaves of tomato plants.

The result was unequivocal: the infectious RNA was very small indeed. Our first estimate was that its molecular weight was about 50,000; more refined measurements eventually gave a value of about 130,000. This finding immediately raised a question that has yet to be answered: How could such a small RNA be a viable infectious agent? It is a well-established generalization in virology that most viruses have a genome, or total content of DNA or RNA, with a molecular weight of at least about a million. That much nucleic acid seems to be required for a virus to take over the genetic machinery of a host cell and subvert it to the cause of virus proliferation, and also to code for the necessary virus-specific proteins. A virus with less genetic information is termed "defective"; it cannot multiply on its own and must depend on genetic information provided by another virus—a "helper" virus—in the same cell.

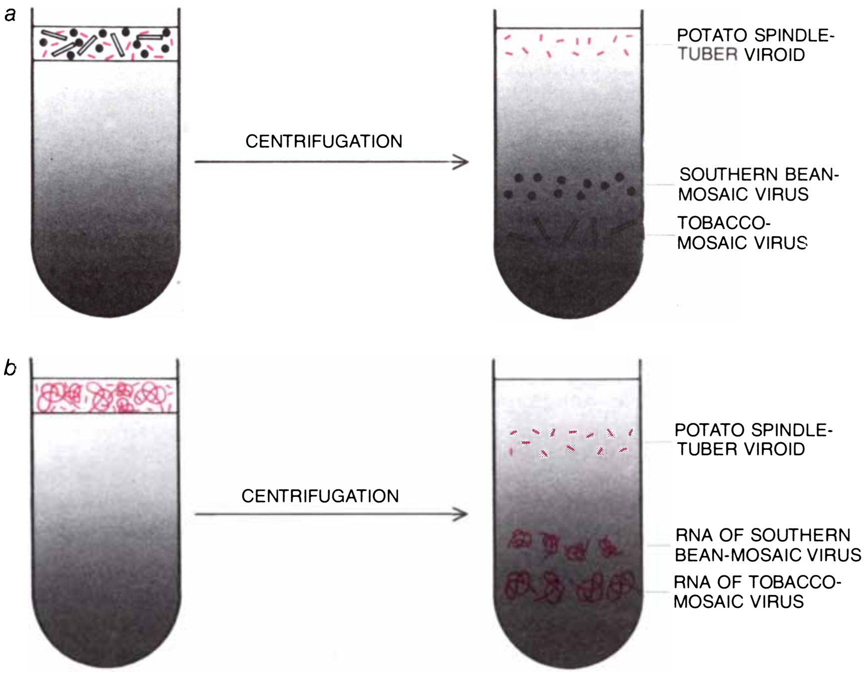
We recognized that the potato spindle-tuber RNA did not have to code for the coat proteins characteristic of a virus, so that it could presumably make do with less genetic information than a typical virus can. A genome of the size we had estimated was only large enough, however, to code for protein with a total molecular weight of about 10,000, which is less than the size of a very small enzyme. The spindle-tuber RNA would have to depend largely, if not completely, on preexisting host enzymes for the synthesis of its progeny. This notion was difficult to accept, however, because plant cells were not thought to have any enzymes that could synthesize new RNA with the spindle-tuber RNA as a template. Such RNA-directed RNA polymerases had been identified only in cells infected by viruses and were apparently encoded by viral genes.

It seemed plausible, then, to regard the infectious RNA as a coatless defective virus that multiplied in plant cells with the aid of a helper virus. We therefore searched, in uninfected tomato plants, for any virus that might provide the necessary supplemental genetic information. All such efforts were fruit-less. I finally became convinced that the infectious RNA was somehow able, in spite of its small size, to multiply independently of any helper virus. It was just possible, however, that the RNA we had detected represented not a single molecular species but a population of different RNA molecules of about the same size, which could assemble to provide a more typical genome. We investigated this possibility by testing the effect of dilution and of ultraviolet irradiation on the infectivity of our RNA preparations; in both cases the slope of the infectivity curves seemed to rule out any possibility that several different molecules were involved. Evidently the agent of potato spindle-tuber disease differed drastically from viruses and was the first representative of a newly recognized class of subviral pathogens. In 1971 I proposed that such agents be called viroids.

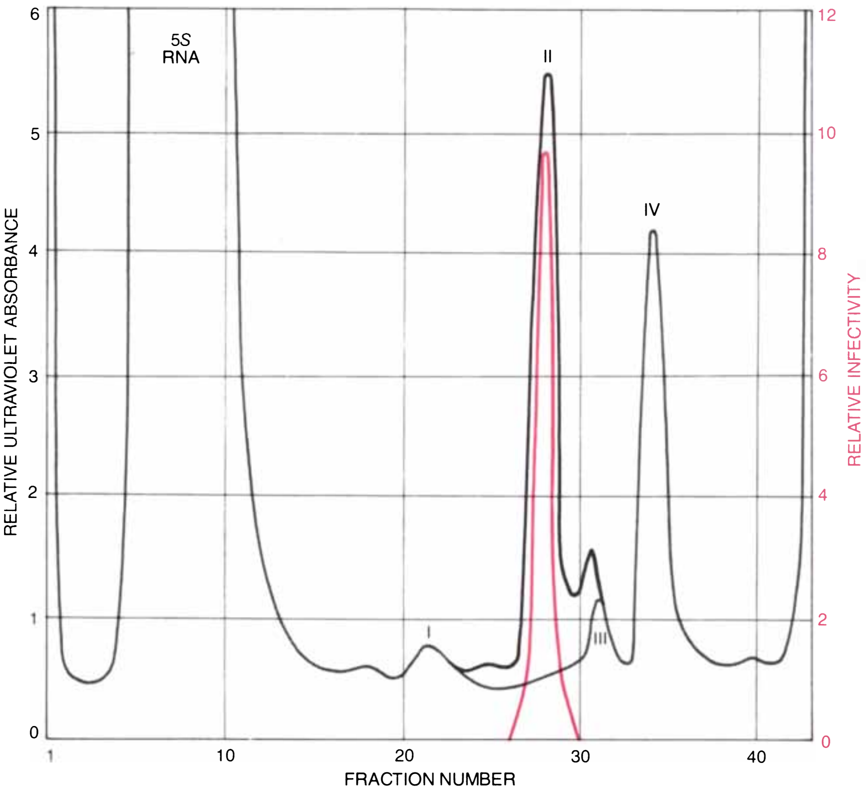
By 1972 Joseph S. Semancik and Lewis G. Weathers of the University of California at Riverside showed that the agent of exocortis, an infectious disease of citrus trees previously assumed to be caused by a virus, was in fact a viroid. Then Roger H. Lawson of the Beltsville center found that the agent of a serious disease called chrysanthemum stunt had properties similar to those of the potato spindle-tuber viroid (PSTV). Collaboration between Lawson and me established that the agent is indeed a viroid.

At least five other plant diseases now appear to be caused by viroids. H. J. M. van Dorst and Dirk Peters of the Agricultural University at Wageningen in the Netherlands found that a viroid is responsible for cucumber pale-fruit disease. Charles P. Romaine and R. Kenneth Horst of Cornell University identified the agent of another chrysanthemum disease, chlorotic mottle, as a viroid; Matsuo Sasaki and Eishiro Shikata of Hokkaido University did the same for the agent of the stunt disease of hops. John W. Randles of the University of Adelaide found evidence that viroidlike RNA's are associated with cadang-cadang, a disease that has killed millions of coconut trees and caused large economic losses in the Philippines. Evidence that avocado sun-blotch disease may be caused by a viroid has been presented recently by N. A. Mohamed and Wayne Thomas of the New Zealand Department of Scientific and Industrial Research and by Peter Palukaitis and Robert H. Symons and their colleagues at the University of Adelaide. Undoubtedly viroids will be implicated as agents of still more plant diseases whose cause is not yet known.

Before the physical-chemical properties of viroids could be examined it was necessary to separate them from the nucleic acids of the cells they infect and to purify them. The low concentration in plant tissues of the viroid RNA com-



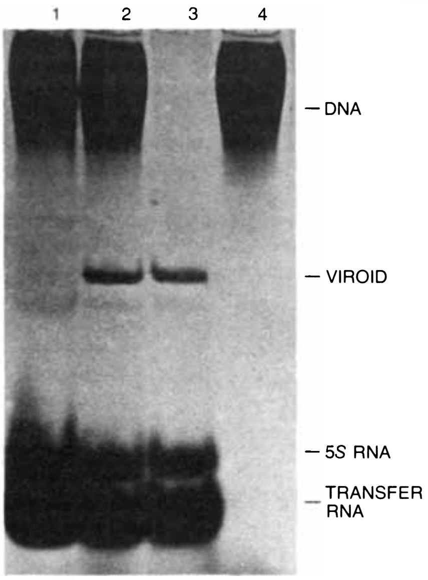
CENTRIFUGATION in a sucrose density gradient showed the agent of potato spindle-tuber disease had to be a small molecule of naked nucleic acid. In the centrifuge tube (*a*) a mixture of the viroid and two viruses is layered on a solution in which the sucrose concentration increases from the top to the bottom. Centrifugation for three hours at about 60,000 *g* separates the constituents by size into three bands, which are identified by infectivity assays; the PSTV remains at the top of the tube. When the viroid is centrifuged for 16 hours with the RNA of the two viruses (*b*), the viral RNA's move well down in the tube but the smaller viroid stays near the top.



SMALL RNA'S from the leaves of healthy and of infected plants were separated according to size by gel electrophoresis and were then detected by the extent to which they absorbed ultraviolet radiation. Preparations from both healthy and infected leaves showed absorbance peaks for a ribosomal RNA (*5S RNA*) and three cellular RNA's (*peaks I, III and IV*). The infected preparation also showed a unique fifth peak (*II*). Absence of peak II from healthy plants and its coincidence with the infectivity peak (*color*) indicated that it must represent the viroid RNA.

pared with host RNA made the separation difficult, and conventional purification methods applicable to protein-coated virus particles were of no help. By adopting improved methods of separating and purifying RNA and by working with large amounts of infected tissue we were eventually able to isolate viroids. We fractionated the small RNA's in extracts of healthy tomato leaves and in leaves infected by PSTV on electrophoresis gels and then measured the absorption of ultraviolet radiation by each RNA fraction. A prominent absorbance peak was present in the profile for infected leaves that was not present in the profile for healthy leaves; when the same fractions were tested for their ability to infect healthy plants, the peak of the infectivity distribution was found to coincide precisely with the unique absorbance peak.

Taken together, these observations represented the first recognition of a viroid as a physical entity. As we succeeded in increasing the concentration of the viroid RNA it became possible to make the viroid fraction visible. It forms a distinct band on a gel treated with a dye that stains nucleic acids; the band is present only in infective extracts and is eliminated by the enzyme ribonuclease. To obtain PSTV essentially free of contaminating nucleic acids and in amounts sufficient for biophysical and biochemical analysis we proceed to cut the PSTV-containing slice out of a large number of gels, to extract the RNA with a solvent, to reconcentrate the RNA and to subject it to an additional electrophoresis cycle or two.



VIROID NUCLEIC ACID was eventually visualized and identified unequivocally as RNA by gel electrophoresis. Tomato-plant nucleic acids were placed in troughs at the top of a gel. Under the influence of an electric potential they migrated toward the positive pole (*bottom*) at a rate inversely proportional to the logarithm of their molecular weight and hence were separated into bands made visible by staining. Track No. 1 shows nucleic acids from healthy plants; nucleic acids from infected leaves were run in tracks Nos. 2, 3 and 4. The viroid is not seen in track No. 1. It is visible in track No. 2. It is not affected by the enzyme deoxyribonuclease, which digests the DNA (*track No. 3*), but it is itself digested, along with the other RNA, by the enzyme ribonuclease (*track No. 4*).

Once purified viroid preparations became available it became possible to investigate the infectious molecule's structure. To begin with, was it a single-strand RNA or a double-strand one? Both forms are found in viruses in which RNA constitutes the genome; cellular RNA's, which play various roles in the translation of genetic information from the cell's DNA genome into proteins, are almost invariably single-strand. Some early experiments (based, for example, on the rate at which a PSTV preparation's infectivity was reduced by treatment with ribonuclease) had suggested the viroid was a double-strand RNA, but contradictory results had been obtained by different analytical methods. If purified viroids could be seen clearly in an electron micrograph, it was hoped, then simply measuring their width should resolve the ambiguity. The first successful micrographs, made from our preparations by José M. Sogo and Theo Koller, working at the Swiss Federal Institute of Technology in Zurich, showed a uniform population of rods with an average length of 50 nanometers (millionths of a millimeter), in good agreement with our earlier size estimate. The width of the viroids seemed to be similar to that of double-strand viral DNA seen in the same micrograph, suggesting the viroid too was double-strand. This turned out not to be the case, as we found by determining the denaturation properties of the viroid.

To briefly review the nature of DNA and RNA, a strand of nucleic acid is a chain composed of four subunits called nucleotides. Each nucleotide is characterized by a projecting chemical group called a base. In DNA the bases are adenine (*A*), guanine (*G*), cytosine (*C*) and thymine (*T*); in RNA the first three bases are the same but thymine is replaced by uracil (*U*). The bases are complementary, so that in RNA *A* pairs with *U* and *G* pairs with *C* and sometimes with *U*. It is the pairing of complementary bases by hydrogen bonding that links two single strands to make a double-strand molecule. Base pairing can also take place in a single strand: one region of a strand can fold back on a complementary region, forming a hairpinlike loop. Regions linked by base pairing are "denatured," or separated, when a nucleic acid molecule is heated to break the hydrogen bonds between complementary bases, and the rate and the range of temperatures at which denaturation takes place vary with the structure of the molecule. The extent of thermal denaturation is most conveniently determined by measuring a nucleic acid preparation's ultraviolet absorbance, which increases with denaturation.

There is a characteristic curve relating ultraviolet absorbance to temperature for single-strand RNA and a different characteristic curve for double-strand RNA. The curve we derived for viroid RNA was not at all like the one for double-strand RNA; it was more like the curve for single-strand RNA but significantly steeper. We concluded that in their "native," or undenatured, state viroids are single-strand molecules folded into a hairpinlike configuration, with extensive regions of intrastrand base pairing. This explained why viroids in electron micrographs had appeared to be double-strand molecules, and it also explained the contradictory results of the earlier experiments. A viroid is a single-strand RNA molecule with such extensive intrastrand pairing that it displays some of the characteristics of a double-strand RNA.

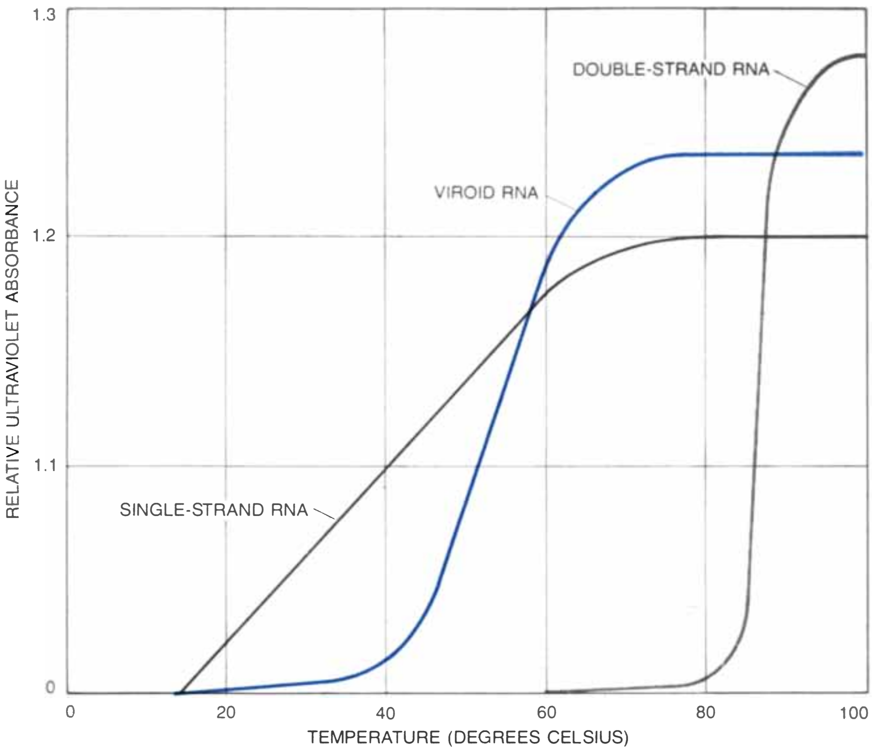
Soon two groups of investigators succeeded in making electron micrographs of completely denatured viroid molecules. William L. McClements and Paul J. Kaesberg of the University of Wisconsin reported that their PSTV micrographs show a mixture of two kinds of molecules: a majority of threadlike linear molecules roughly twice as long as the native rods and a minority of circular molecules whose circumference is about the same as the length of the linear ones. On the other hand, a group headed by Heinz L. Sänger of the University of Giessen in West Germany reported that viroids are all circular; Sänger and his colleagues think the rare linear molecules in their micrographs are artifacts of the purification process. The two groups agree, however, that the circular molecules are closed loops formed by covalent bonds of the nucleotide chain, not linear molecules whose two ends are linked simply by intrastrand base pairing between complementary regions at the two ends.

We undertook to find out if both the linear and the circular forms are present and are biologically significant in living plants. Robert A. Owens of my laboratory managed to separate denatured linear molecules from circular ones by subjecting purified, denatured PSTV to gel electrophoresis under conditions that prevented renaturation; the linear molecules moved faster than the circular ones, so that the two forms ended up in two discrete bands. In collaboration with Russell L. Steere and Eric Erbe we were able to show that both forms were infectious. Then Ahmed Hadidi introduced a radioactive isotope of phosphorus into PSTV-infected tomato plants and monitored the incorporation of the isotope into RNA as the viroid replicated. After a short period of replication most of the radioactively labeled PSTV was in the circular form; as replication and isotope incorporation proceeded, an increasing number of the labeled molecules were found to be linear. This confirmed that both the circular and the linear forms are present in infected plants. It also suggested that the circular molecules may be precursors of the linear ones.

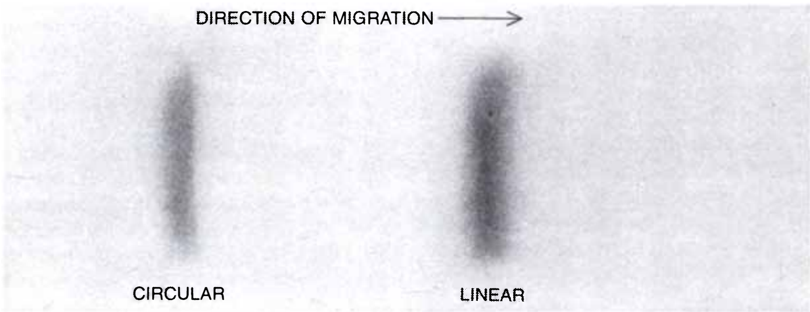
Meanwhile other workers were beginning to determine the chemical composition of viroids, first by the technique known as fingerprinting. An RNA is digested with an enzyme that cleaves the nucleotide chain only at particular sites and the resulting fragments of from one nucleotide to perhaps a dozen nucleotides are separated in two dimensions by electrophoresis and chromatography. The fragments form a two-dimensional pattern of spots that depends on the base sequence (the linear arrangement of the bases *A, G, C* and *U*) and hence is different for different molecules. Before fingerprinting was possible there had been some doubt about whether the viroids causing different diseases were in fact different molecules. Now Elizabeth Dickson and her colleagues at Rockefeller University showed the base sequence of PSTV and that of the citrus exocortis viroid were quite different from each other: the two viroids are distinct RNA species. As other viroids have been analyzed it has become clear that each is a distinct molecule with its own characteristic nucleotide sequence. On the other hand, different strains of a single viroid such as PSTV have sequences that differ from one another at only a few positions along the chain. Such results indicate that viroids are functional genetic systems whose characteristics, like those of other genetic elements, are determined by their nucleotide sequence.

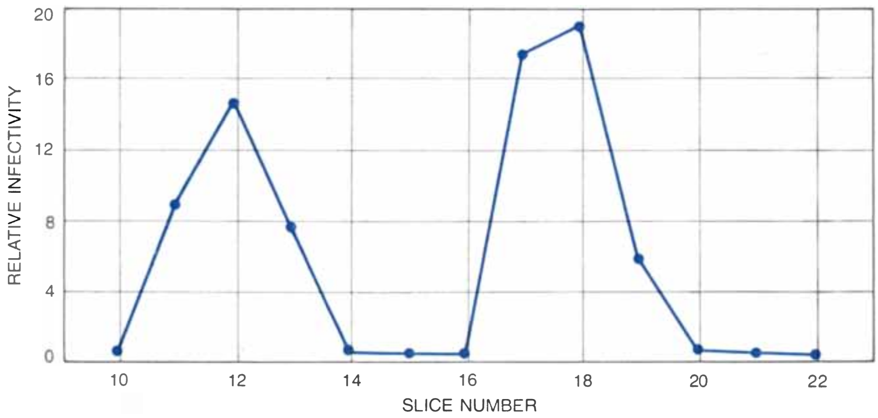
In 1978 the complete primary structure—the entire nucleotide sequence—of PSTV was worked out by a group headed by Hans J. Gross of the Max Planck Institute for Biochemistry in Munich. The viroid is a chain of 359 nucleotides: 73 *A*'s, 77 *U*'s, 101 *G*'s and 108 *C*'s. (The excess of *G*'s and *C*'s seems to be characteristic of viroids in general, judging from estimates of the base composition of several other viroid species.) Gross and his colleagues went on to propose a secondary structure, or conformation, for native PSTV by folding the primary sequence to allow the largest possible extent of base pairing and to account for other properties of the molecule, such as its resistance to cleavage by ribonucleases.

Their model is in good agreement with what had been predicted from

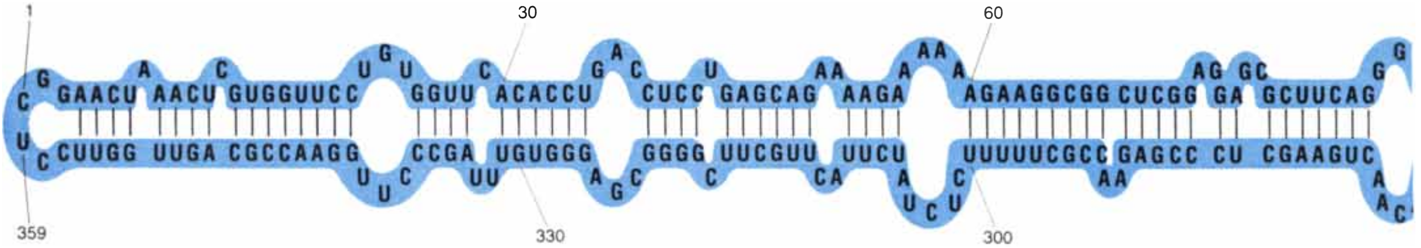


DENATURATION CURVES measure the change in an RNA's absorption of ultraviolet radiation as its base-paired regions separate when the temperature is increased. Here the characteristic curves for single-strand and double-strand RNA are compared with curve for viroid RNA. Viroid is shown to be not double-strand RNA but rather an unusual single-strand RNA.





INFECTIVITY of circular and linear viroid molecules was established by denaturing PSTV and subjecting two samples to electrophoresis under conditions preventing renaturation. One gel was stained, revealing two PSTV bands (*top*); electron microscopy showed circular PSTV in one band, linear PSTV in the other. Corresponding slices in the other gel were ground up and tested for infectivity. Both circular and linear forms were shown to be infective (*bottom*).



STRUCTURE OF PSTV was proposed by Hans J. Gross and his colleagues at the Max Planck Institute for Biochemistry in Munich, who worked out the 359-nucleotide sequence of the viroid and arranged the sequence to maximize base pairing. The vertical lines rep-

physical observations: a closed single strand of RNA in which short double-strand regions (actually regions of intrastrand base pairing) alternate with still shorter single-strand regions (mismatched loops of unpaired bases). This is a novel and perhaps unique structure. Closed, circular, single-strand RNA molecules have not previously been observed in nature. The high degree of intrastrand base pairing, which gives rise to the collapsed-circle or hairpin structure, is also most unusual. It is hard to believe these structural peculiarities do not have some connection with the biological activity of viroids.

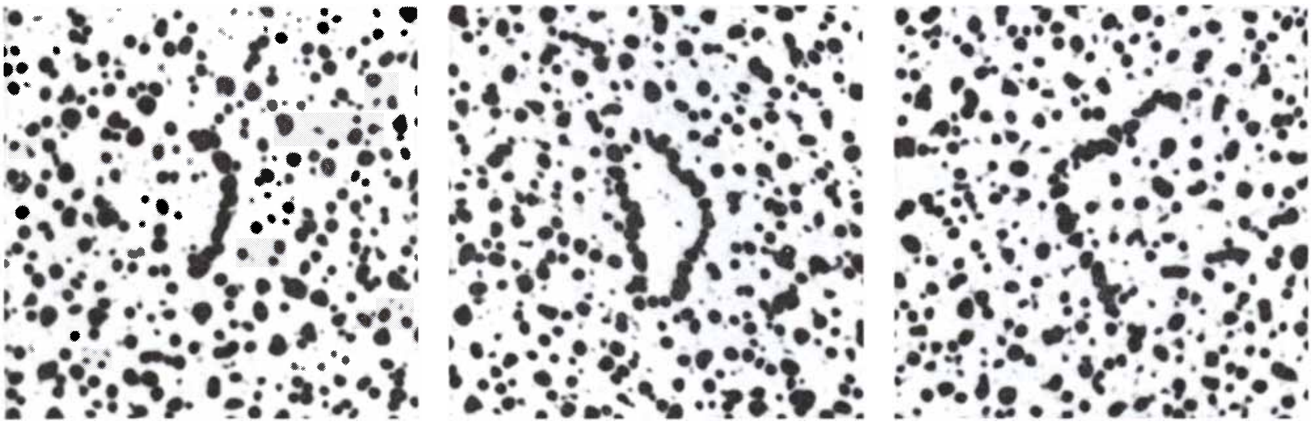
The connection is not at all clear, however. Knowledge at the molecular level of the biological properties of viroids, and in particular how they replicate in a host cell and how they cause disease, is very limited. What is certain is that when a viroid is introduced into a host cell, it replicates without the assistance of a helper virus.

How does such a small RNA molecule induce its own synthesis in an infected cell? Might the viroid RNA encode the manufacture of a polypeptide (a short protein chain) that is essential for replication? Or is the viroid replicated solely by enzymes already present in a healthy plant? There is general agreement that viroids themselves are not translated into protein, that is, they do not act as messenger RNA's encoding specific proteins. In laboratory preparations designed for the translation of RNA into protein they not only are inactive but also do not interfere with the translation of genuine messenger RNA's in the same preparation. Moreover, in plants infected with PSTV or the exocortis viroid no proteins can be found that are not also present in healthy plants. The synthesis of certain proteins is enhanced in infected plants, but Vicente Conejero and his colleagues at the Polytechnic University of Valencia have shown that those proteins are host proteins, not viroid-specific ones (at least in the case of plants infected by the exocortis viroid).

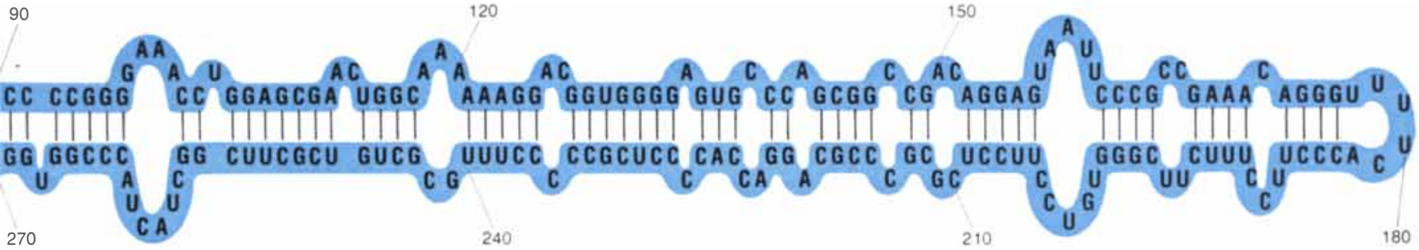
If viroids are not translated, they cannot encode enzymes. Presumably, then, their replication must rely entirely on enzyme systems of the host plant. Contrary to earlier assumptions, there are indeed enzymes in uninfected cells of a number of plant species, including the tomato, that can replicate RNA on an RNA template. These enzymes are obvious candidates for mediating the synthesis of new viroid RNA, but their involvement has yet to be demonstrated.

Whatever the source of the replicating enzymes, they can assemble new viroid strands only along a template: a nucleic acid having a sequence of bases complementary to that of the viroid. Is the template an RNA intermediary transcribed from the viroid or is it a DNA transcribed from the viroid by the enzyme called RNA-directed DNA polymerase? One ought to be able to find evidence for the presence of one or the other kind of template in infected cells. RNA sequences complementary to part of the citrus exocortis viroid can be found in infected cells but not in healthy ones, as was first shown by Larry K. Grill and Semancik. So far, however, it has not been shown that these complementary sequences represent the complete viroid, which presumably would be necessary to provide viable templates. On the other hand, we have identified some sequences complementary to PSTV in DNA from both healthy and PSTV-infected plants, but so far our findings have not been confirmed by other workers. In any case there has been no demonstration that such DNA sequences are involved in viroid replication.

This line of research depends primarily on hybridization experiments. Viroid RNA (or DNA with the same nucleotide sequence as the viroid), labeled so that it can be recognized, is mixed with an extract from infected plant tissue. Under conditions that promote base pairing the labeled nucleic acid probes "find" any complementary strands present in the plant extract and anneal to them to form



DENATURED VIROID MOLECULES seen in electron micrographs made by William L. McClements and Paul J. Kaesberg of the University of Wisconsin are enlarged here some 440,000 diameters. Native PSTV was denatured by formaldehyde treatment and prepared for microscopy under conditions largely preventing the reassociation of base pairs. Some PSTV molecules nonetheless reassumed their native structure (*left*). Of the denatured molecules, some were circular (*middle*), but most denatured molecules were linear (*right*).



resent hydrogen bonds between permissible base pairs: *A–U, G–C* and *G–U;* the colored band represents the covalently linked backbone of the molecule. Short base-paired regions alternate with shorter unpaired regions, resulting in structure resembling double-strand RNA.

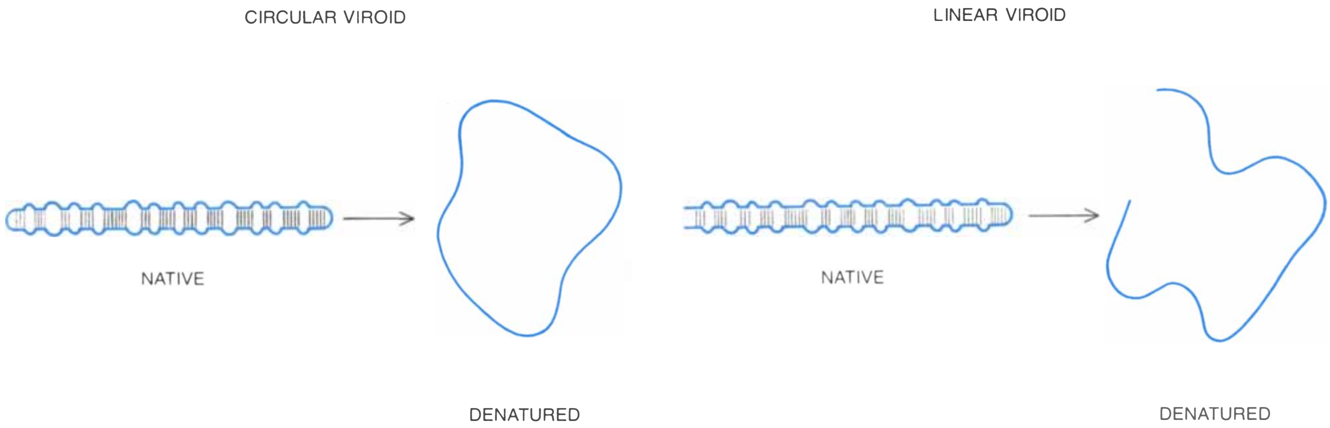
hybrid double-strand molecules, which can be recognized by their labeling. In most of the investigations reported so far the probe has been a viroid isolated from an infected plant, purified to the extent possible and labeled. With such probes it is hard to exclude the possibility that any hybridization one observes may result from the base pairing of a host RNA or DNA with some complementary plant RNA that still contaminates the viroid probe, rather than with the viroid itself.

Recently several investigators have been able to prepare probes by synthesizing strands of DNA that are complementary to a viroid. In my laboratory Owens and Dean E. Cress synthesized DNA complementary to almost the entire sequence of PSTV, made a second DNA strand complementary in turn to the first one, spliced the double-strand DNA into a plasmid (a small circle of bacterial DNA) and introduced the plasmid into the bacterium *Escherichia coli,* where it proliferated. This molecular cloning technique provides large quantities of an unambiguous probe, with which we hope soon to clarify the molecular mechanisms of viroid replication.

There is another fundamental question that has yet to be answered: How do viroids give rise to symptoms? And why do they cause disease only in certain susceptible plants, whereas they multiply in other plants without doing any discernible harm? Many of the symptoms are disturbances of growth and may therefore be the result of imbalanced growth-hormone activity. How viroids might bring that about is still an enigma. Viroids are found predominantly in the nuclei of infected cells, where genes are regulated—turned on or off. It is therefore at least possible that viroids function as abnormal regulatory molecules, somehow interfering with the control of genes encoding particular hormones.

So far viroids have been clearly identified only in higher plants, but they (or very similar agents) may function in other forms of life as well. It seems reasonable to search for viroids wherever an infectious disease has been assumed to be caused by a virus but no virus particles have yet been identified. That is the case for a group of brain diseases in animals known as the subacute spongiform encephalopathies. They include kuru and Creutzfeldt-Jakob disease in man, scrapie in sheep and goats and transmissible encephalopathy in mink. All four have been shown to be infectious nonbacterial diseases. They are commonly blamed on what are called "slow" viruses. Yet in spite of intensive efforts by many investigators no virus has been identified with any of them. Indeed, the causative agents have been shown to have properties that are either unknown or very rare in conventional viruses: great heat stability, extreme resistance to ultraviolet and ionizing radiations and unusual resistance to certain chemicals.

On the basis of comparisons between the properties of PSTV and the agent of scrapie I proposed in 1972 that the latter too may be a viroid. So far no clearly infectious nucleic acid has been isolated from the brain of animals with scrapie. Gordon D. Hunter and Richard H. Kimberlin and their colleagues at the Agricultural Research Council Institute for Research on Animal Diseases in England have, however, reported evidence for the presence of a small DNA molecule in infected tissue that is not present in the brain of healthy animals. Recently Richard F. Marsh and his colleagues at the University of Wisconsin have reported evidence that a component essential for scrapie infection can be inactivated with deoxyribonuclease, suggesting that the scrapie agent may be a coatless DNA molecule. Their findings have yet to be confirmed, and further characterization of the DNA will be required. It begins to appear, however, that viroidlike nucleic acids may be responsible for infectious diseases not only in plants but also in animals, perhaps including man.



CLOSE SIMILARITY of the collapsed-circle structure of native circular molecules (*left*) and the hairpin structure of native linear ones (*right*) makes them impossible to distinguish by electrophoresis or in electron micrographs. When they are denatured, however, the difference is apparent, as is shown in these schematic drawings. Denatured circular molecules can be distinguished from denatured linear ones in electrophoresis gels (*see bottom illustration on page 71*) and electron micrographs (*see bottom illustration on opposite page*).

The Total Artificial Heart

Mechanical substitutes for the natural heart are steadily getting better. In time they will be ready for human beings needing them, a development with social implications as well as medical ones

by Robert K. Jarvik

Heart disease continues to take a heavy toll of human life, being responsible for nearly a million deaths a year in the U.S. alone. The toll has been reduced somewhat by improved medical procedures, new drugs and pacemakers, and it could be reduced far more if individuals assumed greater responsibility for controlling such proved risk factors as smoking, inadequate exercise, high blood pressure, obesity and excessive stress. In this arsenal of therapy and prevention an additional weapon may soon be available: the total artificial heart.

The prospects for the total artificial heart are much better than they seemed to be a decade ago. At that time no animal with an implanted artificial heart had survived for more than three days. Today the record stands at 221 days, or more than seven months. The record is held by the Jarvik-7 heart, one in a series of artificial hearts I have designed at the University of Utah College of Medicine. Neither the Jarvik-7 nor any of the several other total artificial hearts being developed is yet ready to permanently replace a human heart, even on a trial basis, but the pace of improvement in the technology suggests that the day may not be long in coming.

The artificial heart has the potential of reaching a much larger population than heart transplantation. In 1978 only 31 people received transplanted hearts, whereas a panel of the National Heart, Lung, and Blood Institute has estimated that from 17,000 to 50,000 people per year could be given artificial hearts. The main problems with heart transplantation are that too few donors are available and that the tendency of the recipient's immune system to reject a transplant can be overcome only with difficulty. Moreover, the drugs employed to suppress the immune response increase the patient's risk of death from infectious disease. Artificial hearts could be made in large numbers, and so any level of demand could be met. Because they would be made of biologically inert polymers and metals they would also be free of the problems that are often associated with immunological rejection.

The extent to which the artificial heart might prolong life is suggested by the record of experience with transplanted hearts. Of people with transplanted hearts 65 percent survive at least one year and about half live five years or more. Of the latter group 80 percent lead essentially normal lives and half are able to return to work. Some recipients alive today have survived for more than 10 years.

The concept of replacing the function of the heart with a mechanical device is far from new. Initially the mechanical heart was envisioned as a substitute for only a part of the circulation: the supplying of blood to certain organs and tissues. In 1812 Julien-Jean-César La Gallois noted that "if one could substitute for the heart a kind of injection [of arterial blood], one would succeed easily in maintaining alive indefinitely any part of the body."

Many experiments with mechanical perfusion were done in the 19th century. They included the perfusion of muscle in 1828 and the perfusion of the isolated heart and lung by Henry Martin in 1880. (Martin's work, which became known as the heart-lung preparation, made an important contribution to the understanding of the basic physiology of the heart.) In 1855 two workers in Germany developed an artificial lung. By 1920 several other kinds of oxygenator had been developed and the concept of the heart-lung machine had emerged. More than 30 such machines had been devised by 1951. One that was widely publicized was assembled at the Rockefeller Institute for Medical Research in the early 1930's by Charles Lindbergh and Alexis Carrel. Their perfusion pump drew the attention of journalists, who gave it such names as "robot heart" and "glass heart." Lindbergh and Carrel suggested an intriguing scheme that has yet to be carried out. "We can perhaps dream," they wrote, "of removing diseased organs from the body and placing them in the Lindbergh pump as the patients are placed in a hospital. There [the organs] could be treated far more energetically than within the organism and, if cured, replanted in the patient."

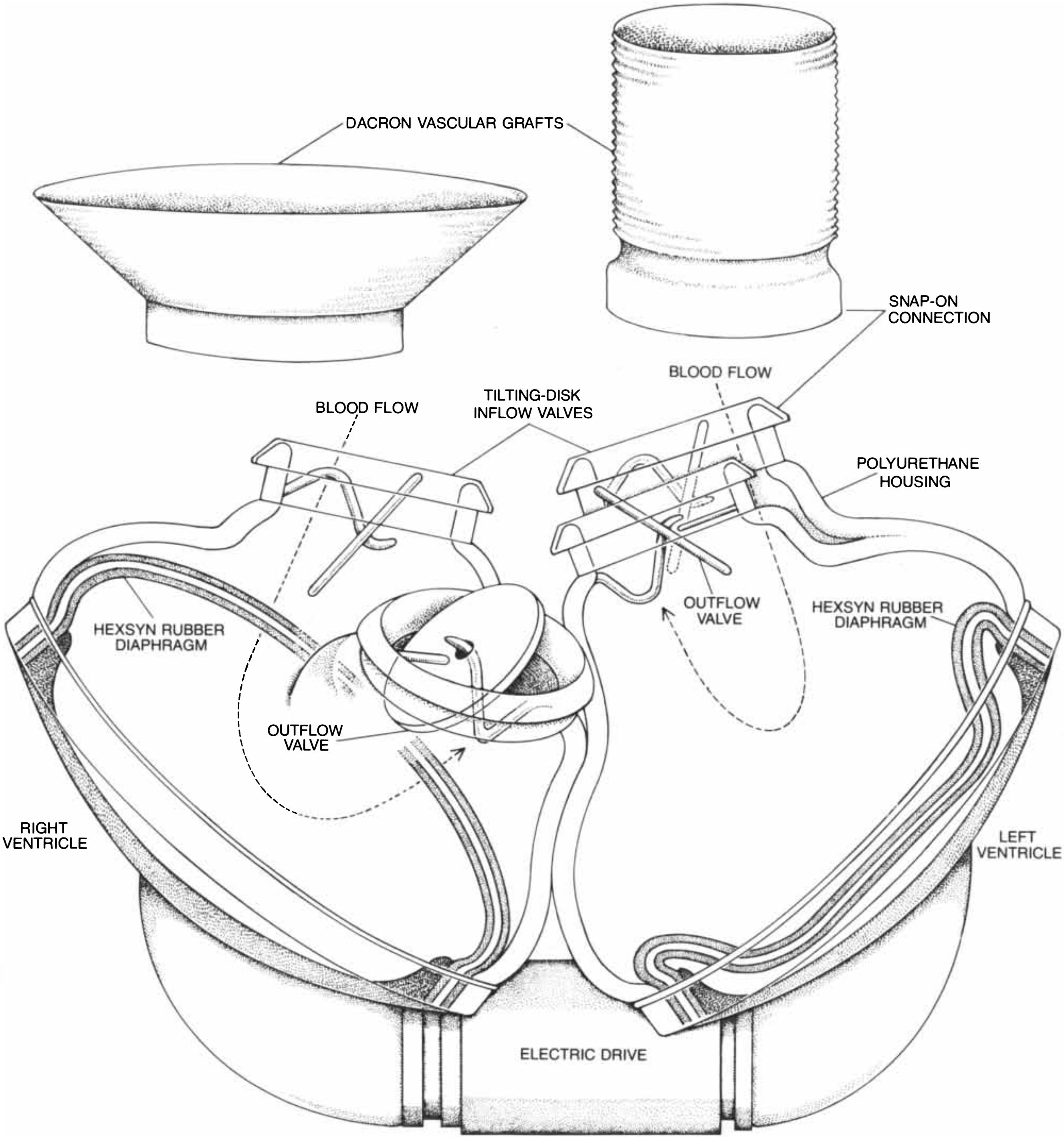
Another early device, which was built by H. H. Dale and E. H. J. Schuster in England in 1928, was a pump intended to replace the function of both the right and the left sides of the natural heart and to provide "complete circulation in the heartless animal." The aim was to temporarily bypass the heart so that surgical procedures could be done on it. In 1952 a mechanical heart developed by F. D. Dodrill of Research Division of the General Motors Corporation drew considerable notice when it was employed to bypass the heart of a human patient for 50 minutes during surgery. As matters turned out, however, open-heart surgery was made possible not by a mechanical heart, which still requires that the patient's lungs oxygenate the blood, but by the heart-lung machine, which supplants both the heart and the lungs.

Experience with the perfusion of organs and with the bypassing of the heart suggested the further step of totally replacing the heart. Here the goal is to install in the chest a device that can take over the function of the heart for the rest of the patient's life. Experiments with a total artificial heart were done in 1957 at the Cleveland Clinic by Willem J. Kolff and Tetsuzo Akutsu. Experiments by both investigators have continued to the present and have been augmented by work in other laboratories. Today workers in about a dozen laboratories (in Argentina, Austria, China, Czechoslovakia, France, Germany, Italy, Japan, the U.S. and the U.S.S.R.) are developing artificial hearts.

When the first experiments were undertaken, certain functional criteria for the artificial heart were already apparent. The device had to be small enough to fit in the space made available by the removal of the natural heart. It had to provide enough output to support the entire body and it had to vary the output according to the body's needs. It had to be readily sterilized and it had to be very durable. Of particular importance, it had to pump blood gently enough to avoid hemolysis, or the destruction of red blood cells. These criteria and others are best understood with the function of the natural heart in mind.

The natural heart has four chambers: two atriums and two ventricles. It has four valves; two allow the blood to move from the atriums to the ventricles but not the other way and two allow the blood to move from the ventricles into the large arteries that carry the blood to the rest of the body. The cardiac cycle is divided into two periods: diastole, when blood enters the ventricles, and systole, when the ventricles contract and blood is pumped out of the heart. The atriums serve mainly as a reservoir for the blood returning to the heart during systole.

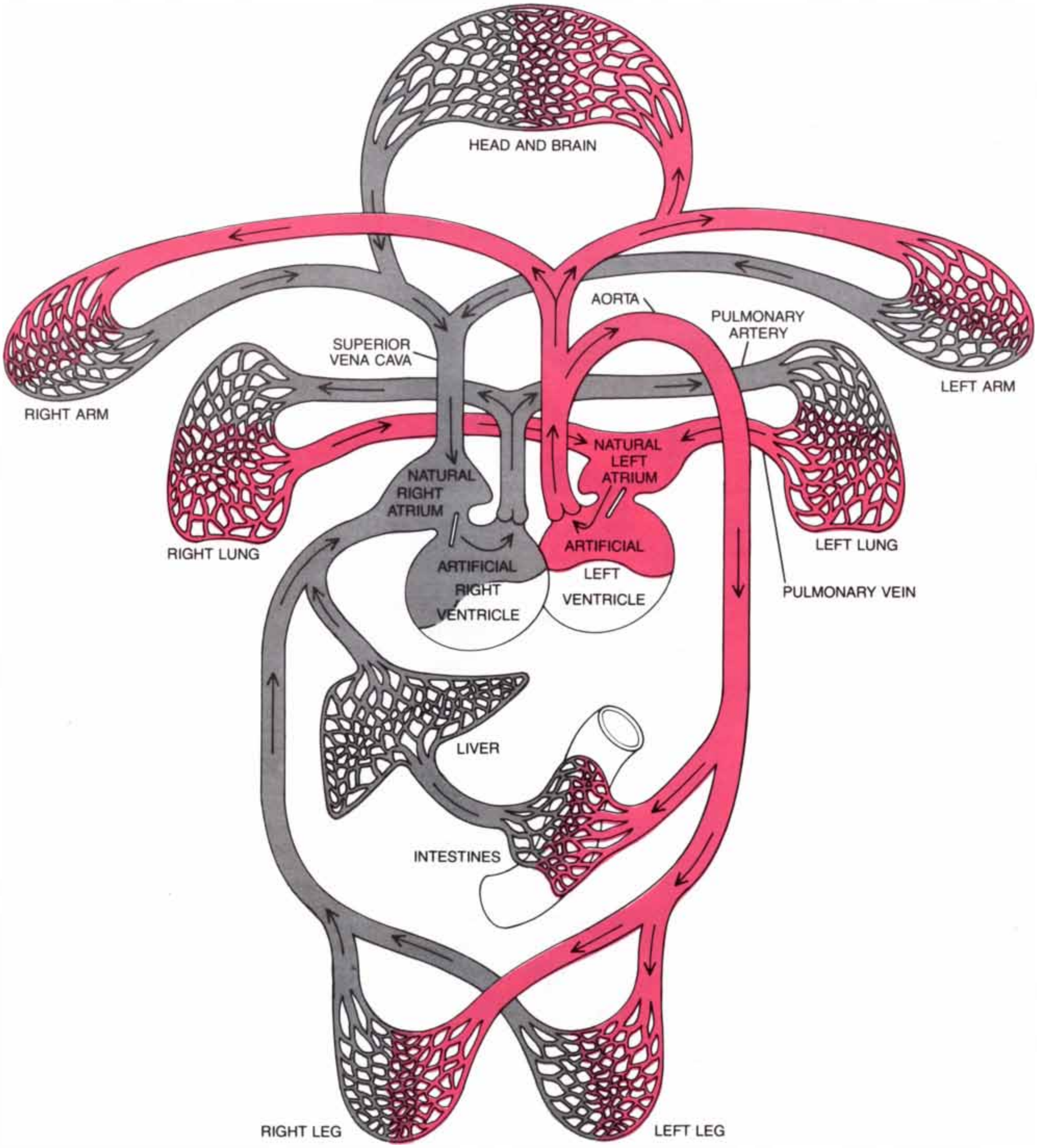
The total power output of the heart is about two and a half watts. About 80 percent of the work is done by the left ventricle, which pumps blood into the arteries serving the organs and tissues of the body and which must generate arterial pressures exceeding 100 millimeters of mercury. The right ventricle does less work (about half a watt) because it delivers blood at much lower pressure to



ARTIFICIAL HEART is the Jarvik-7, the most recent in a series of such devices designed by the author. It is intended to fit the human anatomy but has been tested experimentally in calves, one of which survived for more than seven months. The right and left ventricles are made of polyurethane supported on aluminum bases. Polycarbonate rings support tilting-disk valves. The Dacron grafts depicted above the heart provide the means of coupling the device to the circulatory system. This heart is shown with a self-contained electric drive, which has been tested only in animals. The hearts implanted in long-surviving animals have run on an external pneumatic drive.

the artery leading to the lungs. The atriums contribute a negligible amount of energy.

The volume of blood pumped by the right ventricle must be approximately equal to that pumped by the left ventricle, but both volumes must change to meet the demands of the body. In a normal 150-pound individual the cardiac output at rest is from five to six liters per minute. The output doubles with moderate exercise; with strenuous exercise by a well-conditioned athlete it can reach 30 liters per minute. These increases during exertion result partly from faster heartbeats and partly from an increased stroke volume (the amount of blood pumped with each beat). The mechanism of the heart's response to exercise is an indirect one. Exercise increases the volume and the pressure of the blood returning to the heart, which in turn results in increased cardiac output. This regulatory behavior is called



HUMAN CIRCULATORY SYSTEM is depicted schematically to indicate the role of an artificial heart. Arterial blood (*color*) and venous blood (*gray*) pass through the atriums, or storage chambers, of the natural heart and are pumped by the artificial ventricles. As with the natural heart, blood is pumped by the left ventricle into the aorta. After passing through the various capillary beds, where it supplies the tissues with oxygen and nutrients and removes wastes, it returns via the vena cavas to the right atrium. From there it enters the right ventricle and is pumped through the lungs, where the blood gives up carbon dioxide and is oxygenated. In the next stage of the cycle it passes by way of the left atrium into the left ventricle, from which it is pumped out once again, thereby setting in progress a new cycle.

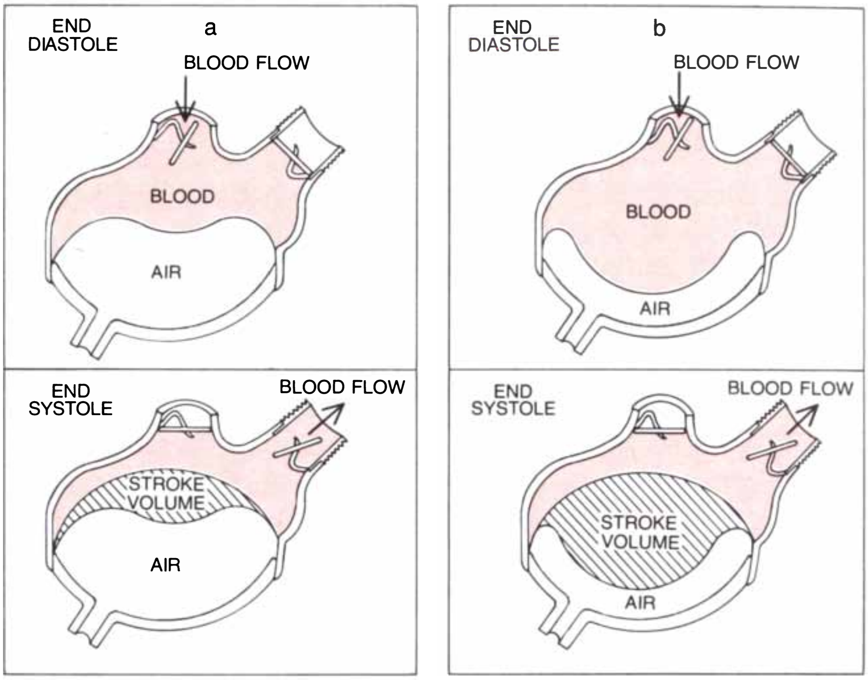
Starling's law, after the British physiologist Ernest Henry Starling.

At rest the normal human heart beats from 60 to 100 times per minute, slower in conditioned athletes and faster with exercise. In a year it beats about 40 million times. In the average lifetime it beats about three billion times and pumps some five million liters of blood, enough to fill a garden hose stretching around the earth at the Equator.

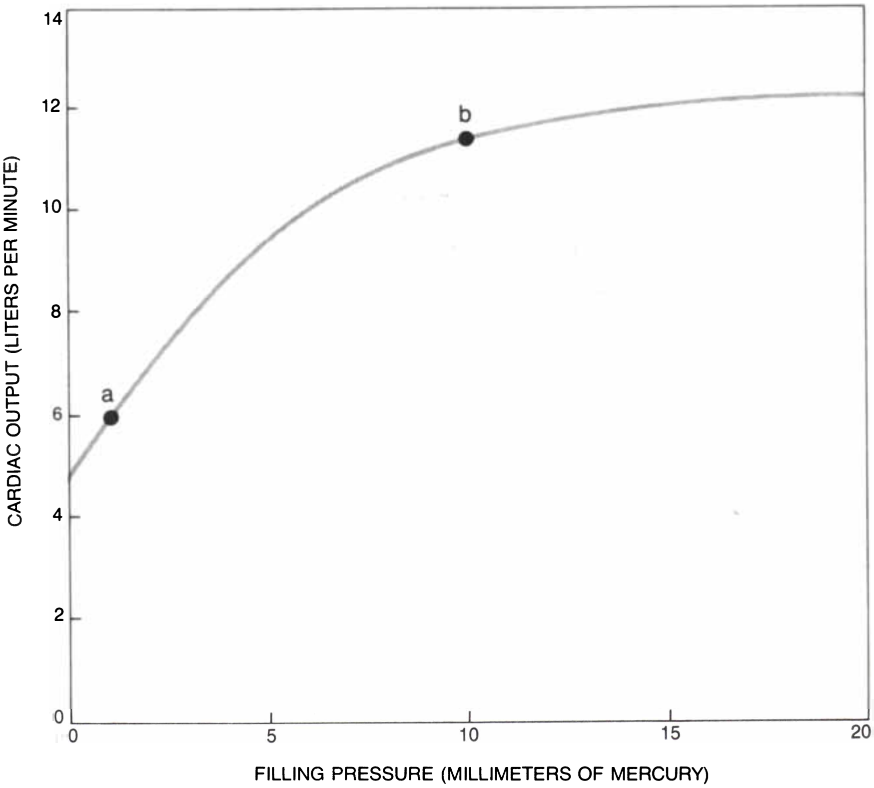
When the development of a total artificial heart was begun, many people considered the goal unattainable. The natural heart was recognized as a remarkable organ, difficult to duplicate even from the standpoint of durability alone. The complex regulatory mechanisms of cardiac function, which call on poorly understood interactions between the heart and the nervous system, seemed still more challenging. In the late 1950's heart transplants had not yet been attempted, and it was not yet known that the heart can regulate itself adequately even if it is isolated from the nervous system. Open-heart surgery was in its infancy and work on artificial valves for the heart had just begun.

The first experiments with total artificial hearts were done in dogs. With a heart made of the plastic polyvinyl chloride and driven by compressed air from an outside source, Kolff and Akutsu were able to keep dogs alive for about an hour and a half in 1957. Over the next few years Kolff's group at the Cleveland Clinic developed and implanted several other kinds of artificial heart driven by electricity. One such device employed five solenoids that displaced oil; the oil in turn compressed polyurethane sacks that held the blood. Animal survivals of three hours were obtained with this heart. In another electrically driven heart an electric motor drove a roller that compressed a blood-carrying tube against a foam-lined housing. The heart needed only outflow valves, but it caused excessive hemolysis and sustained life for only two hours. In the pendulum heart a pivoting electric motor alternately compressed two blood-containing sacks, thereby forcing blood out of the ventricles. Several dogs survived for from four to six hours with this device, but its output was inadequate and it caused excessive hemolysis.

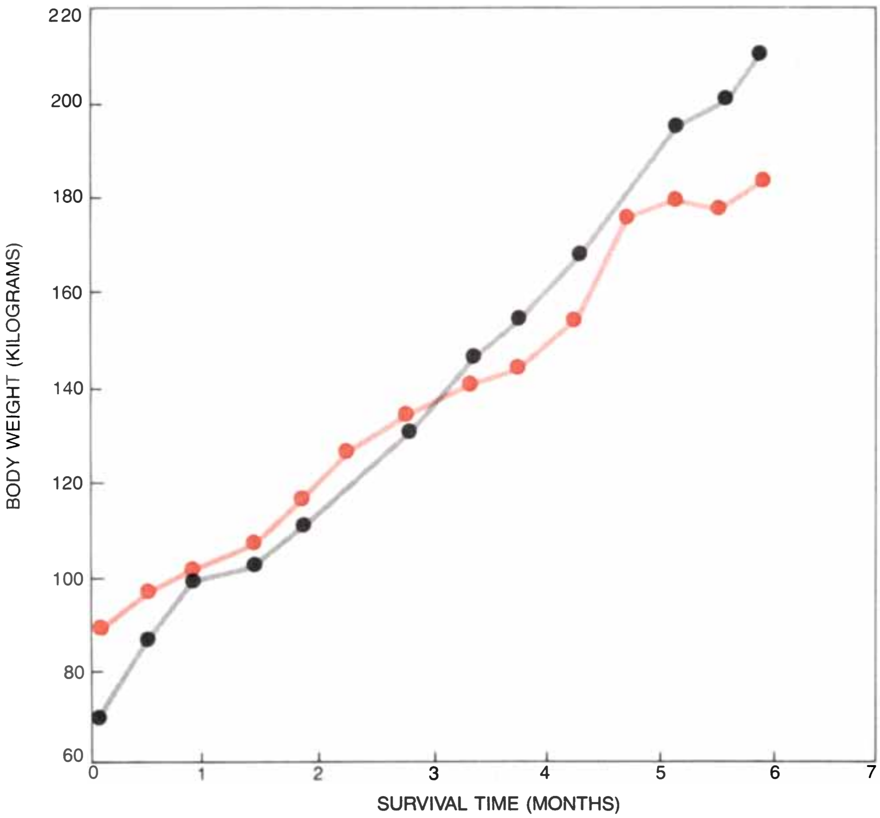
In these early experiments the surgical techniques for implanting an artificial heart were poorly developed, so that survivals of even a few hours were a considerable accomplishment. The recipient animals, however, hardly regained consciousness after surgery. The investigators recognized the difficulty of incorporating the energy converter and the blood-handling mechanism in the available space, and so they returned to the use of compressed air from an external supply. The decision simplified the machines that had to be implanted in the



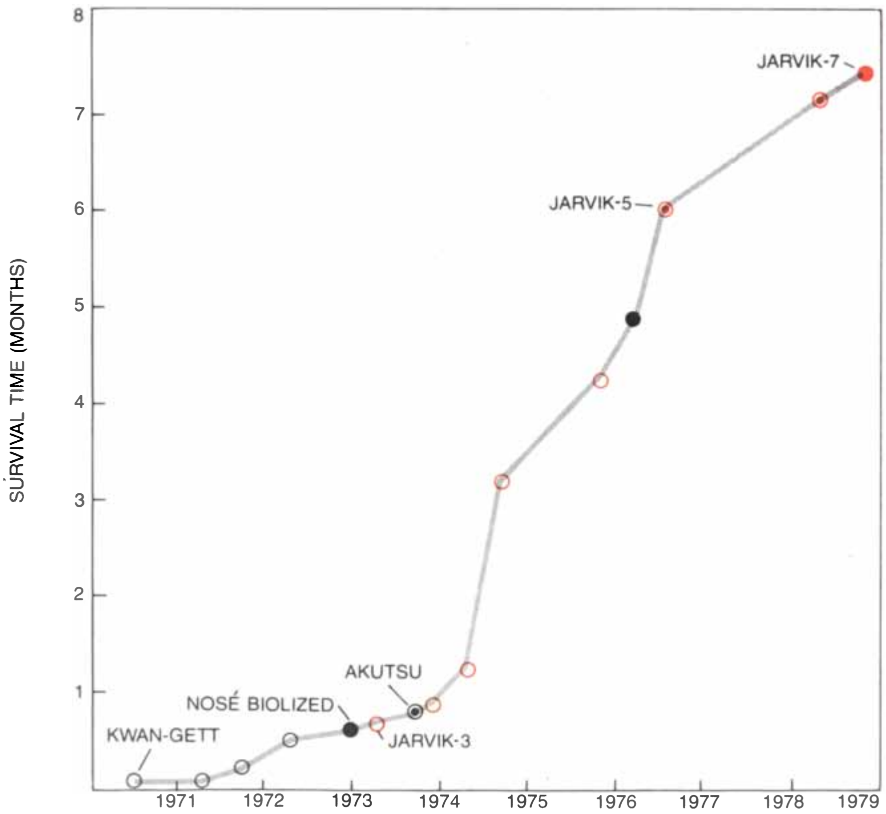
REGULATORY MECHANISM of an artificial heart causes the output of the device to rise when the venous return rises. Here an air-driven artificial heart is shown in two conditions of diastole (the filling stage) and systole (the pumping stage). With a normal flow of venous blood (*a*) the output of the heart is about six liters per minute; with greater activity by the animal in which the heart has been implanted (*b*) the venous pressure increases and the output increases to about 11 liters per minute. Since the animal's vascular system, which is regulated by nervous and hormonal mechanisms, primarily determines the volume of venous return to the heart, the body controls the cardiac output; the device itself responds passively, without changing the heart rate, by varying the volume of the stroke (the amount of blood pumped in each beat).



OUTPUT AND PRESSURE are charted for an artificial heart under conditions where the venous return of blood, as reflected by the filling pressure, is rising. Points *a* and *b* on the curve correspond to the two diastole conditions depicted in the illustration at the top of this page.



WEIGHT GAIN of several calves with artificial hearts is plotted (*color*) in comparison with the growth of a normal control calf (*gray*) over an equal period of time. The data for the calves fitted with the implanted hearts represent the average gain by as many as 11 animals in the earlier months of the experiments; in the later months the number was considerably smaller.



INCREASE OF SURVIVAL of animals with implanted total artificial hearts was considerable in the 1970's. The curve reflects a series of world records set by six different artificial hearts.

chest, but it did little to improve the survival period.

Notwithstanding these discouraging results, the concept of an artificial heart gradually drew the interest of other investigators and organizations. In 1963 the National Aeronautics and Space Administration participated in the development of a computer-based control system that regulated the output of an artificial heart according to various physiological conditions, such as pressure in the atriums. Congress appropriated money for an artificial-heart program to be conducted by the National Heart, Lung, and Blood Institute. Seven organizations joined in an assessment of the technology of the artificial heart, of the need for the device and of the problems associated with its development.

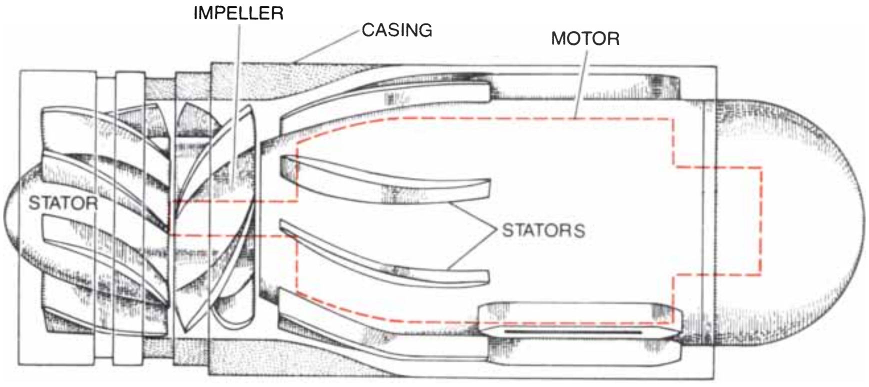
Meanwhile the work with animals continued at the Cleveland Clinic. By the mid-1960's Kolff and Yuke Nosé had kept calves alive for a little longer than a day with hearts of the sack type. Other groups took up the task, some of them introducing novel approaches. For example, William S. Pierce at the University of Pennsylvania School of Medicine tried to work with only a left ventricle, bypassing the right one with a shunt from the vena cava to the pulmonary artery (so that the blood flowed directly from the venous circulation into the lungs, without passing through the heart).

In 1968 the National Heart, Lung, and Blood Institute entered into contracts with three companies to develop nuclear-powered hearts. Two other companies participated in a fourth nuclear-heart program financed by the Atomic Energy Commission. The goal was to develop a fully implantable artificial heart that would operate for 10 years without external support. The various systems utilized heat from the radioactive isotope plutonium 238 to run an engine of the Stirling-cycle, Rankine or thermocompressor type. The engine in turn drove a blood pump pneumatically, hydraulically or mechanically.

Up to now the nuclear-powered artificial heart has had only limited success. The devices are complicated, bulky and expensive. No total artificial-heart system driven by nuclear power has supported an animal for more than two days, although engines for such systems have functioned in calves for a month and blood pumps designed for the nuclear heart but driven by electric motors have sustained calves for as long as 38 days.

In 1969 an artificial heart was installed in a human subject for the first time by Denton A. Cooley of the Texas Heart Institute. The heart, driven by compressed air, had been developed by Domingo Liotta and Charles W. Hall. It sustained life for some 64 hours until a natural heart became available and was transplanted. No further trials in human patients have been undertaken.

By the end of the 1960's the achievements of artificial hearts were still modest. Animal survivals remained less than three days. One major problem was mechanical breakdown. Another was that the devices did not fit well and tended to obstruct the flow of venous blood into the right atrium. Fluid would accumulate in the lungs because the artificial left ventricle did not function adequately. Small clots formed throughout the circulatory system, using up so much of the clotting factors in the blood that uncontrolled bleeding from external or internal wounds became a risk. The devices also caused excessive hemolysis.



MINIATURE PUMP for an electrohydraulic artificial heart, which could be implanted internally and would replace the external pneumatic drive, has only one moving part. It has two stator sets and utilizes an impeller driven by a brushless direct-current motor. The unit has the volume of a medium-size flashlight battery and can generate a flow of 45 liters per minute.

By 1970 the feasibility of the artificial heart still appeared doubtful. A series of developments beginning about then, however, made the prospects for the device look better. A major advance was the design of a heart whose pumping element was a diaphragm. This principle was introduced by Clifford S. Kwan-Gett, who had been with Kolff at the Cleveland Clinic and who moved with him to the University of Utah College of Medicine in 1967. The Kwan-Gett heart was free of problems with mechanical breakage and hemolysis. With it survivals of one week were achieved in 1971 and two weeks in 1972.

The first models of the Kwan-Gett heart were made of smooth silicone rubber. The material apparently caused excessive clotting and so gave rise to uncontrolled bleeding. In order to overcome the problem tiny fibers of Dacron were glued to the surface of the rubber to keep small clots in place. The hope was that the inside of the heart would become coated with a smooth layer of fibrin (the major constituent of blood clots) and the formation of small blood clots would cease. On the fibrin layer endothelial cells might then grow, making the interior surfaces of the device similar to those of a natural heart. The fibers helped in the short run, but eventually the buildup of fibrin on the inner surface became so thick that it impaired the filling of the atriums or blocked the motion of the diaphragm.

A different approach to the clotting problem was taken by Nosé. He developed what he called a "biolized" heart, in which the surfaces that came in contact with blood were made from natural tissues (such as pericardium) modified by treatment with chemical fixatives to make them tougher and immunologically inert. Early in 1973 a heart of this kind functioned in a calf for 17 days.

A problem with the Nosé heart and with many other designs was right-side failure resulting from the poor fit of the artificial device to the natural anatomy. The blood vessels leading into the device tended to kink, obstructing filling and resulting in inadequate output. In 1972 I designed the Jarvik-3 heart with the aim of overcoming these deficiencies. As with earlier improvements in the artificial heart, this design raised the maximum time of survival a bit more. In 1973 a group of investigators led by Akutsu, who had moved to the University of Mississippi School of Medicine, did an animal experiment in which the survival time was 25 days. A year later I fitted the Jarvik-3 heart with a highly flexible three-layer diaphragm made of smooth polyurethane. Earlier hearts had had a single-layer diaphragm of polyurethane, which had ruptured too easily. The material I chose for the diaphragm was Biomer, a medical grade of Lycra, which is an elastic material used in brassieres and girdles. The heart with the three-layer diaphragm soon increased survival times to as much as four months. Gradually, in the latter part of the 1970's, the survival period was increased further by various groups: to five months by the workers at the Cleveland Clinic, to six and a half months by a group in West Berlin and finally to the 221 days achieved at the University of Utah with the Jarvik-7 heart in 1979.

The Jarvik-7 was designed to fit the human anatomy, although it has been experimentally implanted in the calf, which has substantial anatomical differences. (My colleagues and I are now switching to sheep because they do not grow to a body size far beyond what the heart was designed for.) The two ventricles of the Jarvik-7 are made of polyurethane supported on aluminum bases. Rings of polycarbonate support tilting-disk valves. During the installation of the heart the animal is put on a heart-lung machine, the natural ventricles are removed and polyurethane cuffs are sutured to the two remaining atriums and the two large arteries that connect with the heart. The separate right and left ventricles of the artificial heart are then snapped onto the cuffs and connected to the external driving system by means of two tubes.

In the record survival the heart sustained the growth of the calf from 200 pounds at surgery to more than 350 pounds seven months later. The heart was intended for a person weighing about 150 pounds. The calf was able to walk on a treadmill for an hour even after it had grown to twice the size of the intended recipient. The animal's blood chemistry remained normal for most of the experiment.

A serious problem, which was discovered at autopsy, was the formation of pannus: the uncontrolled growth of connective tissue at the suture points. The tissue had spread across the inflow opening of the artificial heart, reducing its size and limiting the cardiac output. The phenomenon is rarely seen in people who are given prosthetic heart valves, and so it seems likely that the formation of pannus in calves is related to the fact that the animal is growing rapidly. Presumably it will not be a problem in human beings who receive an artificial heart. At the University of Tokyo the experimental animal is the goat, and no pannus formation has been found in goats that survived for as long as eight months with an artificial heart connected outside the body.

How long will it be before total artificial hearts are routinely implanted in human beings? Probably at least a decade, although clinical trials on a small scale may begin sooner. The experience of six laboratories that have maintained animals on an artificial heart for five months or more is encouraging. The devices themselves, the surgical techniques and the postoperative care have all been highly developed.

Certainly the feasibility of an artificial heart has been demonstrated. When such hearts were first implanted in animals more than 20 years ago, it was hypothesized that the heart functions solely as a mechanical pump and that an appropriate artificial device could sustain that function for long periods of time, maintaining the recipient in good health. Today these concepts can be considered facts rather than hypotheses.

Why, then, will it be so long before artificial hearts are routinely implanted? The answer is to be found mainly in ethical considerations and in the allocation of the limited money available for research on the artificial heart. The pneumatically powered artificial hearts that have proved successful in animals are not portable. The animal is confined to a cage, tethered to a large drive system and exercised only on a treadmill. Such conditions would be unacceptable for human beings. Even if compressed-air drives were made portable, the large pneumatic tubes that enter the chest would be uncomfortable and would carry a high risk of infection at the points of entry.

Perhaps many people would accept such limitations if the alternative was death. It may not be. A good deal of work has been done on devices that assist the natural heart by supplementing the function of the left ventricle. More than 40 such devices have been installed temporarily in individuals with severe heart disease. The hearts of several of the patients have healed.

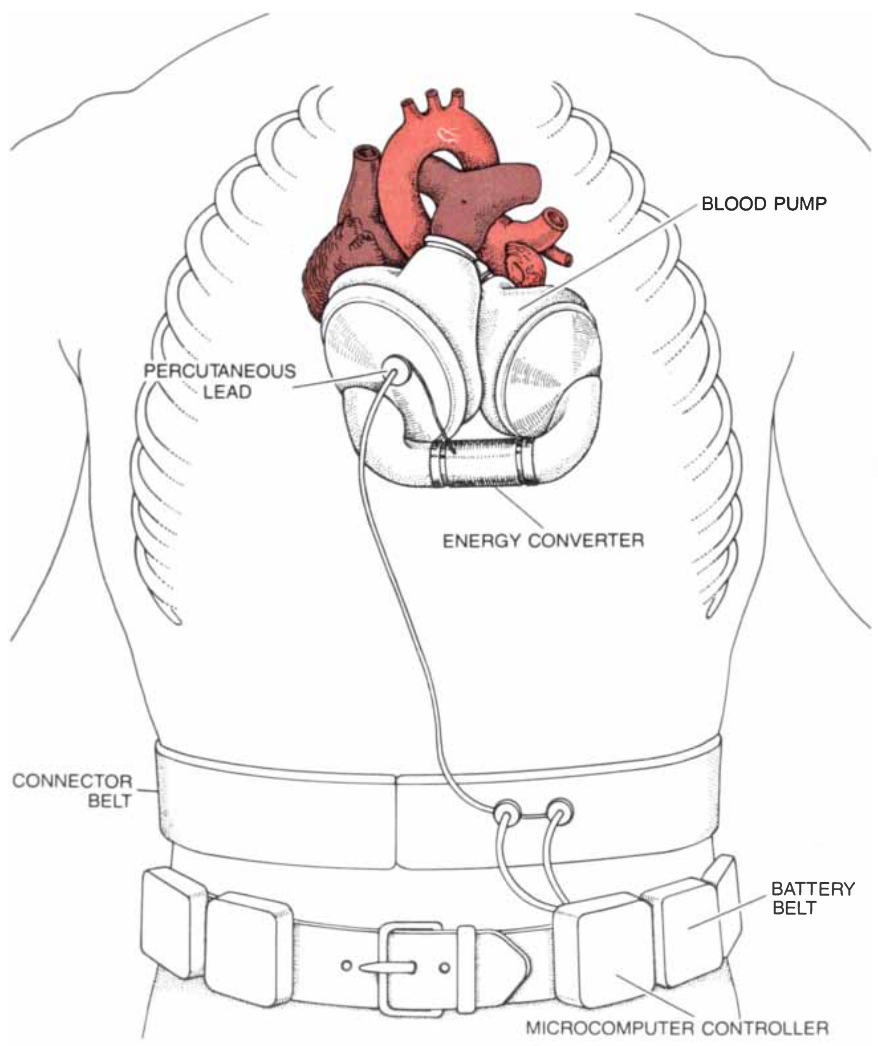
The trials of the left-ventricle-assistance devices have been carried out by William Bernhard of the Children's Hospital Medical Center in Boston, by Pierce, who had moved to the Pennsylvania State University College of Medicine, by John Norman of the Texas Heart Institute and by Marko Turina and Åke Senning of University Hospital in Zurich. The Cardiology Advisory Board of the National Heart, Lung, and Blood Institute has recommended that the development of such devices be given a higher priority than the development of a total artificial heart. Although the decision is a controversial one, it has directed the funding of work on artificial devices under contracts from the National Institutes of Health exclusively to left-ventricle-assistance apparatus.

One line of work that applies to both the left-heart devices and the total artificial heart is the development of electric-energy converters and blood pumps. My colleagues and I at the University of Utah, in collaboration with Milton Isaacson of Nu-Tech, Inc., are working on an electrohydraulic energy converter that has only one moving part. The impeller of an axial-flow pump is attached to the rotor of a brushless direct-current motor, with the impeller and rotor supported by a single hydrodynamic bearing. Reversing the rotation of the pump reverses the direction of the hydraulic flow. The hydraulic fluid (silicone oil of low viscosity) actuates the diaphragm of a blood pump just as compressed air does. In a left-heart device the axial-flow pump moves the hydraulic fluid from a reservoir sack into the blood pump and back. In a total artificial heart the hydraulic fluid is pumped back and forth between the right and the left ventricles.

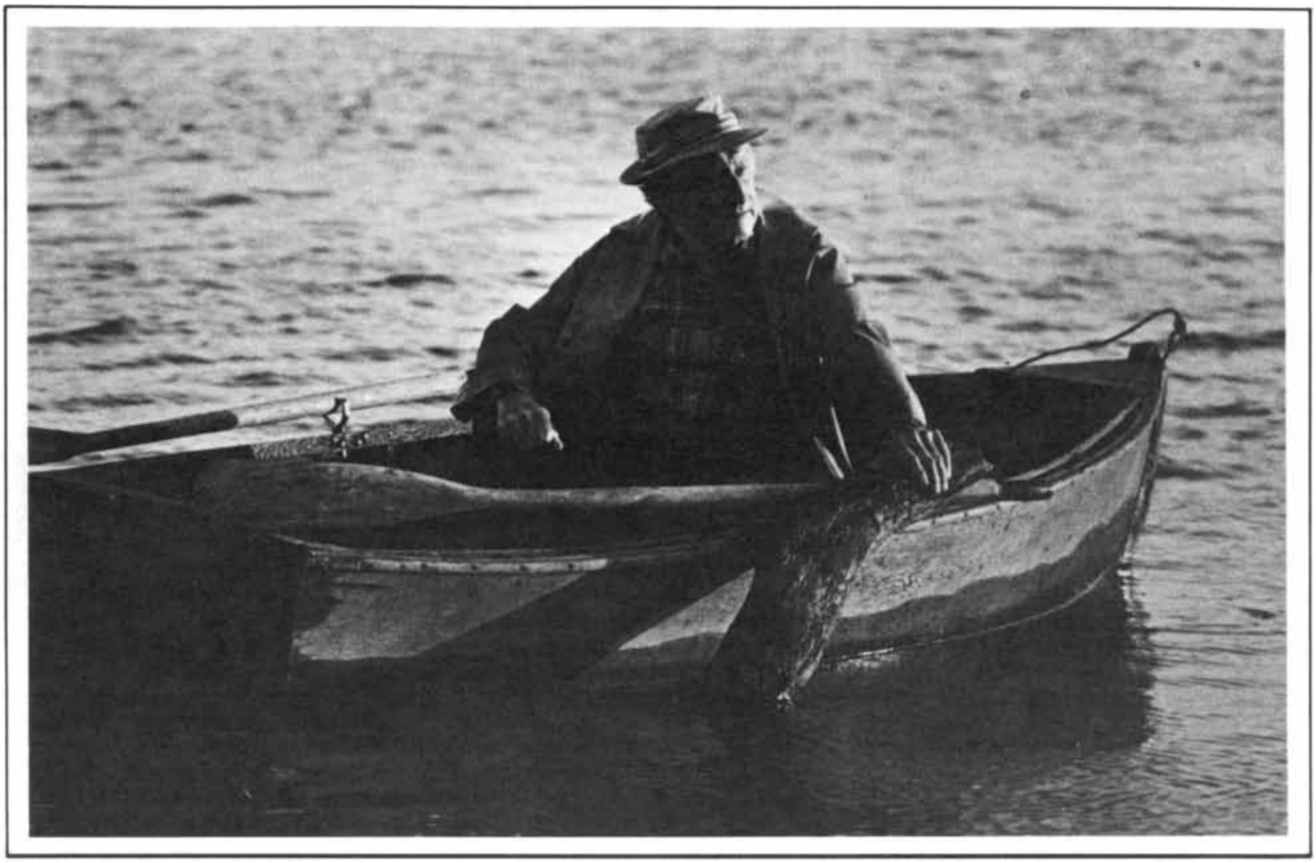
The energy converter is so small and so mechanically simple that it can be implanted without impinging on vital structures. It weighs 85 grams and occupies about 30 cubic centimeters, or approximately the volume of a medium-size flashlight battery. The converter requires an external battery and electronics package, which is connected to the heart by a small cable that passes through the chest. The batteries weigh from two to five pounds and could be worn in a vest or on a belt. They would have to be replaced with recharged batteries once or twice a day. The great advantage of such a power source is that the patient would be fully mobile. Although this device illustrates many of the desirable features of a portable electric artificial heart for human use, it has been tested only in a model circulatory system and in a few calves. Years of further work will be needed to demonstrate the reliability and durability of the system before it can be tried in human patients.

It should be evident from this account that substantial technical problems remain to be solved before a total artificial heart can be routinely utilized in human patients. Ethical, social and economic considerations must also be dealt with. When the artificial heart has been perfected, it must be made available in sufficient quantity to serve a large number of people. The criteria for selecting recipients must be defined clearly and objectively to ensure a short period of hospitalization and a rapid recovery for a high percentage of the recipients. Patients will need sociological and psychological counseling to help them adapt to a situation new in human experience. Nowhere else will the dependence of life on technology and machines be more apparent.

If the artificial heart is ever to achieve its objective, it must be more than a pump. It must also be more than functional, reliable and dependable. It must be forgettable.



PROSPECTIVE ARRANGEMENT of an electrically driven artificial-heart system is portrayed. The power would come from batteries worn on a belt and would be transmitted through a lead that carried a small cable through the skin. The cable would also carry the control signals from the microcomputer controller. The artificial-heart patient would be fully mobile.



Some people prefer to push a button and go zipping across the water.

Some don't.

If you are that way about photography, write us a letter.

Although we owe a century of success in business to having made photography blissfully easy, we make no such claim about the film on which this photograph was taken, Kodak technical pan film 2415. It comes in 36-exposure 135 cartridges, 35 mm long rolls, and 4″ x 5″ sheets. Though it is not generally available in 8″ x 10″ sheets, you would judge that it was made with an 8 x 10 camera if the press that printed this magazine did not have to come between your eye and the original print. Such a judgement would be made by a person who looks upon photography as a fine art that goes on making its demands on the photographer *after* the button has been pressed.

Actually, photographer Bob Clemens found an SLR more manageable on the marshy shore than an 8 x 10 view camera would have been. With this film and a special low-contrast compensating developer (which we will tell you how to make up if you write us), he sacrificed nothing to graininess. We estimate that enlargement would have to go to 25X to show the same graininess as an 8″ x 10″ print from a 35 mm negative on Kodak Tri-X pan film.

Seen as a gallery print, this photograph would really show off the breadth of its tonal range. Yet this self-same film is exactly what we recommend for such high-contrast uses as lecture slides to project type matter or a diagram in bright white light on a dark black screen! It's all in how it's handled.

Formerly, this film was available under the designation "SO-115." But not very available, by comparison with better-known Kodak films. Last year's sales figures convinced us how wrong we were in thinking of it as just a technical film for such specialties as sunspots and holograms. Now, as "2415," we've made it an over-the-counter item; but, before shopping for it, request further information from Gordon P. Brown, Scientific and Technical Photography, Eastman Kodak Company, Rochester, NY 14650. Just mention this ad.

Bear in mind, though, that you must enjoy doing what ordinary mortals just don't bother with.

Kodak

SCIENCE AND THE CITIZEN

Five-Year Checkup

The rapid advances of recent years in both biomedical knowledge and medical technology have not been paralleled by improvement in world health. There is a gap—perhaps an increasing one—between medical capabilities on the one hand and the prevention and cure of disease on the other. In the developed countries as well as the underdeveloped ones poverty continues to be at the root of the most pressing health problems. The traditional approach to improving medical care, emphasizing the application in hospitals of technology to cure acute diseases of selected patients, cannot deal with the massive health problems of the poor countries and is becoming too expensive for the rich ones. These are among the conclusions of the World Health Organization's *Sixth Report on the World Health Situation,* which covers the years from 1973 through 1977.

To those aware of the decline of infectious diseases in industrial nations and the preeminent role of heart diseases and cancer in mortality it comes as a surprise that "the same basic complex of infectious, parasitic and respiratory diseases, compounded by nutritional deficiencies, still accounts for most of the world's deaths." Apart from one success, the eradication of smallpox, there has been "little or no progress" in reducing the prevalence of the diseases that most severely afflict the underdeveloped countries, which have three-fourths of the world's population. The incidence of most of those diseases "can be dramatically reduced at relatively modest cost," and so "the suffering and death they cause must be accounted as unnecessary and preventable."

One of the most striking developments of the five-year period was a serious resurgence of malaria, particularly in Southeast Asia. The number of reported cases more than doubled as many countries reduced their antimalaria activities after a partially successful earlier drive to eradicate the disease. In tropical Africa, it is estimated, at least a million children under 14 die of the disease every year. The transmission of malaria can be interrupted only by the elimination of the malaria parasite's *Anopheles* mosquito vectors; the ubiquity of the mosquitoes in many parts of the world, the cost of insecticides and the increasing resistance of the mosquitoes to them combine to make the elimination extremely difficult. Other approaches include the widespread use of antimalaria drugs; again cost is a problem, along with the malaria parasite's increasing resistance to some drugs. As for a vaccine, according to the WHO report it may take as much as 10 more years to develop one.

In addition to malaria a number of other parasitic diseases that are known in the developed countries only as strange-sounding names are daily facts of life in many underdeveloped countries. Schistosomiasis is endemic in 71 countries; its causative agent, a snail-borne worm, infests some 200 million people; other worms cause filariasis, which afflicts at least 250 million people, and onchocerciasis, a major cause of blindness in Africa.

Tuberculosis and five childhood diseases—diphtheria, whooping cough, tetanus, measles and poliomyelitis—are preventable by immunization, but fewer than 10 percent of the 80 million children born each year are being fully immunized. The six diseases are thought to cause some five million deaths per year among children less than five years old and to blind, cripple or otherwise disable an additional five million. There has been an ominous increase in the number of cases of paralytic poliomyelitis in particular. A major WHO objective is to provide immunization against all six diseases for all children by 1990.

In the developed countries diseases of the heart and blood vessels remain the leading cause of death among men and the second or third among women. The good news is that the death rate from cardiovascular diseases has remained level in most of those countries and has declined in several, notably the U.S., for reasons that are still obscure (see "The Rise and Fall of Ischemic Heart Disease," by Reuel A. Stallones; SCIENTIFIC AMERICAN, November, 1980). The bad news is that cardiovascular disease has begun to be a significant health problem in some underdeveloped countries. There the major problem has been not heart attacks, as it is in the developed countries, but heart-valve damage as a sequel to rheumatic fever, which in turn is a complication of streptococcal infections. Rheumatic fever can be prevented by treating the primary infection with penicillin; long-term maintenance on penicillin also protects children or young adults with a history of rheumatic fever from reinfections that would cause or exacerbate damage to the heart valves. The crippling and often fatal effects of rheumatic heart disease are therefore preventable. The task is to identify streptococcal infections, administer penicillin and then make sure that penicillin is given monthly into young adulthood. In most parts of the world that is not easy.

Cancer, until recently a serious public-health concern only in the developed countries, is also beginning to be a major cause of sickness and death in underdeveloped countries, such as China, where infectious and parasitic diseases are being brought under control. The actual prevention of cancer is never likely to be simple, the WHO report points out; even when a direct-acting risk factor is identified, eliminating it may call for changing deep-seated cultural habits. "Excessive exposure to sunlight is now recognized as the cause of much of the skin cancer that occurs in Caucasians, but discouragement of sunbathing is an unpopular measure." The most clearly implicated risk factor is cigarette smoking, which significantly increases the risk of not only lung cancer but also cancers of the larynx, esophagus and bladder (as well as bronchitis, emphysema and coronary heart disease); it is estimated to be responsible for more than 40 percent of all male cancer deaths in the United Kingdom. Another recent WHO publication argues that cigarette smoking should be considered a pandemic, "since smoking is an unnatural form of behavior, resulting in impaired well-being and often in severe illness, and is spread throughout the world by a process of psychological contamination."

Perpetual Menace

With the U.S. and the U.S.S.R. apparently poised on the verge of a major acceleration of the nuclear-arms race, a sobering reminder of the nuclear facts of life and death has been issued by an international group of experts convened under the auspices of the United Nations. The group, composed of governmental and nongovernmental members from 12 countries and headed by Anders I. Thunborg, the permanent representative of Sweden to the UN, was appointed two years ago by Secretary General Kurt Waldheim to carry out a comprehensive study of nuclear weapons for submission to the 1981 session of the General Assembly. Their unanimous final report, the first such review conducted for the UN in more than a decade, provides a synopsis of available information on present nuclear arsenals, on trends in the development and deployment of nuclear-weapons systems and on the probable effects of their use. The study also discusses the implications for international security and negotiated disarmament of two critical factors: (1) the doctrine of deterrence and other theories concerning nuclear weapons and (2) the continued quantitative increase and qualitative improvement of nuclear weapons.

The authors of the UN report invoke as their text the "prophetic words" of Niels Bohr, who in 1944 wrote (in identical letters to President Roosevelt and to Prime Minister Churchill): "The fact


BLENDED SCOTCH WHISKY • 80 PROOF • IMPORTED BY CALVERT DIST. CO. N.Y.C.
PARIS
PASSPORT
SCOTCH/750 ML
ATHENS
Season's Greetings
PASSPORT
SCOTCH
Because you enjoy going first class.
This season, give the Scotch that's always in good taste. Passport. Made of Scotland's finest whiskies and honored in 126 countries around the world. It's always more rewarding to give a first-class gift.
Passport Scotch.

of immediate preponderance is that a weapon of unparalleled power is being created which will completely change all future conditions of warfare.... Unless, indeed, some agreement about the control of the use of the new active materials can be obtained in due time, any temporary advantage, however great, may be outweighed by a perpetual menace to human society."

With Bohr's words having gone unheeded, mankind now finds itself, according to the UN group, in a historically unique predicament: "Whereas most if not all previous instances of competitive arms buildup and rivalries in weapon development have eventually culminated in conflict on the battlefield, the present situation makes such a dénouement unthinkable.... Never before have States been in a position to destroy the very basis of the continued existence of other States or regions; never before has the destructive capacity of weapons been so immediate, complete and universal; never before has mankind been faced, as today, with the real danger of self-extinction."

In spite of the drastic alteration brought about by the advent of nuclear weapons in the very basis of all strategic military thinking, the report continues, "the nuclear arsenals in the world have continued to grow in numbers and in their destructive capability. There exist today at least 40,000 to 50,000 nuclear weapons, the combined explosive power of which is believed to be equivalent to that of more than one million Hiroshima bombs or, to put it differently, some 13 billion tons of TNT, which represents more than three tons for every man, woman and child on earth." Because of the continued increase not only in the number of independently targetable warheads but also in the accuracy with which these weapons can be delivered, "the lethality and effectiveness of the arsenals are enhanced much more than a numerical comparison of strategic launchers or warheads would indicate. Large numbers of nuclear weapons can now be used strategically in situations other than a mass attack on the homelands of the superpowers. There is also a growing capacity for 'theater' use of tactical nuclear weapons, which pose a threat to many States, for instance in Europe. Thus the nuclear-weapon powers are today prepared for rapid use of their tactical nuclear weapons in a war and for escalation of the level of nuclear violence. Nuclear overkill is everywhere."

A major part of the new UN review is devoted to a recitation of the catastrophic consequences that would ensue if the nuclear arsenals of the present or future nuclear-weapons states were ever to be unleashed in war. Although the nuclear-weapons states might be expected to suffer the heaviest casualties and the most damage in such a war, the report points out, "all nations in the world would experience grave physical consequences. Radioactive fallout could be a serious problem especially in countries adjacent to the belligerent States, and during the decades after a major nuclear war fallout would take a toll of millions worldwide, in present and future generations. Even more serious than radioactive fallout, however, would be the global consequences of a large nuclear war on the world economy and on vital functions of the international community. The sudden collapse of many of the world's leading trading nations as well as of established mechanisms for international transactions would lead to profound disorganization in world affairs and leave most other nations, even if physically intact, in desperate circumstances. Widespread famines could occur, both in poor developing countries and in industrialized nations. Those starving to death might eventually outnumber the direct fatalities in the belligerent countries."

In view of the enormity of the destruction that would result from a nuclear war, the UN experts express their concern over the supposed stability of the nuclear "balance of terror." They write: "The argument of the stability of the balance is one which gives the proponents of deterrence great difficulty. In order to claim that it is possible to continue, forever, to live with nuclear weapons, the balance must be maintained at all times irrespective of any technological challenges that may present themselves as a result of the arms race. In addition, there must be no accidents of a human or technical nature, which is an impossible requirement.... For these and other reasons it is not possible to offer a blanket guarantee of eternal stability of the deterrence balance, and no one should be permitted to issue calming declarations to this effect. The consequences of being wrong are too great. The chances of being wrong are too obvious."

Even if the balance of terror were an entirely stable phenomenon, the group adds, "there are strong moral and political arguments against a continued reliance on this balance. It is inadmissible that the prospect of the annihilation of human civilization is used by some States to promote their security. The future of mankind is then made hostage to the perceived security of a few nuclear-weapon States and most notably that of the two superpowers. It is furthermore not acceptable to establish, for the indefinite future, a world system of nuclear-weapon States and non-nuclear-weapon States. This very system carries within it the seed of nuclear-weapon proliferation. In the long run, therefore, it is a system that contains the origins of its own destruction."

In spite of the many arguments put forward in favor of negotiated disarmament in such international forums as the UN, "some countries have chosen to base their security perceptions on nuclear-weapon systems, in the hope that the balance of deterrence may remain stable. In particular, the superpowers perceive that nuclear weapons support their national security both by deterring a direct conflict between them and by increasing their influence in other areas of the world. At the same time, both are concerned that the other might achieve nuclear superiority. In the absence of verifiable measures of disarmament, these concerns are projected as justifying further quantitative increases and qualitative developments of their nuclear arsenals. But the doubtful stability of deterrence may well decrease as a result of the nuclear-arms race, even if both sides have agreed to seek nuclear parity. It is therefore highly questionable whether the security of the nuclear-weapon States—however defined—can be maintained on the basis of an arms race."

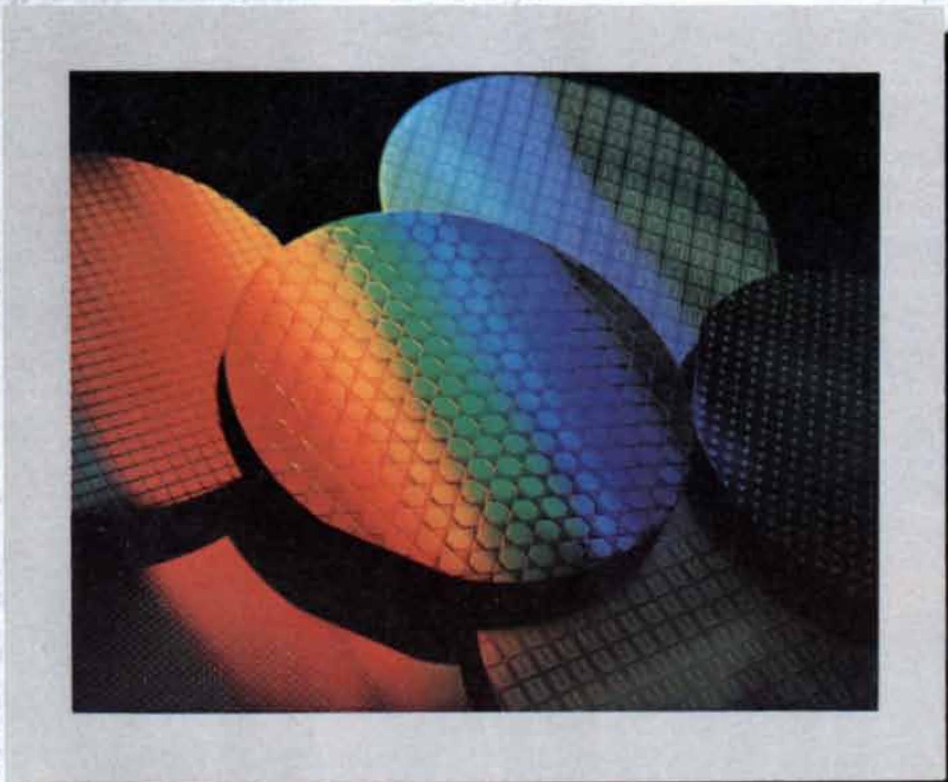
Citing nuclear weapons as without question "the most serious threat to international security," the UN report concludes: "So long as reliance continues to be placed upon the concept of the balance of nuclear deterrence as a method for maintaining peace, the prospects for the future will always remain dark, menacing and as uncertain as the fragile assumptions upon which they are based. Fortunately this is not the only alternative that is available to mankind. We have, in the United Nations, an institution which should be utilized for all the purposes and stages that are relevant to the process of disarmament—negotiation, agreement, implementation, verification and ratification where necessary. What is needed is the creation of a strong public opinion which should, in time, create the political will among all States to transfer their security reliance from the nuclear-weapon system to another universally accepted system. Only a system of international security based on the observation of the principles of the United Nations Charter and of other universally accepted instruments of international law can provide a mutually acceptable basis of security. This must therefore be the goal on the road to nuclear disarmament. The Charter and nuclear weapons date their existence from the same time. The future road should point to a full reliance on the Charter and to the elimination of all nuclear weapons."

Decline and Rise

No new order for a nuclear electric-power plant has been placed in the U.S. since 1978. The fact remains that for the present and the near future the


NEW DATSUN 280-ZX.
LUXURY IN THE FAST LANE.
Remember yesterday's image of a sports car? It had to be as noisy as a foundry. Ride like a rock. And be as spartan as a monk's cell.
Well, the Datsun 280-ZX has turned it all around. Up above you see its lavishly appointed interior. Every luxury has been well attended: from power windows to climate-control air conditioning. And for the first time, an optional T-bar roof is available on the roomy 2+2.
Even long-time owners of Cadillacs and Mercedes have seen the wisdom of opting for a sleek new Z-Car. It gives them all the luxury they're accustomed to, plus the mileage and range they need today.
This year's 280-ZX boasts a 10% increase in horsepower, just for the thrill of it. Yet this fuel-injected, 6 cylinder OHC powerhouse delivers 21 EPA estimated mpg, 32 estimated highway mileage with 5-speed transmission. Based on these figures, an estimated MPG range of 443 miles and a highway range of 675 miles await you. Use estimated MPG for comparison only. Your mileage and range may differ depending on speed, trip length and weather. Highway mileage will probably be lower. California highway estimates are slightly less.
Buy or lease one today. And get to know what AWESOME is. The Datsun 280-ZX is another example of superior workmanship from Nissan Motor Co., Ltd., a worldwide company whose name stands for quality.
NISSAN
DATSUN
WE ARE DRIVEN

To Jay Maisel, this is a sea of breathtaking tranquillity.



To industrial photographer Steve Grohe, so is this.

Both were taken with new Polacolor ER instant film.

When highly acclaimed photographer Jay Maisel wanted a faithful rendition of a scene he discovered while vacationing in Maine, he selected the same medium Steve Grohe used to capture the vivid detail of silicon wafers:

Polaroid's new Polacolor Extended Range instant film.

Not surprising. The same qualities that make new Polacolor ER film a great tool for artistic photography make it a great tool for industrial and commercial applications, too.

Its medium contrast emulsion and extended dynamic range enable it to not only handle scene brightness ranges up to 1:48 with excellent detail in both highlight and shadow areas, but also to provide increased exposure latitude where the scene brightness range is less than 1:48.

On top of that, Polacolor ER film's new magenta dye helps produce color quality that will satisfy the sensibility of the most demanding photographer.

And its excellent skin tones and broad tonal range offer a useful neutral position from which you can easily change to versions with varying degrees of contrast, depending on what's called for.

Needless to say, Polacolor ER film provides pictures that are ready on-the-spot for studying, filing, attaching to reports or whatever.

Like any good film, Polacolor ER has its own singular personality. You'll find its look and feel, its versatility, sensitivity and responsiveness make it another good choice for finished art.

Polacolor ER film is balanced for daylight and electronic flash and is rated at 125 ASA. It is available in two formats: 8x10 (Type 809) and 4x5 (Type 59).

For further details see your Polaroid professional film dealer, or in the continental U.S. call: 800-225-1618 (in Massachusetts call collect: 617-547-5177).

Polaroid
Instant Professional Films

share of electric power contributed by nuclear reactors in the U.S. will continue to rise. According to a recent report issued by the Atomic Industrial Forum, the nuclear-power industry's trade association, that share will reach 22 percent in the next decade. Indeed, in four states nuclear power already accounts for 50 percent or more of all the electricity generated.

The U.S. now has 74 nuclear-power reactors licensed to operate; their combined capacity is about 55,000 megawatts, which is 9.2 percent of the electric-power capacity in the U.S. and 11.4 percent of the electricity actually generated in 1979. In 1970 the figures were 1.9 percent of capacity and 1.5 percent of generation. The prediction is that by 1990 nuclear power will account for 22 percent of generation and 16 percent of capacity.

As of September 15, 1980, there were, in addition to the 74 nuclear plants with operating licenses, 87 for which construction permits have been issued, two with limited work authorizations and 19 on firm order. They represent respectively a generating capacity of 95,839 megawatts, 2,300 megawatts and 22,282 megawatts.

Even before 1978 the decline in orders for power reactors in the U.S. was precipitous. In 1973, the peak year, the number of orders was 41. In 1974 it was 26; in 1975, four; in 1976, three; in 1977, four, and in 1978, two. Moreover, between January 1 and September 15, 1980, work on 48 projects was delayed, eight cancellations were announced and one reactor was retired. The average time from the formal filing of an application for a construction permit for a plant to the start of operation rose from 5.1 years in 1970 to 10.7 in 1978.

The four states in which nuclear power now accounts for at least half of the electricity generated are Vermont (78 percent), Maine (60 percent), Connecticut (53 percent) and Nebraska (50 percent). Regionally New England relies on nuclear power for 34 percent of its power generation, the Middle Atlantic states for about 17 percent, the South Atlantic for 15 percent, the East North Central and West North Central for 14 percent each, the Pacific for 6 percent, the West South Central for 2 percent and the Mountain states for less than 1 percent.

In many other developed countries nuclear-power development continues at a rapid pace. Among the countries that rely on nuclear power for a substantial part of their generating capacity are Belgium, Finland, France, the Federal Republic of Germany, Japan, Sweden, Switzerland, Taiwan and the U.K. The figures range from 17.5 percent for Switzerland to 10 percent for Finland, France and the U.K. According to the Atomic Industrial Forum report, it is predicted that by 1985 the figures will be 28 percent in Sweden, 38 percent in Belgium and 50 percent in France.

Gassing Up

The conversion of biological wastes into fuel is contemplated throughout the world, but in China's Sichuan province a massive program is already under way. The province's Office for the Popularization of Biogas estimates that some five million converters (called digesters) had been constructed as of August, 1979. In certain rural areas of the province nine-tenths of all the families had built one for their household. Somewhat larger digesters are operated communally. Each digester is an airtight chamber in which the fermentation of a mixture of animal dung, human excrement and crop residue such as straw yields a clean-burning gas that is one-half to three-fourths methane. The gas powers not only burners and lamps but also pumps and motors for such purposes as irrigation and the grinding of grain. Experiments are in progress toward converting the gas into a fuel to power small tractors.

An overview of the program in Sichuan is provided by E. Ariane van Buren of the International Institute for Environment and Development, who writes in a report of the United Nations Environment Programme and also in an article in the Swedish journal *Ambio*. In Sichuan, van Buren attended a course by which more than 200,000 "biogas technicians" have been trained in the past five years. The trainees have dispersed throughout the province as part-time teachers of the technology. In Beijing she attended an agricultural fair at which she saw a film, "The Flower of Biogas." Its message, she reports, "is clear. This is a technology to which China is committed." The film "is straightforward and explicit: even the less attractive aspects of manure handling . . . are shown as simple and easily performed."

A typical digester, van Buren observes, is a squat cylinder with a slightly domed roof. It encloses a volume of 10 cubic meters, and it can be constructed from roughly 1,000 bricks of low-grade concrete. The building supplies for such a digester are 300 kilograms of slaked lime, 70 kilograms of cement, 300 kilograms of cinders, 900 kilograms of sand and 4,000 kilograms of clay. The first two items cost 20 yuan ($14); the last three are simply dug up. The digester is built underground, a strategy that saves aboveground space, reduces the required thickness of the walls and provides thermal insulation. Often the digester is built directly under a pigsty.

In one demonstration a newly built digester was loaded with several hundred kilograms of a mixture of straw and manure that had been composted for a few weeks. Bucketfuls of "active sludge taken from another digester and from a human excreta storage pit" were added, along with water. A space of 60 to 80 centimeters at the top of the converter was left for gas to accumulate in. The strategy of loading the digester is under intensive study. The effort is to ensure that the fermenting material attains an optimum temperature and *p*H and the accumulating gas attains an optimum pressure and composition. A 10-cubic-meter digester can yield as much as two cubic meters of biogas per day, a quantity said to be sufficient for a family of five to cook its meals and boil 15 liters of water.

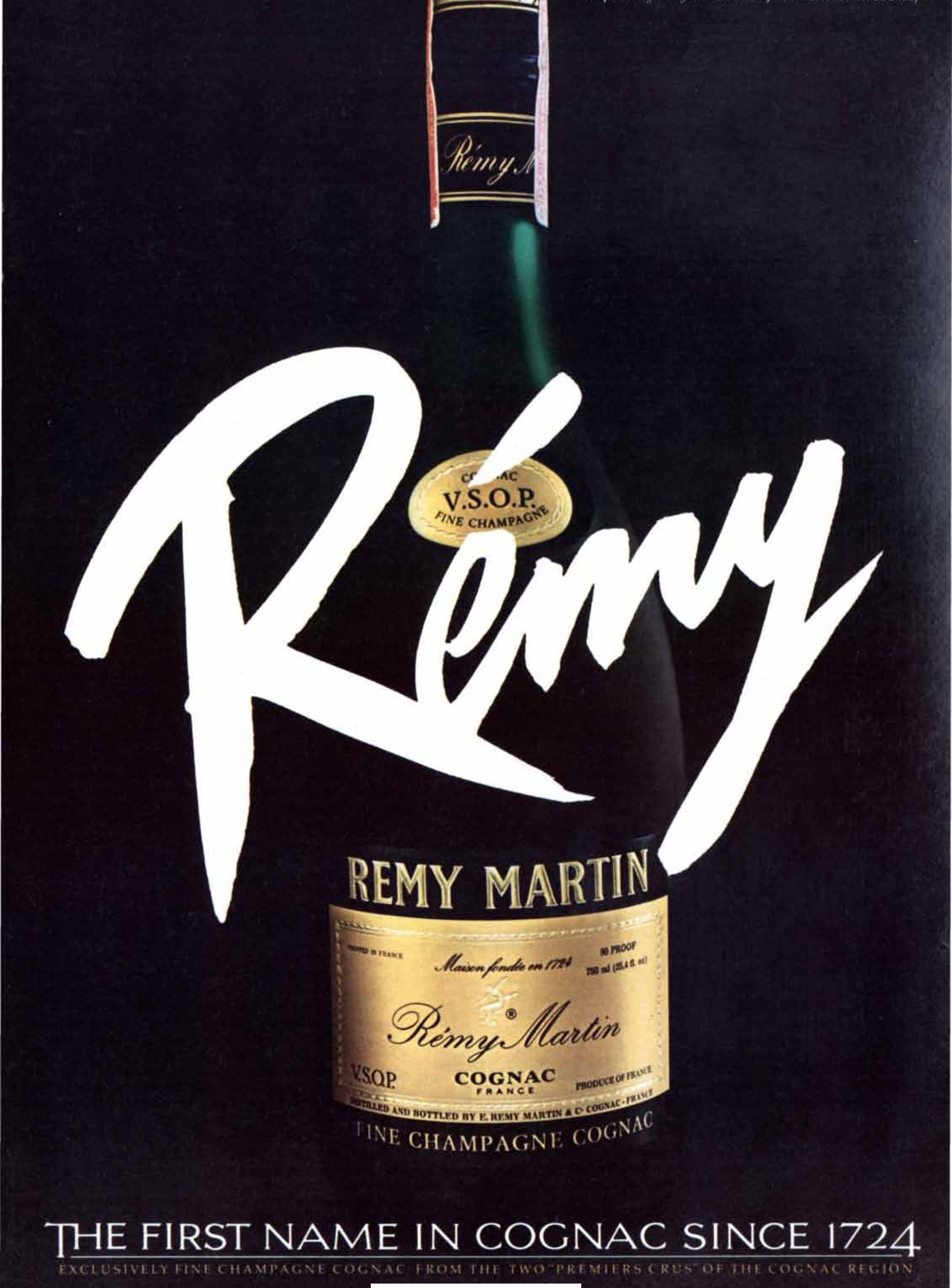
The digester has one further function. As time passes the parasitic organisms in the fermenting wastes tend to die off. Hydrocarbons are being liberated as gas, but nitrogen, phosphorus and other plant nutrients remain in the solid material. Twice a year at planting time, therefore, the sediment in the digester is removed and mixed into the soil as a fertilizer that is more hygienic than the unprocessed excrement was.

To remove the sediment a cover is opened at the top of the digester. It is the only part of the chamber that is aboveground. The rest of the year the cover is sealed by a gasket of clay whose integrity is maintained by a design in which the cover and the gasket are at the bottom of a shallow pool of water. Van Buren found that aquatic plants such as hyacinths are grown in the pool and then fed to pigs, whose manure goes into the tank. "Our suggestion of putting the aquatic plants directly into the digester was greeted," she reports, "with surprise." The Chinese thought it inefficient to bypass "a potential stage of nutrition in the cycle of resource use."

The Detoxification of Isaac Newton

In 1692 and 1693 Isaac Newton avoided contact with other people and suffered from severe insomnia, loss of appetite, absentmindedness, depression and occasional delusions of persecution. The cause of his distemper has been a subject of controversy. In 1979 two papers in *Notes and Records of the Royal Society of London* proposed that his illness was chiefly the result of poisoning by the mercury he used in his chemical and optical experiments. (The papers were discussed in this department for December, 1979.) Now in the same journal R. W. Ditchburn of the University of Reading argues that Newton's breakdown was not caused by any kind of poisoning but was what would be called today clinical depression or depressive illness.

The symptoms of Newton's disorder have been learned from letters he wrote to friends. For example, in a letter


Rémy
COGNAC
V.S.O.P.
FINE CHAMPAGNE
REMY MARTIN
Maison fondée en 1724
90 PROOF
Rémy Martin
V.S.O.P.
COGNAC
FRANCE
PRODUCE OF FRANCE
DISTILLED AND BOTTLED BY E. REMY MARTIN & Cie COGNAC • FRANCE
FINE CHAMPAGNE COGNAC
THE FIRST NAME IN COGNAC SINCE 1724
EXCLUSIVELY FINE CHAMPAGNE COGNAC FROM THE TWO "PREMIERS CRUS" OF THE COGNAC REGION

to Samuel Pepys dated September 13, 1693, Newton wrote: "I am extremely troubled at the embroilment I am in, and have neither ate well nor slept well this twelve month, nor have my former consistency of mind. I never designed to get anything by your interest, nor by King James's favour, but am now sensible that I must withdraw from your acquaintance, and see neither you nor the rest of my friends any more."

Three days later Newton wrote to John Locke: "Being of the opinion that you endeavoured to embroil me with women and by other means, I was so much affected by it as that when one told me you were sickly and might not live I answered twere better if you were dead. I desire you to forgive me this uncharitableness. For I am now satisfied that what you have done is just and I beg your pardon for having hard thoughts of you."

The case for attributing Newton's illness to mercury poisoning, which was independently proposed by Myron L. Wolbarsht of the Duke University Eye Center and Larry W. Johnson and by P. E. Spargo of the University of Cape Town and C. A. Pounds of the Atomic Weapons Research Establishment at Aldermaston in England, is based on three arguments. First, the symptoms of the illness appear to be those of mercurialism. Second, in his alchemical experiments Newton is known to have been exposed to the toxic vapors of a wide variety of metals, including mercury, lead, arsenic and antimony. Third, four surviving locks of Newton's hair show high concentrations of lead, arsenic and mercury.

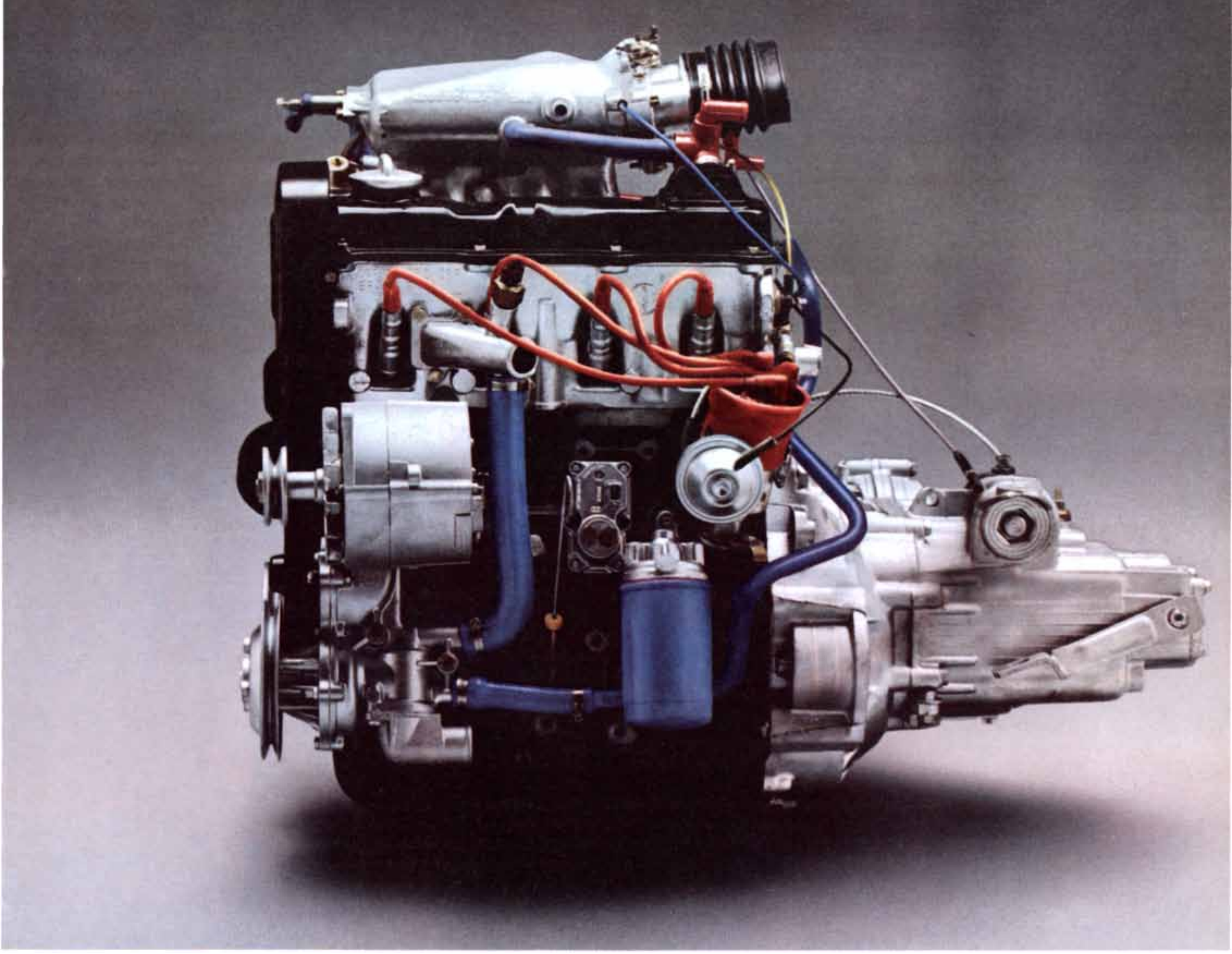
Ditchburn disputes all three arguments. He maintains that the symptoms of Newton's illness match neither the physical nor the psychological effects of mercurialism. According to a report published by an expert committee of the Common Market Commission, the physical symptoms of mercury poisoning include weakness and pain in the limbs, tremors in the fingers, paresthesia ("pins and needles") in the fingers and toes, inflammation of the mouth and gums and the loss of teeth. There is no evidence that Newton showed any such physical symptom. Moreover, it is known with some certainty that he did not show some of the symptoms. For example, in his adult life he lost only one tooth. And surviving samples of Newton's handwriting over the two years of his breakdown do not reveal any sign of finger tremor.

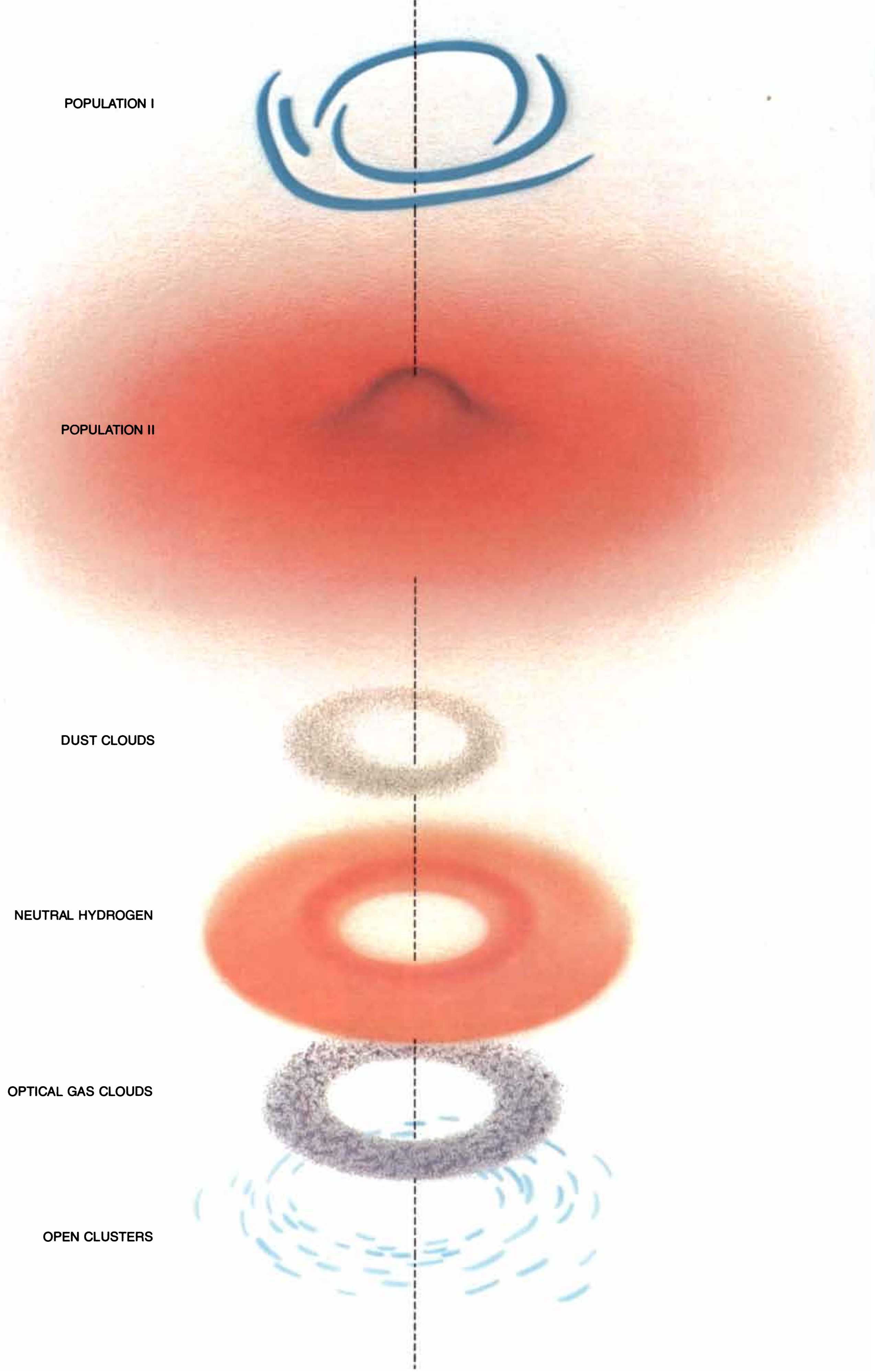
The chief psychological symptom of mercury poisoning is irritability. According to Ditchburn, Newton did not exhibit this aggressive kind of behavior. Indeed, his illness was characterized mostly by passivity and withdrawal. The only symptoms he had in common with mercurialism were depression, insomnia and delusion, but such effects do not always accompany mercury poisoning and are not specific to it.

In any event, Ditchburn asserts, Newton was not exposed to enough mercury to have been poisoned. Newton is thought to have conducted "several hundred" chemical experiments over a period of 20 years. Ditchburn argues: Suppose Newton did 800 experiments, half of which involved mercury. His notebooks indicate that most of his chemical investigations took less than an hour. Therefore he was not exposed to mercury for more than 20 hours per year, which is only 1 percent of the exposure time of industrial workers (in the mirror and hat industries) who are known to have developed mercurialism.

What about the four locks of Newton's hair in which Spargo and Pounds have found unusually high concentrations of toxic metals by employing the ultrasensitive modern techniques of neutron-activation analysis and atomic-absorption spectrophotometry? Ditchburn does not dispute the authenticity of the hair samples, one of which was found in the library of Trinity College, one in a book in Newton's own collection and two in the library of the current Earl of Portsmouth. He maintains that they do not bear on the cause of Newton's illness because nothing is known about when they were taken from his head. Hair grows about 10 centimeters per year, and so unless the locks were taken in 1692 or 1693 any toxic material that might have been responsible for his illness would have grown out. Spargo and Pounds say only that "the locks were taken from Newton's head after his death though this is by no means certain." Newton died in 1727, so that his hair would have grown some three and a half meters, and hence would have been cut many times, between his breakdown and his death.

Even on the remote chance that the samples of hair were taken in 1692 or 1693, writes Ditchburn, there are good reasons to doubt the interpretation of the modern chemical analysis. He points out that Spargo and Pounds's measurements show that Newton's locks also contain an abnormally high concentration of chlorine and bromine. The amount of these two elements implied by their concentration would have had serious physiological effects, none of which Newton seems to have shown. Moreover, the chemical analysis of one sample revealed a concentration of sodium far lower than normal. This suggests that the chemical composition of the locks changed as they lay pressed in books for more than two and a half centuries. The measured concentration of mercury is therefore suspect. All in all, Ditchburn concludes, the cause of Newton's illness was not physical but psychological.




POPULATION I
POPULATION II
DUST CLOUDS
NEUTRAL HYDROGEN
OPTICAL GAS CLOUDS
OPEN CLUSTERS

The Andromeda Galaxy

The large spiral galaxy nearest our own, it has been a laboratory for the study of the evolution of stars and galaxies. Even today it presents puzzles; for example, how are its spiral arms arrayed?

by Paul W. Hodge

The year was 1611. Night watchmen made their rounds throughout the towns of Europe. The flames of their candles were protected from wind and rain in lanterns whose windows had a thin covering of horn. The Bavarian astronomer Simon Marius directed his telescope toward a nebulous patch of light in the region of the sky occupied by the constellation Andromeda. He likened the nebulosity to a "candle seen at night through a horn."

His description gives a good idea of how the object now called the Andromeda galaxy looks to someone with a small telescope, but it fails to suggest the place the Andromeda galaxy has held in the history of astronomy. Through large modern telescopes the Andromeda galaxy is seen to be a giant spiral galaxy. It is thought to be similar in form to our own galaxy, the Milky Way. One difference is that the Andromeda galaxy seems to be twice as large: it may include as many as 400 billion stars. At a distance of two million light-years from the solar system it is the spiral galaxy nearest our own, and it is the only giant spiral close enough for us to view it in detail. The Milky Way itself is less open to an overall inspection because, from the vantage of the earth, dust clouds hide its structure. The Andromeda galaxy has therefore been responsible for breakthroughs in the understanding of such matters as the evolution of stars, the rotation of galaxies and the scale of distances in the universe. Sources of X rays in the galaxy and of emissions in the radio part of the electromagnetic spectrum are now being examined. With the construction of new telescopes on the earth and in orbit the Andromeda galaxy becomes a prime target for exploration in powerful new ways.

Proof of Distance

The first modern study of the Andromeda galaxy was made almost 100 years ago, when photography first presented a way to record light too dim for the eye to see, and thereby to probe deep into space. With a 20-inch telescope Isaac Roberts took the first photographs of the Andromeda galaxy that showed its spiral structure. The photographs also suggested the presence of faint stars in the outer parts of the spiral, but that clue to the nature of the object was not understood at the time. Instead the Great Nebula in Andromeda was taken to be a cloud of gas that might eventually condense to form a star with a planetary system. It seemed to be the largest, the brightest and therefore probably the nearest among hundreds of similar nebulas. It was thus thought to be relatively close to the solar system.

The idea that star systems lay beyond the Milky Way soon occurred to several investigators, including Edwin P. Hubble of the Mount Wilson Observatory. In 1925 Hubble demonstrated that the small, inconspicuous nebula NGC 6822 is a distant aggregation of stars. Meanwhile the great spiral in Andromeda occupied much of his attention. It was the subject of his landmark paper, "A Spiral Nebula as a Stellar System," published in 1929. Hubble's many photographs of the object showed a huge, amorphous bulge of light surrounded by tightly wound spiral arms consisting of dust clouds, star clusters and thousands of points of light, each one a star.

Hubble's proof of the great distance of the Andromeda nebula derived from his discovery of 40 pulsating stars in the spiral. The examination of successive photographs of the galaxy showed that the stars were periodically brightening and dimming. Hubble recognized the stars as Cepheid variables, which are also found in the Milky Way. Harlow Shapley had already shown that Cepheid variables could serve as astronomical yardsticks: the intrinsic luminosity of a Cepheid variable is proportional to the period of its brightening and dimming, whereas its apparent luminosity can of course be measured directly. The ratio of the two luminosities is proportional to the square of the distance of the star. Using Shapley's calibration of the relation between period and intrinsic luminosity, Hubble concluded that the Andromeda spiral must be almost a million light-years from the earth, far beyond the edge of the Milky Way. Subsequent work has shown that the Andromeda galaxy is actually more than twice as far as Hubble calculated. I shall return to this matter below.

By comparing a succession of photographs Hubble also discovered 63 stars that flared up in luminosity and then slowly dimmed. These are novas. It is now thought the flares of light are emitted when gas escapes from a giant star, falls to the surface of a hot, dense companion star and explodes in a sudden episode of nuclear fusion. In effect the

EXPLODED DIAGRAM shows the distribution of the principal types of object in the Andromeda galaxy. For each tier of the illustration the vertical broken line marks the center of the galaxy and the horizontal direction corresponds to the major axis of the galaxy's disklike image in the sky. Stars of Population I (*top tier*) are young and blue. They are arrayed in segments that suggest a highly imperfect spiral pattern. The pattern in the illustration reflects the author's interpretation of the distribution of the stars. Stars of Population II (*second tier*) are old and red. They make up the central bulge of the galaxy and to a lesser extent the galactic disk. They extend in the illustration to a radius of some 130,000 light-years, which places them beyond the optical image of the galaxy in the picture on the next page. Stars of Population II also make up the aggregations known as globular clusters (not shown), which are scattered above and below the plane of the galaxy. Dust clouds (*third tier*) are visible because they redden or even block the light from the stars behind them. Their distribution is greatest toward the northeast (the left in the illustration). Neutral (un-ionized) hydrogen (*fourth tier*) is detected by its emission in the radio part of the electromagnetic spectrum. Its distribution has the shape of a doughnut. Evidence of star formation is provided by optical gas clouds (*fifth tier*), which are made luminous by their proximity to hot, young stars. Their distribution is more or less coextensive with the region of spiral-arm segments. Open clusters (*bottom tier*) are loose aggregations of stars of Population I. Their distribution, however, does not evoke a clear spiral pattern.

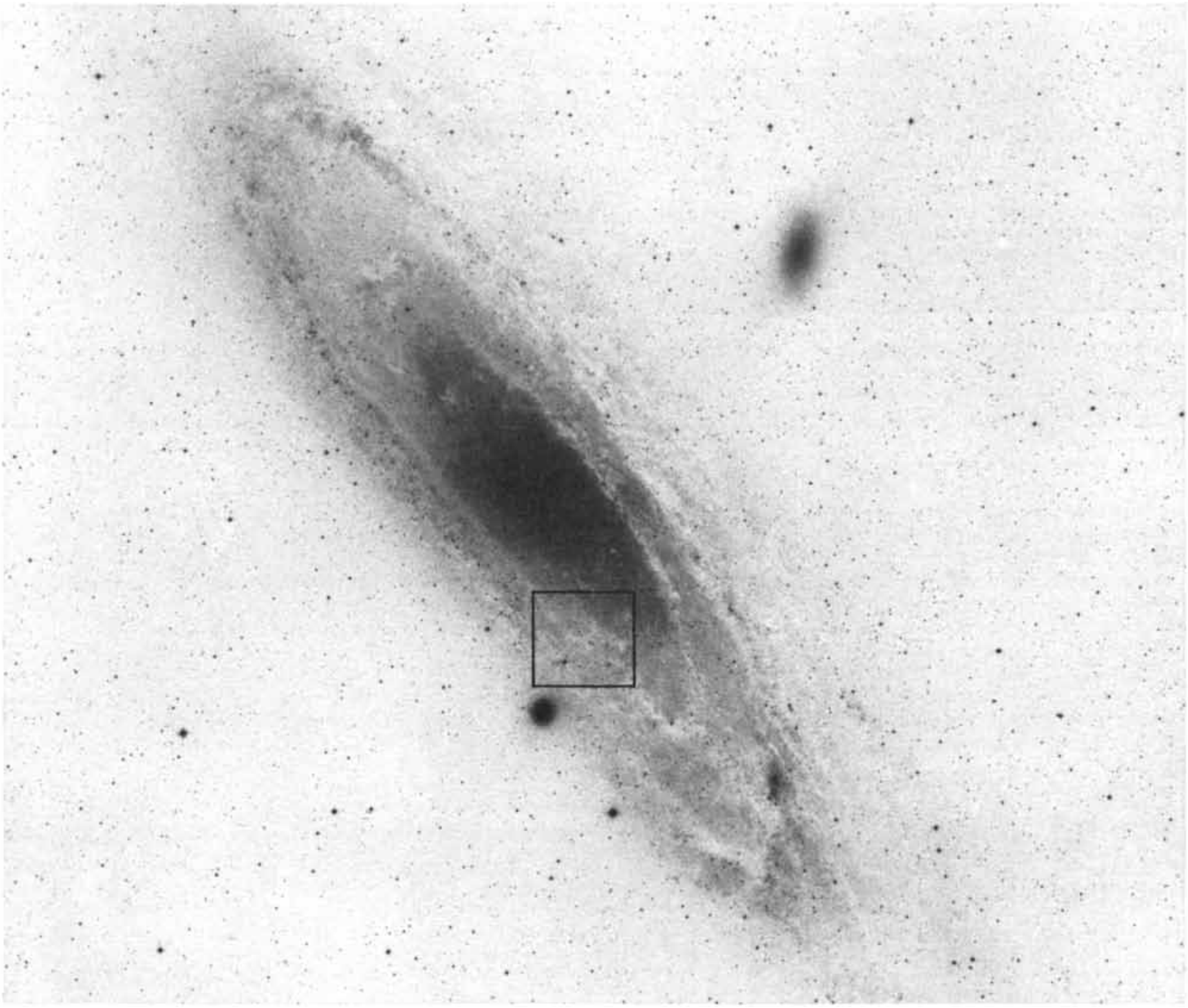
surface of the companion star becomes a hydrogen bomb. The novas in the Milky Way show that at maximum brightness these objects are approximately equal in luminosity to the brightest normal stars. In the Andromeda galaxy Hubble found that the novas were quite faint but were nonetheless similar in luminosity to the brightest stars in that galaxy. This was further proof of the galaxy's great distance.

In 1885 another event identified as a nova had been observed near the center of the Andromeda nebula. At its brightest it was extraordinarily luminous. Indeed, it was almost visible to the unaided eye. In later debates the event, called S Andromeda, was cited as evidence that the nebula must be within our galaxy. Otherwise, it was argued, S Andromeda would have been much too bright to be a nova. Hubble's study of the many novas in the Andromeda galaxy convinced him that S Andromeda belonged to an exceptional class of stellar explosions just being recognized at the time. They are now called supernovas. Quite unlike ordinary novas, supernovas are explosions that destroy a star. So far S Andromeda is the only supernova observed in the Andromeda galaxy. Statistics in other galaxies suggest that we should not be surprised to see another one any day.

Baade's Discoveries

The next important figure in the exploration of the Andromeda galaxy was Walter Baade, who systematically explored the galaxy with the 100-inch telescope on Mount Wilson. His work led to two discoveries that thoroughly revised our understanding of stars and our scale of distances in the universe.

The first discovery was made in 1944. Baade was a German national at the time, and so he was not able to join the American war effort. This left him almost the sole user of the 100-inch telescope and allowed him ample time for experimenting with new photographic emulsions and with optical filters of various colors. During the war the nearby city of Los Angeles held occasional defensive blackouts. On those nights the sky was unusually dark, and Baade was able to extend his photography of the Andromeda galaxy to record progressively fainter features.



ANDROMEDA GALAXY is arrayed from northeast to southwest in the sky above the Northern Hemisphere. Its plane is oblique with respect to the earth, so that the nearest part of the galaxy is the northern edge of its disk. The photograph shown here is printed as a negative; hence the dark parts of the galaxy consist of stars. The white parts consist of dust, which conceals the stars behind it. A prominent dust lane lies to the right of the galaxy's central bulge. The dark spot below the bulge is M32, a companion galaxy. A second companion galaxy, NGC 205, lies above and to the right of the bulge. Two other companion galaxies are outside the field of view. The diameter of this image of the Andromeda galaxy is about 125,000 light-years, and the galaxy is two million light-years from the earth. The galaxy thus subtends an angle of three degrees in the sky, or six times the angle subtended by the moon. The black rectangle marks the part of the galaxy in the illustration on the opposite page. The photograph was made with the 48-inch Schmidt telescope on Palomar Mountain.

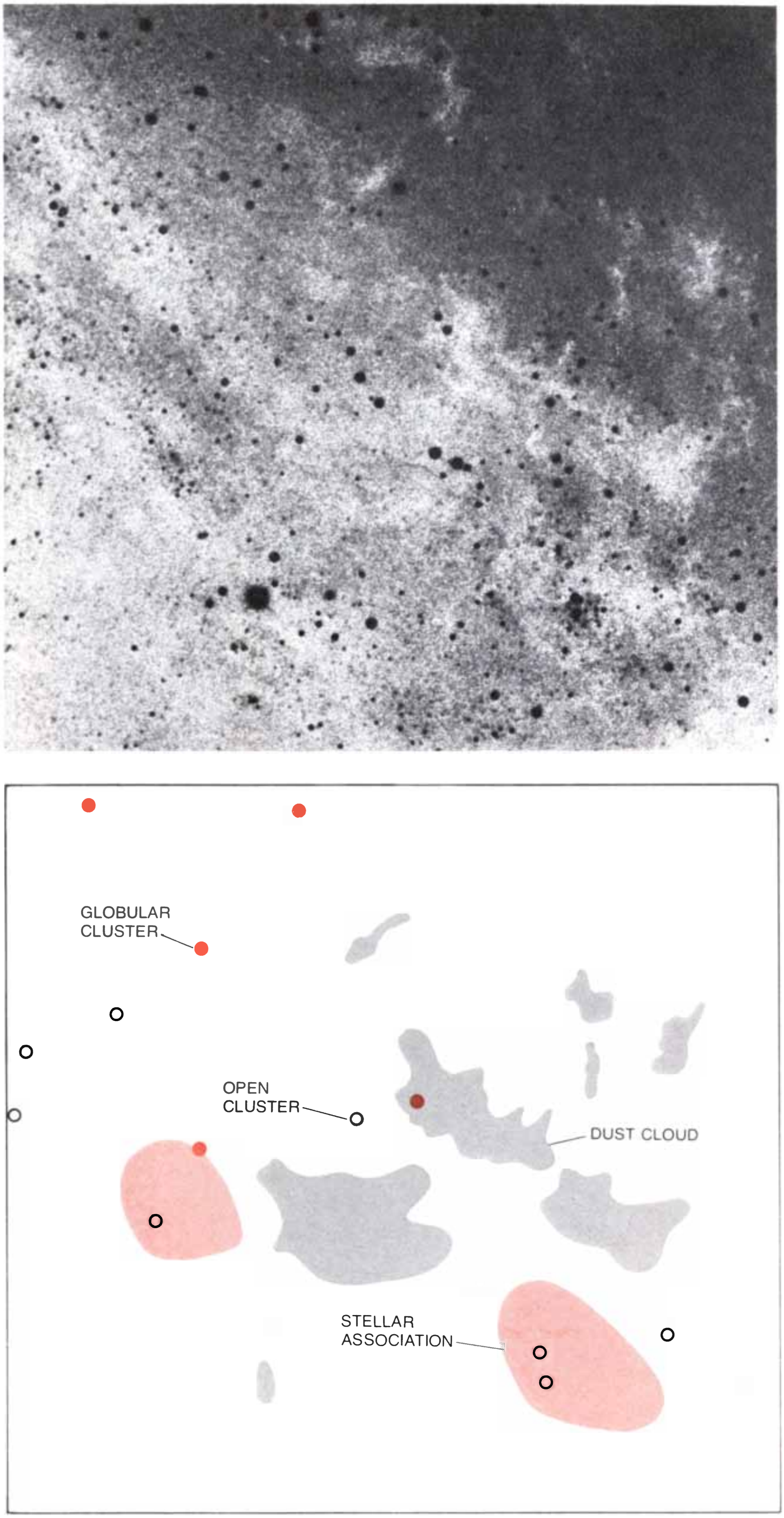
Baade was trying to understand why the spiral arms of the Andromeda galaxy were well resolved into individual stars in photographs of the galaxy, whereas the bright, amorphous central region defied resolution even on Hubble's best photographic plates. The light from the central region doubtless came from millions of stars. Why could none of them be seen? To compound the mystery, four small, somewhat featureless galaxies are companions of the giant spiral (two are closer and two are more distant), and these too could not be resolved into stars.

Eventually Baade tried a fortunate combination: photographic emulsions sensitive to red light, a red filter, perfect atmospheric conditions, a blacked-out Los Angeles and extremely long exposure times. The resulting photographs not only resolved the stars in the central bulge of the Andromeda galaxy and in the four companion galaxies; they also led Baade to distinguish two stellar populations. The elusive stars that appeared on Baade's plates were red-giant stars, large but nonetheless too faint to be resolved by the emulsions previously employed. They belonged to a class of stars that Baade called Population II. Such stars were the same, Baade contended, as the ones that make up the aggregations in the Milky Way called globular clusters.

In number the stars of Population II predominate at the center of the Andromeda galaxy and in that galaxy's globular clusters, which are scattered throughout a spherical volume extending above and below the plane of the galactic disk. In mass (although not in luminosity) they probably predominate throughout the Andromeda galaxy. Subsequent work by Baade, his students and others showed that Population II stars are all about 12 billion years old, which makes them almost as old as the universe. In contrast, Population I includes the bright blue stars that make up the spiral arms of the Andromeda galaxy. These stars are thought to be young. Population I also includes the gas and dust that tend to enmesh bright blue stars.

Baade's second major discovery was also largely dependent on the Andromeda galaxy. The globular clusters of the Milky Way include a few Cepheid variables, and Shapley had used them to calibrate the relation of period to luminosity for the Cepheids generally. Baade noticed, however, that in the globular clusters of the Milky Way the brightest red-giant stars had about the same apparent luminosity as the Cepheid variables whose periods were from 30 to 40 days. In the Andromeda galaxy the brightest red giants were fainter than the Cepheids with those periods that did not lie in globular clusters.

The resolution of this discrepancy was announced at the 1952 meeting of



WEALTH OF OBJECTS in the Andromeda galaxy is suggested by the photograph at the top, which shows a small part of the galaxy. The map at the bottom identifies some of the objects. Both "open clusters" and "globular clusters" are aggregations of stars. In the Milky Way galaxy the former are younger and less compact than the latter, but that distinction is less clear for the clusters in the Andromeda galaxy. Each cluster in the Andromeda galaxy is apparent as simply a patch of light. "Stellar associations" are loose aggregations of very young stars. The photograph was made with the four-meter telescope at Kitt Peak National Observatory.

the International Astronomical Union in Rome, when Baade showed that the Cepheid variables in the globular clusters of the Milky Way are a special type of star found only in the company of stars of Population II. They are not the same as the Cepheid variables in the spiral arms of the Andromeda galaxy or in those of the Milky Way. The Cepheids in globular clusters are intrinsically about four times fainter than spiral-arm Cepheids with the same period of variation. Since the apparent brightness of an object varies as the square of the distance, it follows that the Cepheids studied by Hubble in the Andromeda galaxy must be twice as far away as Hubble had calculated. The putative distances of all the galaxies depended on Hubble's chain of reasoning, and so Baade's second discovery doubled the size of the universe.

Star Formation

Baade's interest in stellar populations led him to note that the gas and dust in the Andromeda galaxy seemed to be associated almost exclusively with Population I stars. This association was circumstantial evidence that the gas and dust mark areas where new stars are forming. In brief, the bright blue stars of Population I are young because a star as luminous as they are cannot have maintained that output of energy for long. Evidently, therefore, the gas and dust provide the raw material from which the stars are formed. The dust particles could be sites of accretion for the gas.

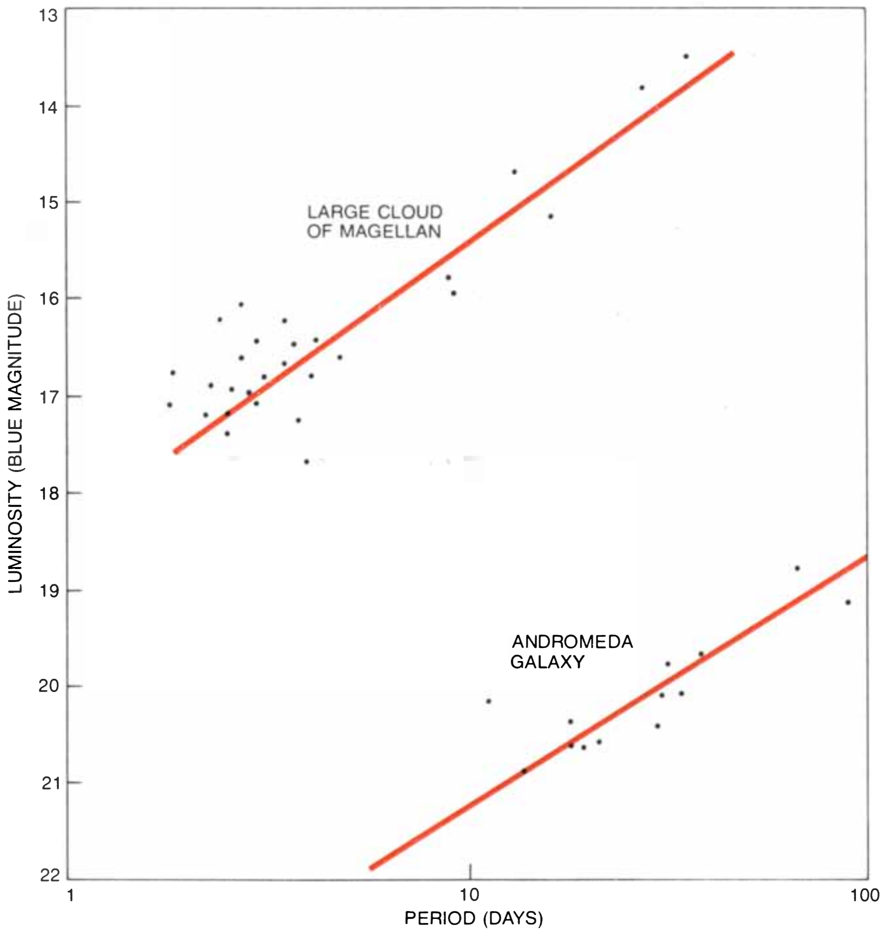
In trying to discover both how stars form and how the regions of star formation are arrayed in a spiral galaxy, Baade examined the distribution of hot gas clouds. Such clouds are illuminated by the bright, young stars embedded in them, but they absorb the light and then emit it only at certain wavelengths, which correspond in energy to excitations of the electrons in the atoms of the gas. By making photographs with filters that allowed only light of those wavelengths to pass, Baade was able to map the locations of 688 gas clouds in the Andromeda galaxy. He found that they were concentrated in the spiral arms. They were most conspicuous at the middle distance along the arms, some 30,000 to 40,000 light-years from the center of the galaxy. That is where the bright blue stars are present in the greatest number, and so it is also in such regions that the newest stars in the galaxy are now forming.

With the development of radio astronomy new methods became available for examining the formation of stars. The most important single advance was the discovery of the radio emission of neutral (that is, un-ionized) hydrogen gas at a wavelength of 21 centimeters. Each photon, or quantum, of the emission arises when the intrinsic angular momentum of the electron in a hydrogen atom inverts with respect to that of the proton in the atom. It is also necessary that the atom be in its lowest energy state, and so the hydrogen atoms emitting the radiation cannot often be in collision. (The collisions would excite the hydrogen atoms to higher energy levels.) Hence the astronomical sources of the 21-centimeter radiation are cold, thin clouds of gas.

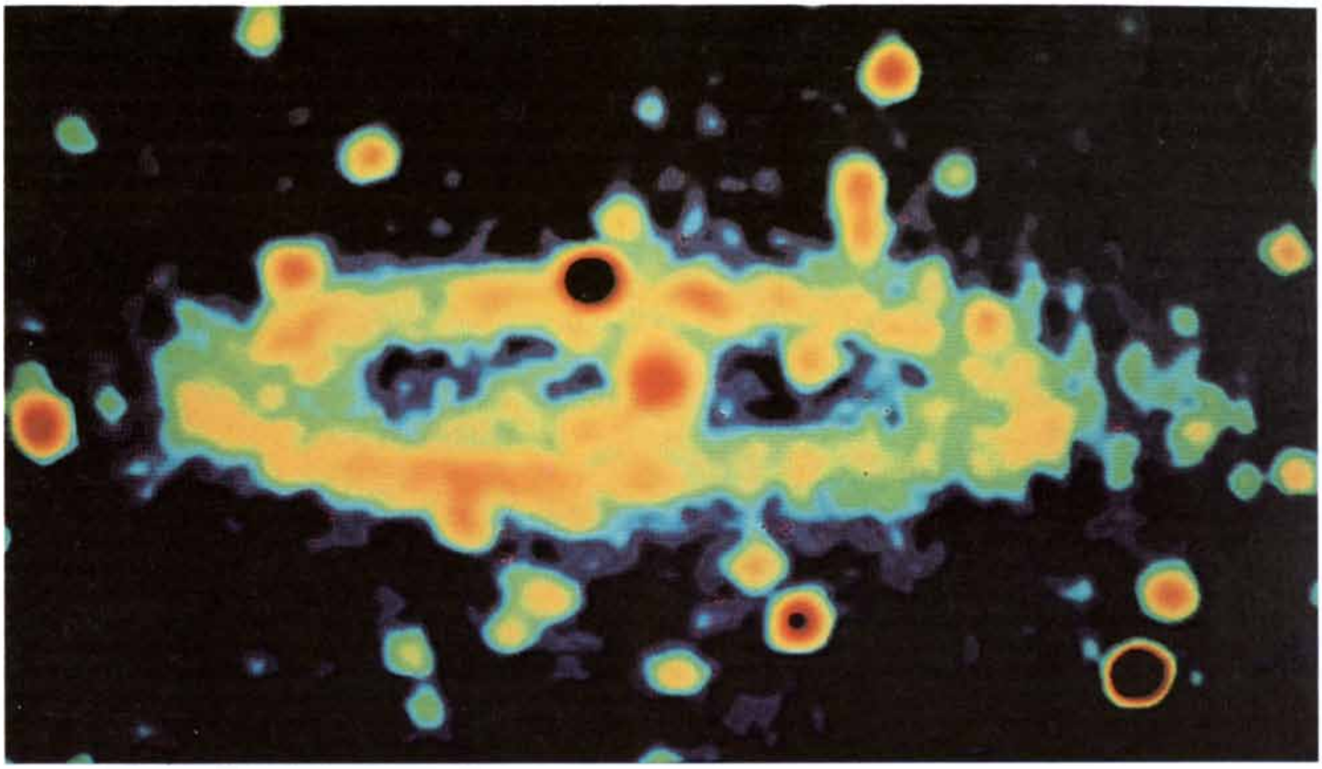
The landmark study of neutral hydrogen came in 1966, when Morton S. Roberts of the National Radio Astronomy Observatory employed the 300-foot radio telescope at Green Bank, W.Va., to amass the data for a high-resolution map of the Andromeda galaxy's radiation at 21 centimeters. Instead of a disk of gas coincident with the galactic distribution of stars, Roberts found that the distribution of neutral hydrogen resembled a giant doughnut, with a hole in the middle and a maximum density about 40,000 light-years from the center, a distance that corresponds to both the brightest parts of the spiral arms and the maximum concentration of the gas clouds mapped by Baade. To this extent the distributions of hot and cold hydrogen match.

Beyond the thickest part of the doughnut the two distributions differ. The hot gas clouds taper off at about 50,000 light-years. The cold neutral hydrogen could be detected out to twice that distance. This kind of pattern is now known to be fairly common in giant spiral galaxies, and it is understood to result from a mixture of causes. In the inner region of such a galaxy stars of Population II predominate. Star formation was complete long ago, and little gas is left. In the middle region, where gas is still abundant, passing shock waves in the density of the gas evidently trigger the condensation of new stars. In the outer region such shock waves are weak or absent, and so the density remains too low for new stars to condense.

In recent years still more has been learned about the distribution of neutral hydrogen, as a result of the development of radio telescopes in which signals



DIFFERENCE IN LUMINOSITY between the Cepheid variable stars in the Andromeda galaxy and the ones in the Large Cloud of Magellan (a small companion galaxy of the Milky Way galaxy) allows a calculation of the distance to the Andromeda galaxy. The luminosity of each such star in the Large Cloud of Magellan varies with a periodicity (*horizontal axis*) that is proportional to its brightness (*vertical axis*). The Cepheids in the Andromeda galaxy exhibit the same behavior, but the stars appear to be fainter as a result of their greater distance. In particular, a difference of six magnitudes indicates a hundredfold difference in brightness and a tenfold difference in distance. The Large Cloud of Magellan is 200,000 light-years from earth.



RADIO EMISSION of the Andromeda galaxy at a wavelength of 11 centimeters is displayed by a computer at the Rheinisches Landesmuseum at Bonn in West Germany. Red signifies the highest intensity and violet the lowest. Thus the strongest source of emission is at the center of the galaxy, a location that probably is notable for the concentration of the remnants of exploded stars. The overall symmetry of the pattern is produced by sources at a radius of about 30,000 light-years from the galactic center. That distance also characterizes the distribution of hot gas clouds, which mark regions where stars form. The data were obtained at the 100-meter radio telescope at Effelsberg in West Germany by Rainer Beck, Elly Berkhuijsen and Richard Wielebinski of Max Planck Institute for Radio Astronomy.

from several parabolic antennas are processed to yield a resolution equal to that of a telescope the size of the distance of their separation. The first studies to be completed with this new generation of instruments include the work of D. T. Emerson and his colleagues at the University of Cambridge, who employed the Cambridge Half-Mile Telescope to show that the hole in the doughnut of neutral hydrogen in the Andromeda galaxy extends to a distance of about 12,000 light-years from the galactic center. From that distance outward the distributions of hydrogen and dust seem to match closely. On the other hand, the Cambridge investigators found hydrogen some 105,000 light-years from the center of the Andromeda galaxy along the southwestern major axis of the galaxy. (The galaxy slants from northeast to southwest in the sky.) That puts it farther from the center than any part of the galaxy's image in an optical telescope.

Newest Hydrogen Maps

An even more detailed study of neutral hydrogen in the Andromeda galaxy has now been made by Estaban Bajaja of the Westerbork Observatory, who employed the Westerbork Aperture Synthesis Telescope in the Netherlands. Although not all the results have yet been published, Bajaja has found a close correspondence between hydrogen and optically visible dust only along the northeastern major axis and therefore in only half the galaxy. On the other side of the galactic center the arrangement of both gas and dust is poorly defined.

Bajaja has also investigated the motion of the neutral hydrogen. He finds that it is deviating from a strictly circular trajectory. The simplest hypothesis about a spiral galaxy's motion is that all its components rotate about the massive galactic center. The stars of Population II are thought to have formed before much of the galaxy had condensed into a disk. Their orbits are highly elliptical. In contrast the orbits of young stars and of gas and dust in the disk of Population I are thought to be almost circular, like the orbits of planets around the sun.

The new discoveries clearly show that this assumption is sometimes quite wrong. Of the three apparent arms of hydrogen in the northeast half of the Andromeda galaxy, parts of the innermost arm are plummeting inward toward the center of the galaxy with a velocity of at least 100 kilometers per second. This is in addition to their motion around the center. The cause of the inward velocity is a mystery. The next arm out does not show it. Is the noncircular motion related to the complex dynamics of a warping of the galactic disk? Is it caused by the companion galaxies, or has some explosive event in the past disrupted this part of the galaxy? The answer probably lies in the increasingly detailed data the radio-telescope arrays are now gathering.

While the instrument at Westerbork has been mapping the radio emission of neutral hydrogen at 21 centimeters, a 100-meter single-dish radio telescope near Bonn in West Germany has enabled Elly Berkhuijsen and her colleagues at the Max Planck Institute for Radio Astronomy to map the emission at 11 centimeters and deduce the nature of its sources. Earlier G. G. Pooley of Cambridge had shown that most of the Andromeda galaxy's 11-centimeter radio emission came from the center of the galaxy and also from the parts of the spiral arms that are bright in optical images.

Piet van der Kruit of the Leiden Observatory and Yervant Terzian and Bruce Balick and their co-workers at Cornell University confirmed Pooley's results at other wavelengths. They also demonstrated from the relative intensities of the radiation at those wavelengths that a large proportion of the radiation could not be thermal. A source of thermal radiation, such as a cloud of hot gas, emits radio noise whose spectrum has a characteristic rise

INTRODUCING

THE NEW WORLD CAR FORD ESCORT

ENGINEERING TEAMS FROM AMERICA AND OVERSEAS JOIN FORCES TO CREATE A NEW CAR WITH BETTER IDEAS FROM AROUND THE WORLD

Top engineers drawn from the worldwide resources of Ford Motor Company teamed up to create Escort. They pooled their expertise... compared, tested, evaluated, experimented—to come up with better ideas.

The result: a high-mileage car built in America to take on the world. Escort will be made in America for American drivers...with other models built and sold overseas.

HIGH MILEAGE THROUGH ADVANCED TECHNOLOGY

Escort has higher gas mileage ratings than subcompacts such as VW Rabbit, Honda Accord, Toyota Corolla hatchback—yet Escort has more room—in fact, Escort has the room of a compact (based on EPA Volume Index).

30 EPA EST MPG* **44** EST HWY*

*Applicable only to sedans without power steering or A/C. For comparison. Your mileage may differ depending on speed, distance, weather. Actual hwy mileage and Calif. ratings lower. Excludes diesels.

ONE OF THE MOST POWER-EFFICIENT ENGINES AVAILABLE IN AMERICA

Escort's new-design compound valve hemispherical head (CVH)

engine gives you high mileage plus power for freeway driving.

FRONT-WHEEL DRIVE AND FOUR-WHEEL INDEPENDENT SUSPENSION

Escort comes with front-wheel drive...four-wheel fully independent suspension...rack-and-pinion steering...stabilizer bar...new all-season steel-belted radials, and other road-control features.

DRIVE A WORLD CAR

Whether you buy or lease, see your Ford Dealer now to order your Escort...3-door Hatchback, 4-door Liftgate...or even a Squire wagon option. And remember to ask about Ford's Extended Service Plan.

Escort Liftgate...4-doors and wagon room when you need it.

BUILT TO TAKE ON THE WORLD

FORD ESCORT

FORD DIVISION Ford

and fall determined by the cloud's size and temperature. The intensities recorded by the Cornell investigators do not conform to that shape. Instead much of the radiation must be generated by some nonthermal mechanism. It might, for example, be synchrotron radiation, which is given off by electrons moving at nearly the speed of light through a magnetic field. Synchrotron radiation includes electromagnetic waves with almost equal intensity from X rays to long radio waves. It arises in places where large amounts of energy are released, such as the vicinity of supernova explosions and the collapsed supermassive objects (neutron stars or black holes) the explosions are thought to leave behind.

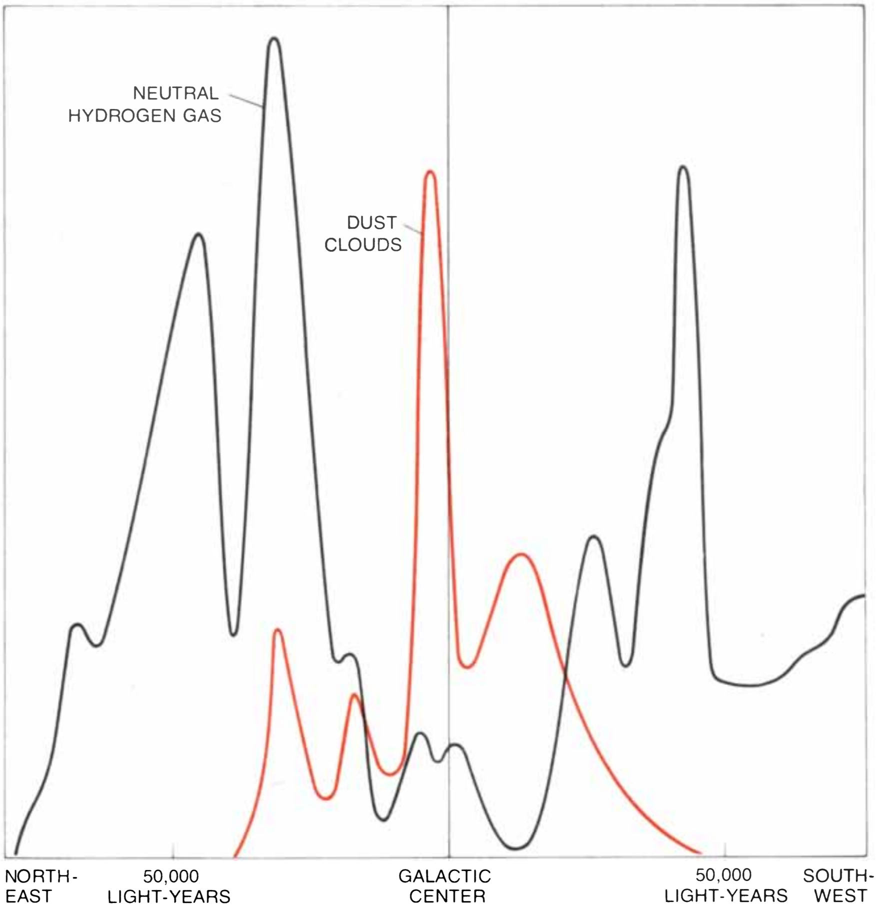
Berkhuijsen's maps showed that thermal radiation comes mostly from the vicinity of the doughnut of neutral hydrogen, about 30,000 light-years from the center of the Andromeda galaxy. It is therefore probably given off by the numerous hot gas clouds that are commonest there. The nonthermal radiation is more broadly based. It has a sharply peaked source at the center of the galaxy, but it extends more or less evenly to a distance from the center of from 40,000 to 50,000 light-years. Supernova remnants in the Milky Way galaxy (and the seven remnants discovered so far in the Andromeda galaxy by Vera C. Rubin, Cidambi K. Kumar and W. Kent Ford, Jr., of the Department of Terrestrial Magnetism of the Carnegie Institution of Washington) have a similar distribution. Thus the nonthermal radiation may be explained largely as the combined noise from the remnants of supernovas, perhaps including the spectacular S Andromeda of 1885.

Other Objects

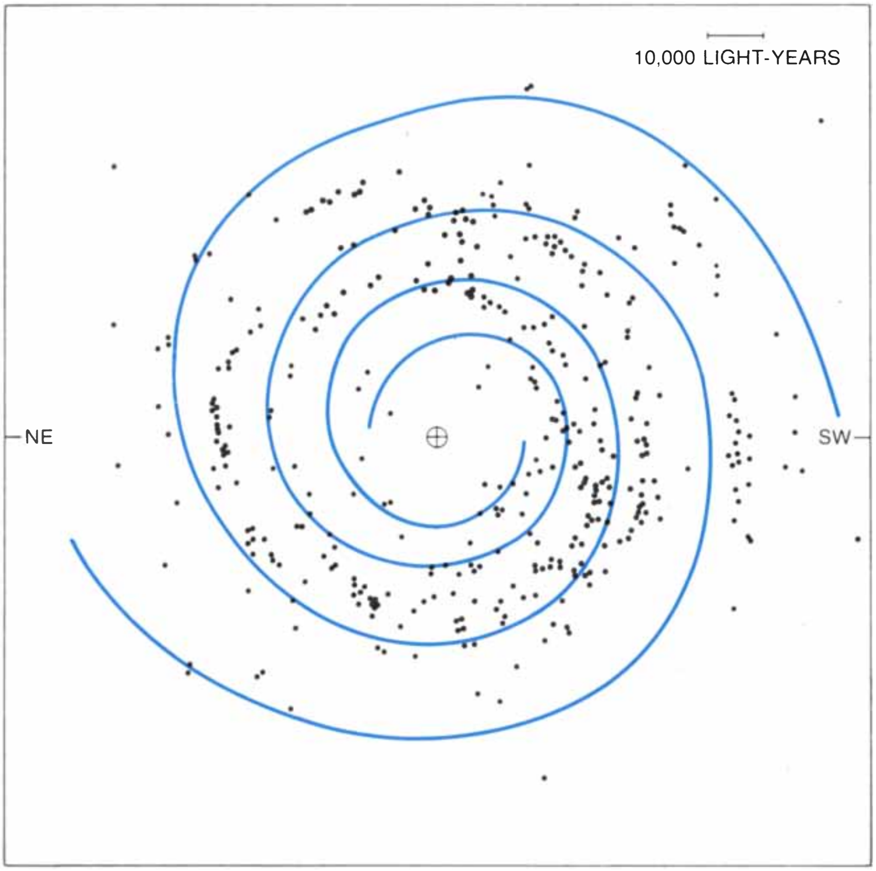
A kind of bright gas cloud that resembles a smoke ring is thought to be the shed outer atmosphere of an aging or dying star. The ring is called a planetary nebula because although it is nebulous, its image resembles that of a planet in that the image is compact and disklike. Although Baade found five such objects in the Andromeda galaxy, they have been discovered there in large numbers only in the past two or three years. Holland Ford and George Jacoby of the University of California at Los Angeles have exploited new techniques of filtering and image intensification to record images that would have been too faint for Baade to record. With the 120-inch telescope of the Lick Observatory they found 315 planetary nebulas. They calculate that the total number in the Andromeda galaxy is roughly 10,000. Since planetary nebulas mark dying stars, such nebulas map the galaxy's decay. In the bulge at the center of the galaxy about five stars per century are apparently dying. Over the past few billion years the gas released in this way should have accumulated in a gaseous disk that rotates about the bulge. Radio astronomers have found a central disk of gas whose mass is just about the expected one.

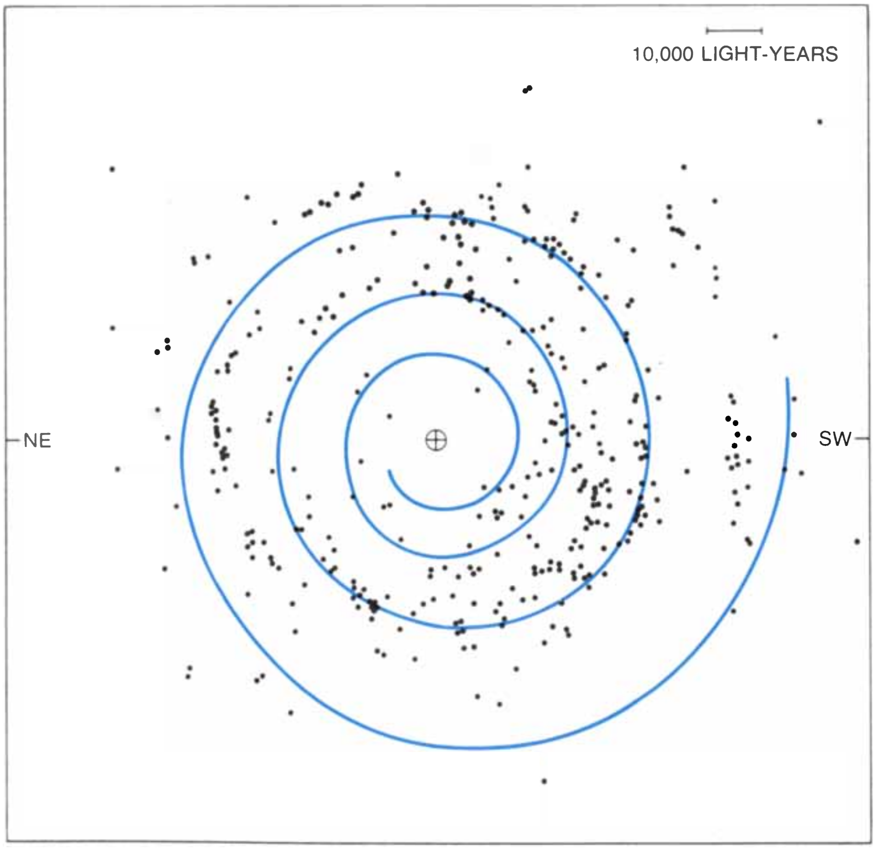
A similar application of modern image intensifiers and filters has led to progress in the exploration of the hot gas clouds in the Andromeda galaxy. A survey much superior to Baade's effort of the 1950's was recently completed by a French group led by G. Courtés at the two-meter telescope of the Haute-Provence Observatory. The measured velocities of the clouds, combined with the velocities of neutral hydrogen, now provide a clear picture of the rotation of the galaxy. Moreover, the velocity of an object in orbit depends on the amount of mass inside the orbit as well as on the mass near the orbiting object, so that from the orbital speeds of objects in the Andromeda galaxy it is possible to deduce the distribution of mass in the galaxy. The most recent attempts to deduce the total mass yield values between 200 billion and 400 billion times the mass of the sun, or as much as twice the mass of the Milky Way galaxy. Even so, the orbital velocity of hydrogen in the outer parts of the Andromeda galaxy hints that large amounts of unseen matter may form a huge halo.

One way to test the current value of the mass of the Andromeda galaxy would be to measure the gravitational interaction of the galaxy's mass with some other objects, preferably objects at great distances from the galactic center. Realizing that the globular clusters in the galaxy's Population II might be the needed objects, F. D. A. Hartwick of the University of Victoria, Sidney van den Bergh of the Herzberg Institute of Astronomy in Victoria and Wallace L. W. Sargent of the California Institute of Technology recently collaborated on a search for the globular clusters on photographic plates made with the new four-meter telescope of the Kitt Peak National Observatory in Arizona. They have catalogued 355 probable globular clusters, doubling the number that had previously been discovered by astronomers beginning with Hubble. The 355 globular clusters are almost three times as many as are known to lie in Population II of the Milky Way galaxy. The abundance of globular clusters is evidence in itself that the mass of the Andromeda galaxy is great. The orbits of the clusters have not yet been deter-



DUST AND GAS DISTRIBUTIONS along the major axis of the Andromeda galaxy are compared. The dust particles provide sites on which the gas can accrete, and so it is thought that both types of matter are required for the formation of stars. The distributions agree moderately well toward the northeast side of the galactic disk but not at all toward the southwest side.





SPIRAL-ARM CONTROVERSY arose when Halton C. Arp of the Hale Observatories saw in the Andromeda galaxy a two-armed spiral form (*top*), whereas Agris Kalnajs of the Mount Stromlo Observatory in Australia discerned a one-armed pattern (*bottom*), with the arm spreading outward in the opposite direction. (The rotation of the galaxy is counterclockwise when the galactic disk is viewed, as it is here, from below.) The dots in the illustration mark positions of open clusters, which consist of young stars. It is young stars that form the spiral-arm segments.

mined, however, and so the question of the total mass remains open.

It is known from a study of the spectra and the colors of the brighter globular clusters by van den Bergh that the early history of the Andromeda galaxy must have been puzzlingly different from the history of the Milky Way galaxy. In the Milky Way galaxy the globular clusters in the outer regions are very poor in heavy elements. The sun is much richer. This difference is usually interpreted as part of a general pattern. The ancient stars of Population II are almost purely hydrogen and helium, the elements created in abundance in the early universe, whereas the younger disk of Population I stars includes debris from a multitude of dying stars, which synthesized heavy elements by thermonuclear fusion. Surprisingly, the spectra of the globular clusters in the Andromeda galaxy reveal a diversity of patterns of heavy-element abundance, and there is no relation at all between the abundance of heavy elements in a cluster and the position of the cluster in the galaxy.

The Spiral-Arm Controversy

Clusters of another kind have also been found in the Andromeda galaxy. They are open clusters, or loose aggregations of stars. Open clusters are younger than globular clusters, and they lie in the plane of the galaxy, along with the rest of Population I. Except for a couple of examples Hubble identified they had not been found in the Andromeda galaxy. Last year, however, I took advantage of the wide field of view of the new four-meter telescope at the Kitt Peak National Observatory to survey the galaxy and identify 403 of them.

Most of the 403 are only 60 light-years or so across. They have nonetheless allowed me to begin to trace the recent history of star formation in the Andromeda galaxy. The crucial point is the likelihood that the stars of an open cluster all form simultaneously. On the other hand, the distribution of various types of stars in the cluster changes over a period of time. Thus a particular set of statistics implies a particular age. In this way I have shown that the rate of star formation has varied across the disk of the Andromeda galaxy. Recently it has been anomalously great about 30,000 light-years from the center, where neutral hydrogen and bright stars are concentrated.

The open clusters are now being invoked in the controversy over the configuration of the spiral arms of the Andromeda galaxy. The first attempt to delineate the arms was that of Halton C. Arp of the Hale Observatories, who based his work on the distribution of gas clouds in the galaxy. Arp found that a two-armed spiral with the arms trailing the direction in which the galaxy rotates fitted the scatter of data points fairly well. He suggested that the imperfections of the fit might be due to the gravitational disturbance of the galactic disk by the mass of M32, one of the four companions of the Andromeda galaxy.

A recent analysis done with the aid of a computer by Gene G. Byrd of the University of Alabama seems to confirm this and at the same time seems to account for the infalling observed in the distribution of neutral hydrogen. On the other hand, Agris Kalnajs of the Mount Stromlo Observatory in Australia has concluded from the distribution of clouds of hydrogen that the galaxy has only one spiral arm and that it spirals in the direction opposite to that of the two arms plotted by Arp. It is thus a leading arm: its free end points in the direction of the galaxy's rotation.

There is no unambiguous instance of a leading one-armed spiral galaxy anywhere else in the universe. Kalnajs has noted, however, that the presence of M32 may be the explanation. If the period of rotation of the Andromeda galaxy and the period of rotation of M32 about the Andromeda galaxy have a common divisor, a gravitational resonance between the two galaxies could give rise to the leading arm. Recently a group of French, Swiss and Greek astronomers examined a large collection of positional data for Population I objects and attempted to discern a spiral structure. They too concluded that the best fit was that of a leading one-armed spiral.

The controversy remains unresolved. The distribution of gas clouds seems to fit a one-armed pattern fairly well. It turns out, however, that the distribution of open clusters fits only a two-armed spiral that is gravitationally warped. Moreover, the dust clouds fit no spiral pattern well. The discrepancies show how imperfect most galactic spiral patterns are, and also how imperfect our understanding of galactic spirals is.

The Andromeda galaxy continues to be the subject both of observations made possible by new techniques and of searches for instances of newly discovered types of object. Maps now plot the distribution of carbon monoxide in the Andromeda galaxy; they show that it is closely correlated with the neutral hydrogen. Moreover, just last year the Einstein Orbiting X-Ray Observatory detected 69 X-ray sources in the Andromeda galaxy. Astronomers at the Center for Astrophysics of the Harvard College Observatory and at the Smithsonian Astrophysical Observatory think some of the sources are globular clusters, some are concentrations of Population I objects and some are supernova remnants. The precise nature of the high-energy processes that give rise to the radiation may soon be deduced. The Andromeda galaxy has provided astronomers with a rich lode of information. It will no doubt reveal much more.

The Wild Gene Resources of Wheat

Modern plant-breeding practices have reduced the genetic variability of the cultivated wheats. The best hope for future crop improvements lies in exploiting the abundant gene pool of the plant's wild relatives

by Moshe Feldman and Ernest R. Sears

World wheat production has increased dramatically during the past few decades. There is no question, however, that it will have to be raised still higher to help feed the world's growing population. Since new arable croplands are not likely to be made available on a large enough scale to do the job, and since the application of such energy-intensive agricultural aids as fertilizers, pesticides and herbicides is becoming prohibitively expensive, the next big increment in production will have to be achieved mainly by further improvements in the productivity of wheat. Accordingly there is an urgent need for the development of new varieties of wheat with either a greater yield or a superior ability to grow in places where wheat is not grown today.

Two factors limit this effort. The first is that the genetic material of the cultivated wheats has already been exploited for breeding purposes almost to its full capacity. The second is that the range of genetic variation of the cultivated wheats has decreased drastically in recent years. The continuing rapid erosion of the gene pool of wheat not only reduces the possibility of further improvements in productivity but also makes the world wheat crop increasingly vulnerable to new diseases and to adverse climatic changes.

The genetic variability of the cultivated wheats, which had accumulated over 10,000 years of cultivation, has been diminished by the introduction of modern, scientifically planned breeding practices. (The same practices are of course largely responsible for the present high productivity of wheat.) For several decades new, improved varieties of wheat have been selected at the expense of the overall genetic variability of the world wheat crop. The new varieties, each consisting of a single genotype, have steadily replaced the traditional "land races," each of which comprised many different genotypes.

The failure to conserve the primitive cultivated varieties of wheat has already resulted in the loss of a substantial reserve of genetic variability. Attempts to increase the variability of the new cultivated wheats by inducing mutations, either by ionizing radiation such as X rays or by chemical treatment, have met with little success. Conservation of the germ plasm of the surviving primitive cultivated wheats can lessen the danger of further genetic erosion. On a large scale, however, the restoration and enrichment of the gene pool of the cultivated wheats can be accomplished only by tapping the vast genetic resources that are to be found in the wild relatives of the wheats.

The cultivated wheats belong to four species of the genus *Triticum,* all of which contain some multiple of the basic haploid (single) set of seven chromosomes per gamete, or reproductive cell. The species with the lowest chromosome number, *T. monococcum,* is normally diploid, that is, it has two times seven, or 14, chromosomes per somatic, or body, cell. (It is equivalent to say that such a species has seven chromosome pairs.) Two of the other species, *T. turgidum* and *T. timopheevii,* are tetraploid: they have 28 chromosomes (four times seven, or 14 pairs) each. The cultivated *Triticum* species with the highest chromosome number, *T. aestivum,* is hexaploid: it has a total of 42 chromosomes (six times seven, or 21 pairs).

Of the four cultivated wheats *T. aestivum* is economically by far the most important. Most of its modern varieties fall under the general heading of common wheat, the flour of which is best suited for bread. *T. aestivum* encompasses more than 20,000 cultivars (cultivated varieties), which are adapted to a wide range of environments.

The cultivation of *T. aestivum* is worldwide. Similarly, the tetraploid wheat *T. turgidum,* particularly its main modern variety *T. turgidum durum* (durum or macaroni wheat), is grown widely in comparatively dry parts of the world, such as the Mediterranean basin, India, the U.S.S.R. and the low-rainfall areas of the North American Great Plains. In contrast, the cultivation of the tetraploid *T. timopheevii* and the diploid *T. monococcum* is quite limited: *T. timopheevii* is grown today only in a few places in the Transcaucasian region of the U.S.S.R., and *T. monococcum* is grown only in some mountainous regions of Yugoslavia and Turkey, where it is used mainly for animal fodder.

The wild relatives of wheat that can be crossed with these cultivated wheats and hence can serve as potential donors of desirable characters (heritable traits) belong to the Triticeae tribe of the Gramineae, or grass family. This tribe, which is economically the most important division of the grasses, has given rise not only to the cultivated wheats but also to rye (*Secale*) and barley (*Hordeum*). The Triticeae tribe includes a total of 14 genera, which on the basis of their morphological characteristics are grouped into two subtribes, the Triticinae and the Hordeinae [*see bottom illustration on page 105*].

Within both subtribes members of different genera can occasionally form hybrids, facilitating gene flow either through the process known as crossing-over (the exchange of genetic material between homologous, or very closely related, chromosomes) or through the creation of an amphiploid species, that is, a fertile interspecific hybrid with a complete set of paired chromosomes derived from each parent species. For example, within the subtribe Triticinae not only the various species of Triticum but also species of certain other genera can be crossed with common wheat or durum wheat to yield viable hybrids. Recently several successful attempts were reported of even more distant hybridizations, involving species of the Hordeinae subtribe and wheat. These findings indicate that the genetic variation of the entire Triticeae tribe is potentially exploitable for the improvement of wheat.

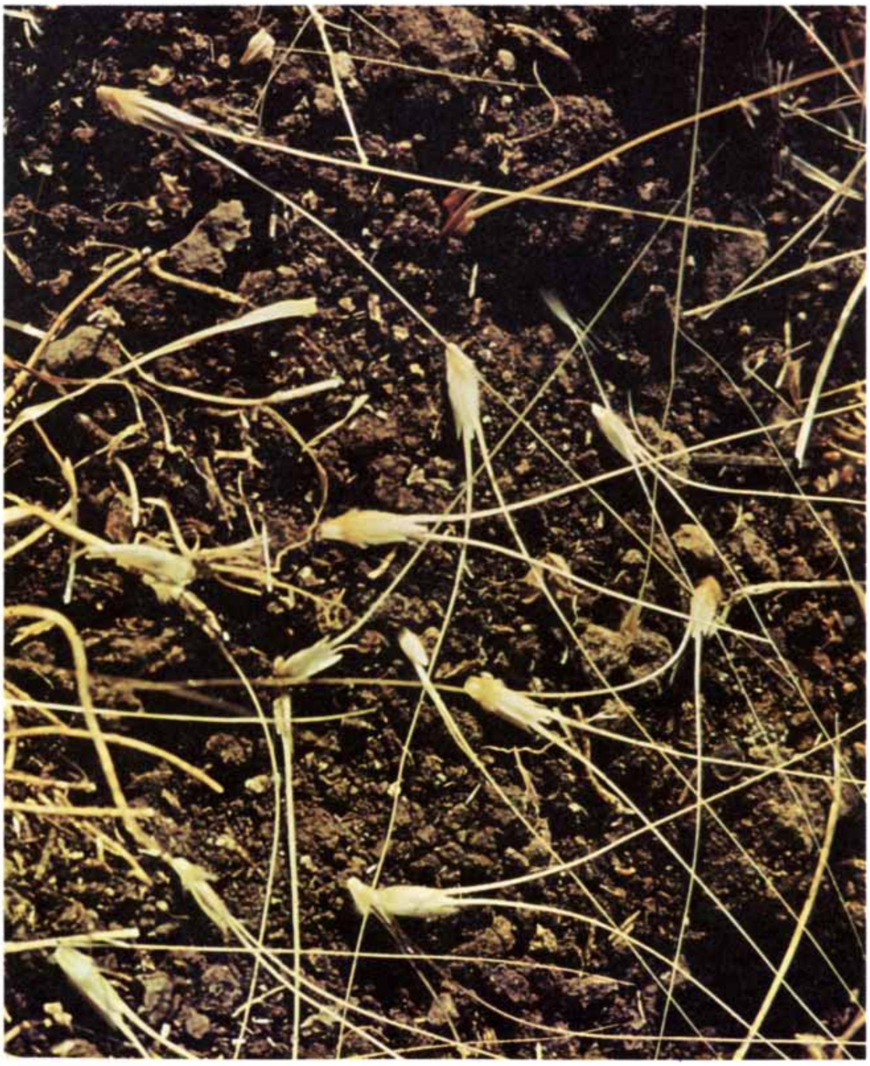
A large store of basic biological information is already available on some of the wild relatives of wheat, particularly those belonging to the genus

WILD WHEAT belonging to the variety *Triticum turgidum dicoccoides* (also known as wild emmer) is seen growing alongside a road near the Sea of Galilee in northern Israel in this photograph made by Shalom Nedam of the Weizmann Institute of Science. The individual spikes of the wild plants vary greatly in color, ranging almost from black to white. Differences in stature are also apparent. A primitive cultivated variety of this species, *T. turgidum dicoccum*, is thought to be one of two progenitors of the common bread wheat *T. aestivum*.

CULTIVATED WHEAT, one of the more than 20,000 modern cultivars of the species *T. aestivum*, was photographed by Nedam in a field in the Yavne'el Valley to the east of the Sea of Galilee. Little or no variation in either color or stature is evident from plant to plant.

Triticum. The various wild species of *Triticum* belong to the same three ploidy levels as the cultivated species: diploid (with seven chromosome pairs), tetraploid (with 14 pairs) and hexaploid (with 21). Although the wild diploid species are presumably descended from a common ancestor, they have diverged considerably from one another not only morphologically but also ecologically and geographically. Chromosome studies have corroborated the taxonomic classification by showing that almost every diploid species has a distinct complement of genes—a distinct genome. The related chromosomes of the different genomes exhibit varying degrees of reduced affinity for one another, and hence they do not pair regularly when the cells formed by an interspecific hybridization undergo meiosis (the type of cell division that reduces the number of chromosomes in the gametes of sexually reproducing organisms). This factor tends to make such hybrids sterile and leads ultimately to the genetic isolation of the various diploid species from one another.

The polyploid species of *Triticum* (those with ploidy levels higher than two) constitute a classic example of evolution through amphiploidy. Their behavior is typical of the class of hybrids known to geneticists as genomic allopolyploids: their chromosomes pair in a diploidlike manner (forming only bivalent, or paired, chromosomes at meiosis), and their mode of inheritance is disomic (through the pairing of essentially identical chromosomes). The allopolyploid nature of the *Triticum* polyploids has been verified by various investigators through the analysis of the genomes of hybrids formed between species of ploidy levels. For example, each tetraploid species can be identified as the product of the hybridization of two diploid species, followed by a spontaneous doubling of the chromosomes. Similarly, each hexaploid species is a product of the hybridization, followed by a chromosome doubling, of a tetraploid species and a diploid one.

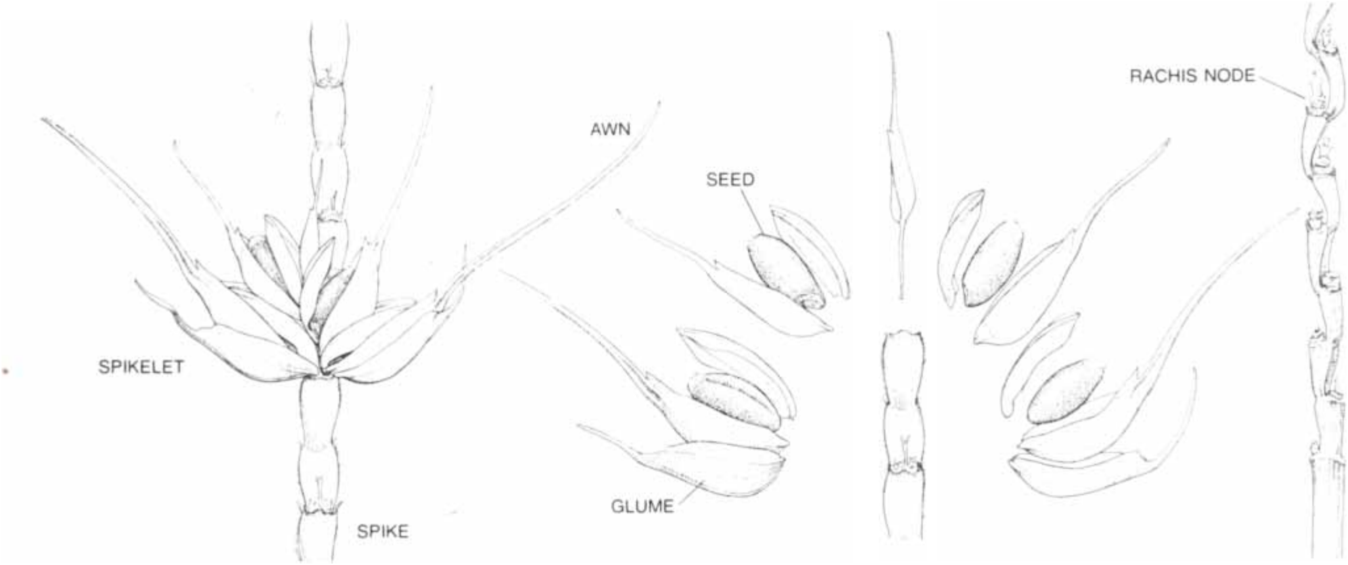


SPIKELETS OF WILD WHEAT lie strewn on the ground following the fragmentation of the spikes at maturity. The main difference between the wild wheats (represented here by *T. turgidum dicoccoides*) and their cultivated derivatives is that on ripening the spikes of the wild species tend to break into their component spikelets and fall to the ground. Some of the spikelets seen in this photograph have partially buried themselves in the soil, through the action of their awns (their elongated, bristlelike terminal processes) in response to wetting; the enclosed seeds are thereby ensured favorable conditions for germination and survival. Although shattering at maturity is essential to the wild wheat as its primary seed-dispersal mechanism, it must have been a major nuisance to the ancient harvesters of wheat. It is therefore not surprising that the early wheat growers selected nonshattering mutants when these appeared, thereby converting the wild wheat into a cultivated one no longer capable of surviving without man's help.

Three groups of polyploids are recognized in the genus *Triticum.* The species in each group have one genome in common and differ in their other genome (or genomes). The polyploids in group *A,* for example, share the genome of their common diploid ancestor *T. monococcum,* those in group *D* share the genome of the wild grass *T. tauschii* and those in group *U* share the genome of another wild grass, *T. umbellulatum.* In basic morphology and particularly in the structure of the seed-dispersal mechanism the polyploids in each group resemble the diploid donor of the common pivotal genome.

This characteristic genetic structure (one common genome and one or two different ones) is what accounts for the comparatively high rate of successful spontaneous hybridization (and hence gene flow) between the polyploids. Hybridizations of this type are facilitated by the shared genome, which acts as a buffer, ensuring some fertility in the resulting hybrids. In such cases the different genomes, which are brought together from different parents, can exchange genetic material and form a new, mixed genome. Thus in each group of polyploid species there is an almost continuous variation between the characteristics of the different species.

The genetic structure of the polyploid species of *Triticum* allows for a rapid buildup of genetic diversity. In such species most of the gene loci are present in four or six doses (duplicate versions), and the accumulation of genetic variation through mutation or hybridization is tolerated more readily than it is in a diploid organism. Thus polyploidy facilitates genetic diploidization, the process whereby existing genes in multiple even-numbered doses can be diverted to new functions. Because the *Triticum* polyploids are true allopolyploids, they show permanent heterosis ("hybrid vigor"), an effect that results from the interaction of certain genes at homologous gene loci in the different genomes.



ANATOMY OF A SPIKELET of common wheat is detailed in this set of drawings. The spike is composed of alternating rachis nodes, each of which bears a single spikelet (*left*). Dissection of the spikelet shows that it contains several seeds, or grains (*center*). The alternation of the rachis nodes is clearer in a side view (*right*). In some varieties the awns extending from the ends of the glumes are absent.

Both the wild and the cultivated polyploid species of the genus *Triticum* are characterized by convergent evolution (that is, the adaptive evolution of superficially similar structures in different species subjected to similar environments) by virtue of the fact that they contain genetic material from two or three different diploid genomes and can exchange genes with one another through hybridization. The evolutionary significance of polyploidy in this context arises mainly from the fact that it has facilitated the formation of a genetic superstructure that brings together the genes of the isolated diploid species and allows them to recombine. Moreover, polyploidy, reinforced by a mating system in which self-pollination is predominant, has proved to be a very successful genetic system. The evolutionary advantage of the polyploids over the diploids is reflected in the wide morphological and ecological variation of the polyploids. No wonder, then, that the cultivated polyploid wheats exhibit a wide range of genetic flexibility and

				GROWTH HABIT			MODE OF POLLINATION				
SUBTRIBE	GENUS	NUMBER OF SPECIES	PLOIDY LEVEL (X = 7)	PERENNIAL	PERENNIAL, ANNUAL	ANNUAL	CROSS	CROSS, SELF	SELF	DISTRIBUTION OF WILD SPECIES	SUCCESSFUL HYBRIDS WITH TRITICUM
HORDEINAE	*Hordeum*	25	2X–6X		■			■		WORLDWIDE (ARCTIC-TEMPERATE)	■
	Elymus	60	2X–12X	■				■		WORLDWIDE (ARCTIC-TEMPERATE)	■
	Asperella	7		■			■			WORLDWIDE (ARCTIC-TEMPERATE)	
	Sitanion	1	4X	■			■			NORTH AMERICA (ARCTIC-TEMPERATE)	
	Psathyrostachys	6		■			■			MEDITERRANEAN BASIN, CENTRAL ASIA	
	Crithopsis	1	2X			■			■	MEDITERANNEAN BASIN, CENTRAL ASIA	
	Taeniatherum	2	2X			■			■	MEDITERANNEAN BASIN, CENTRAL ASIA	■
TRITICINAE	*Agropyrum*	100	2X–10X	■			■			WORLDWIDE (ARCTIC-TEMPERATE)	■
	Haynaldia	2	2X, 4X		■		■			MEDITERANNEAN BASIN, CENTRAL ASIA	■
	Secale	6	2X		■		■			MEDITERANNEAN BASIN, CENTRAL ASIA	■
	Heteranthelium	1	2X			■			■	MEDITERANNEAN BASIN, CENTRAL ASIA	
	Henrardia	2	2X			■			■	MEDITERANNEAN BASIN, CENTRAL ASIA	
	Eremopyrum	5	2X, 4X			■			■	MEDITERANNEAN BASIN, CENTRAL ASIA	■
	Triticum	27	2X–6X			■		■	■	MEDITERANNEAN BASIN, CENTRAL ASIA	■

FOURTEEN GENERA of the tribe Triticeae, a major division of the Gramineae, or grass family, are grouped in this table on the basis of their morphological characteristics into two subtribes: the Hordeinae and the Triticinae. The members of the subtribe Hordeinae, such as barley (*Hordeum*), are distinguished by having two or three spikelets per rachis node; those of the subtribe Triticinae, such as rye (*Secale*) and wheat (*Triticum*), typically have only one spikelet per rachis node. The number of species in each genus is listed, along with the corresponding ploidy level, that is, the number of chromosomes in each somatic, or body, cell expressed as a multiple of the haploid (single) chromosome number (*X*) associated with a gamete, or reproductive cell. (All members of the tribe Triticeae have a basic haploid chromosome number of seven.) The characteristic growth habit, predominant mode of pollination and geographic distribution are also given for each genus. The last column at the right indicates genera that are known to form viable hybrids with species of *Triticum*.

have adapted to a great variety of environments.

The center of origin and the main center of diversity of the genus *Triticum* is southwestern Asia, particularly the open oak-park belts of the hilly country of the Fertile Crescent, extending from the Mediterranean coast in the west around the Syrian Desert to the Tigris-Euphrates plain in the east. In this region there is a rich concentration of *Triticum* species, both diploids and polyploids, each exhibiting a considerable range of morphological and ecological variation. In many cases the species grow in polymorphic, or mixed, populations, a circumstance that increases the frequency of spontaneous hybridization and facilitates interspecific as well as intraspecific gene flow. Such dynamic genetic interconnections lead to the buildup of a wealth of genetic variation and make the region an active center of evolution.

To repeat, the cultivated polyploid wheats consist of three different species: two tetraploids, *T. turgidum* and *T. timo-*

	SPECIES	GENOMIC FORMULA	DISTRIBUTION
DIPLOIDS	*T. monococcum boeoticum*	AA	S.W. ASIA, BALKANS
	T. monococcum urartu	AA	S.W. ASIA
	T. dichasians	CC	S.W. ASIA, GREECE
	T. tauschii	DD	S.W. ASIA, CENTRAL ASIA
	T. comosum	MM	GREECE, TURKEY
	T. tripsacoides	$MtMt$	TURKEY
	T. uniaristatum	$MuMu$	BALKANS, TURKEY
	T. speltoides	SS	S.W. ASIA
	T. bicorne	S^bS^b	S.W. ASIA, N. AFRICA
	T. sharonensis	S^lS^l	S.W. ASIA
	T. longissimum	S^lS^l	S.W. ASIA, EGYPT
	T. searsii	S^sS^s	S.W. ASIA
	T. umbellulatum	UU	S.W. ASIA, (GREECE ?)
POLYPLOIDS SHARING *AA* GENOME	*T. turgidum dicoccoides*	$AABB$	S.W. ASIA
	T. timopheevii araraticum	$AAGG$	S.W. ASIA
	T. aestivum	$AABBDD$	WORLDWIDE (CULTIVATED)
POLYPLOIDS SHARING *DD* GENOME	*T. cylindricum*	$DDCC$	S.W. ASIA, BALKANS, CENTRAL ASIA
	T. crassum	$DDM^{cr}M^{cr}$	S.W. ASIA, CENTRAL ASIA
	T. crassum	$DDD_2D_2M^{cr}M^{cr}$	CENTRAL ASIA
	T. syriacum	$DDM^{cr}M^{cr}S^lS^l$	S.W. ASIA
	T. juvenale	$DDM^{cr}M^{cr}UU$	S.W. ASIA, CENTRAL ASIA
	T. ventricosum	DDM^vM^v	S. EUROPE, N. AFRICA
POLYPLOIDS SHARING *UU* GENOME	*T. triunciale*	$UUCC$	S.W. ASIA, S. EUROPE
	T. macrochaetum	UUM^bM^b	S.W. ASIA, S. EUROPE, U.S.S.R.
	T. columnare	UUM^cM^c	S.W. ASIA
	T. triaristatum	UUM^tM^t	S.W. ASIA, S. EUROPE, N. AFRICA
	T. triaristatum	$UUM^tM^tM^{t2}M^{t2}$	S.W. ASIA, S. EUROPE, N. AFRICA
	T. ovatum	UUM^oM^o	S.W. ASIA, CENTRAL ASIA, S. EUROPE, N. AFRICA
	T. variabile	UUS^vS^v	S.W. ASIA, N. AFRICA, ITALY, GREECE
	T. kotschyi	UUS^vS^v	S.W. ASIA, N. AFRICA

SPECIES OF TRITICUM are listed in this table, together with their genomic formulas and the distribution of their wild species. As a rule only the wild forms are included, except for the hexaploid wheat *T. aestivum*, which has no generally recognized wild variety. The colored matrix at the right ranks the wild gene pools of the *Triticum* species according to their degree of phylogenetic relatedness to common wheat (*solid color*). Among the closely related species with genomes homeologous (similar but not homologous, or identical) to that of

pheevii, and one hexaploid, *T. aestivum*. How and where did these species originate? Intensive cytogenetic studies have suggested that *T. turgidum* (genome *AABB*) was derived from a cross between *T. monococcum* (genome *AA*) and *T. searsii* (genome S^sS^s) or a closely related wild species, whereas the other tetraploid species, *T. timopheevii* (genome *AAGG*), may have come from a cross between *T. monococcum* and *T. speltoides* (genome *SS*). Hexaploid wheat (genome *AABBDD*), on the other hand, clearly originated as a hybrid between *T. turgidum* (most likely the cultivated variety *T. turgidum dicoccum*, or emmer wheat) and *T. tauschii* (genome *DD*).

Archaeological evidence has shown that the domestication of the wheats by Neolithic man took place within the geographical distribution center of the wild progenitors of the wheats. The northeastern part of the Fertile Crescent (specifically the area encompassing southwestern Iran, northwestern Iraq and southeastern Turkey) is the center of distribution of two wild diploid wheats, *T. monococcum boeoticum* and *T. monococcum urartu*, and the wild tetraploid wheat *T. timopheevii araraticum*. It is also where the cultivated varieties of both species, respectively *T. monococcum monococcum* and *T. timopheevii timopheevii*, evolved and were first domesticated. Similarly, the southwestern part of the Fertile Crescent (northwestern Israel, southwestern Syria and southeastern Lebanon) is the center of distribution of the wild tetraploid wheat *T. turgidum dicoccoides* and also the area where the primitive cultivated variety of this species, *T. turgidum dicoccum*, originated. A spontaneous hybridization between *T. turgidum dicoccum* and *T. tauschii*, the two progenitors of hexaploid wheat, *T. aestivum*, probably took place in a farmer's field somewhere in western Iran about 8,000 years ago, when the cultivated wheat was brought into the area of the wild one.

SPECIES WITH GENOMES HOMOLOGOUS TO WHEAT			SPECIES WITH GENOMES HOMEOLOGOUS TO WHEAT	
TETRAPLOID PROGENITOR	DIPLOID DONORS OF *AA* AND *DD* GENOMES	POLYPLOIDS WITH ONE HOMOLOGOUS GENOME	CLOSELY RELATED SPECIES	LESS CLOSELY RELATED SPECIES

T. aestivum **is** *T. searsii*, **the putative diploid donor of the** *B* **genome of common wheat. Not represented in the table are a number of more distantly related species that can also be crossed with the cultivated wheats, such as common rye (*Secale cereale*), which has been successfully crossed with the tetraploid wheat** *T. turgidum* **to create the new manmade crop triticale.**

The existence of a large reservoir of useful genes in the wild relatives of wheat is apparent from the adaptation of the various wild forms to very different environments. They are found in a wide range of climatic regions: from cool, humid mountains to hot, dry valleys; from areas with an annual rainfall of 1,000 millimeters or more to arid regions with as little as 100 millimeters. Wild wheats also grow on many different types of soils, in some places even on salty ones.

The screening of these wild genetic resources for agronomically useful characteristics is only in its initial stages, and the full extent of the genetic treasure trove is not yet completely known. From the information gathered so far it is clear that the wild relatives of the cultivated wheats hold rich pools of genetic variation and carry many genes of great economic potential. Among the desirable characteristics that can be found in the wild relatives are improved quality and quantity of grain protein; increased resistance to fungal and viral diseases, to insects, to drought and to lodging (being beaten down by wind or rain); improved winterhardiness, heat tolerance and salt tolerance, and earlier ripening. The wild relatives may also be able to directly confer higher productivity.

The utilization of wild gene resources for the improvement of the cultivated wheats calls for the development of suitable methods of screening and evaluating the wild genetic material. Methods

SOME WILD RELATIVES OF WHEAT known to be crossable with cultivated tetraploid or hexaploid wheat are depicted. Number in parentheses under Latin name of each species gives the number of chromosomes in the gametes of that species; the somatic cells in each case have double the gametic chromosome number. Note that all the chromosome numbers are multiples of seven, the basic haploid set.

will also have to be devised to facilitate the selective transfer of desirable chromosomal elements from the wild species to the cultivated ones.

The first step in the utilization of the genes of the wild relatives of wheat should be the screening of single-plant collections for agronomically important traits. In such screening the genetic potentials of different collections are evaluated and promising genotypes are selected. The genetic interactions of individual genes or blocks of genes in the selected wild genotypes and the cultivated wheats need to be examined. Several analyses of such interactions have shown that when the genes of the wild relatives are separated from the rest of the wild genome and are introduced into the cultivated genome, they may have effects that cannot be forecast from the apparent traits—the phenotype—of the wild parent.

There are several wild relatives of wheat whose chromosomes (or some of them) are homologous with some of those of common wheat and consequently pair readily with them. Genes on these chromosomes can be transferred with comparatively little difficulty to common wheat. Most of the wild relatives, however, have chromosomes that are sufficiently different from those of wheat for them to ordinarily pair with them rarely, if at all. Special techniques are therefore necessary to introduce de-

SEVEN HYBRID LINES OF WHEAT derived from crosses between Chinese spring wheat, a variety of the cultivated hexaploid species *T. aestivum*, and the diploid wild species *T. longissimum* are arrayed at the right side of this illustration. Each new variety has a different pair of *T. longissimum* chromosomes added to the normal 21-pair chromosome complement of the cultivated parent species.

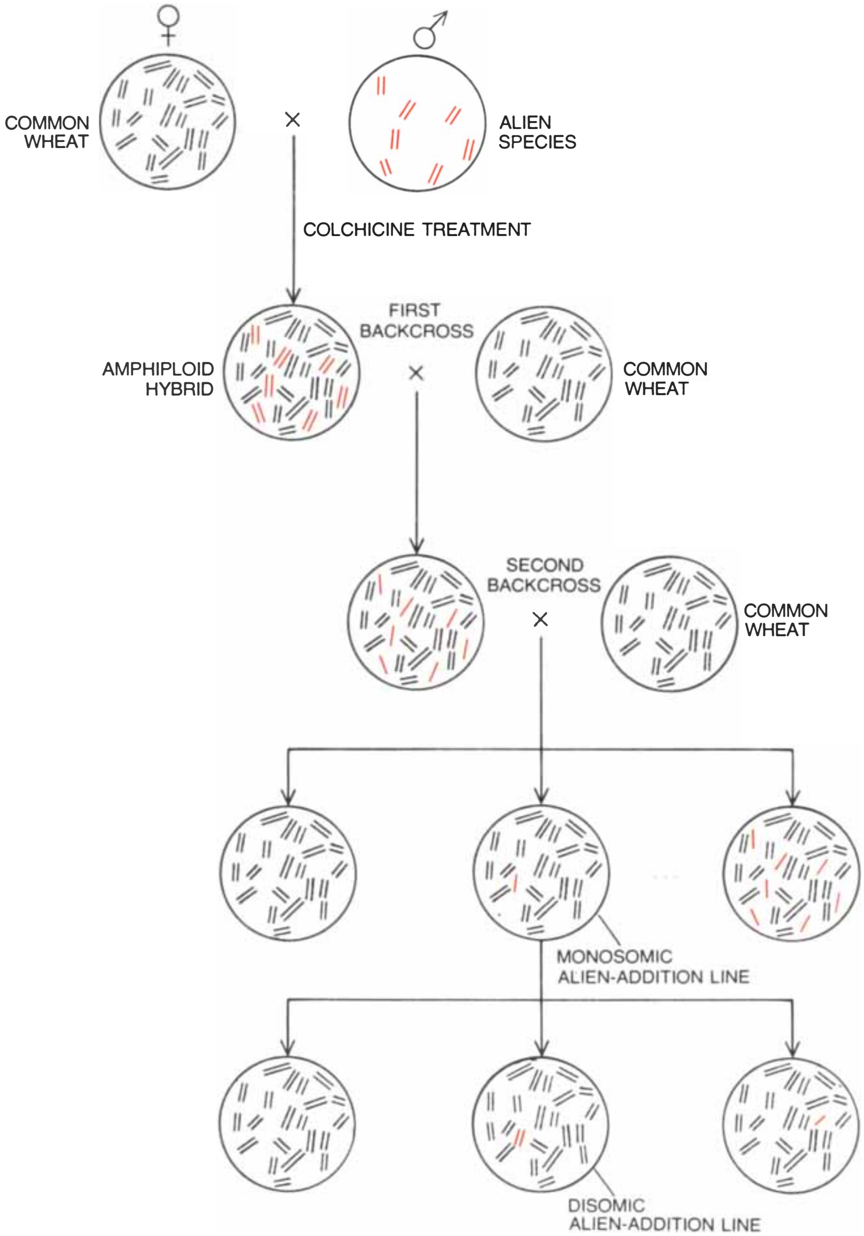
sirable genetic material from these wild relatives into the cultivated wheats.

The addition of an entire genome from a wild species (in other words, the production of a synthetic amphiploid) has been an attractive approach for many wheat geneticists. Erich Tschermak von Seysenegg and Hubert Bleier, working in Vienna in 1926, were the first to obtain an amphiploid species as a result of the spontaneous doubling of the chromosome number of a wheat hybrid, in their case one formed by crossing *T. turgidum dicoccoides* with the wild grass *T. ovatum,* thus demonstrating the possibility of species formation through amphiploidy in the wheat group. The discovery in 1937 that chromosome doubling can be induced by exposing hybrid plants to the alkaloid colchicine opened new possibilities for the production of synthetic amphiploids. Since then the production and exploitation of synthetic amphiploids in the Triticinae subtribe has been greater than it has been in any other group of higher plants, and many amphiploids have been synthesized from various interspecific and intergeneric crosses. The amphiploids are usually fully fertile, in contrast to the nondoubled hybrids, which tend to be sterile, particularly on the male side.

The production of such amphiploids, in which the adaptive hybrid combinations are stabilized by polyploidy and the segregation of parental characters is prevented, opens up many possibilities for the production of new crops. The synthetic octoploids that have been created by crossing hexaploid *T. aestivum* (mainly common wheat) and various diploid species, however, have all been inferior to common wheat. It is possible that ploidy levels higher than six are less favorable than the tetraploid and hexaploid levels.

A good example of a successful synthetic amphiploid at the hexaploid level is the new manmade crop triticale [see "Triticale," by Joseph H. Hulse and David Spurgeon; SCIENTIFIC AMERICAN, August, 1974]. This amphiploid derives from a cross between tetraploid wheat, *T. turgidum,* and common rye, *Secale cereale.* The hybrid combination has the possibility of combining the desirable winterhardiness of rye with some of the useful agronomic properties of wheat. In some regions triticale has even been found to outyield the best commercial wheat varieties. From the standpoint of quality triticale is characterized by a content of grain protein and essential amino acids higher than that of wheat. Hybridizations between triticale (genome *AABBRR*) and hexaploid wheat (genome *AABBDD*) have resulted in the creation of different kinds of mixed amphiploids, plants in which one or more rye-genome chromosomes are replaced by *DD*-genome wheat chromosomes. Some of the mixed amphiploid varieties exceed both parents in yield, quality and adaptability.

It is possible that the hybridization of tetraploid wheat with diploid relatives from semiarid or arid regions would result in new crops that could improve on currently cultivated wheat in drought resistance, salt tolerance or heat tolerance. These new crops would be either the primary amphiploids themselves or derivatives of their crosses to common wheat. Synthetic amphiploids can also serve as an effective bridge over the interspecific and intergeneric crossbreeding barriers that prevent the transfer of genes from wild diploid species to cultivated wheats. Finally, they are used in the creation of alien-addition lines and



ALIEN-ADDITION LINE is a type of hybrid wheat in which one pair of chromosomes from a given wild parent species has been added to the full chromosome complement of a cultivated wheat. In this demonstration of how an alien-addition line is created the cultivated parent is hexaploid; in other words, it has 21 chromosome pairs (*black lines*). The alien line is diploid: it has seven chromosome pairs (*colored lines*). Wild species with 14 or 21 chromosome pairs can also be used. The process begins with the generation of a viable amphiploid hybrid between the wild species and the cultivated wheat. (Interspecific hybrids of this type have a complete set of paired chromosomes derived from each parent species.) Repeated backcrossing with the cultivated species yields a series of monosomic alien-addition lines: hybrids in which the alien chromosome or chromosomes are univalent (unpaired). The chromosome number of any of these monosomic alien-addition lines can then be doubled to obtain a fully fertile disomic alien-addition line, in which the alien chromosome or chromosomes are bivalent (paired).

alien-substitution lines, two other important approaches in the repertory of the wheat geneticist.

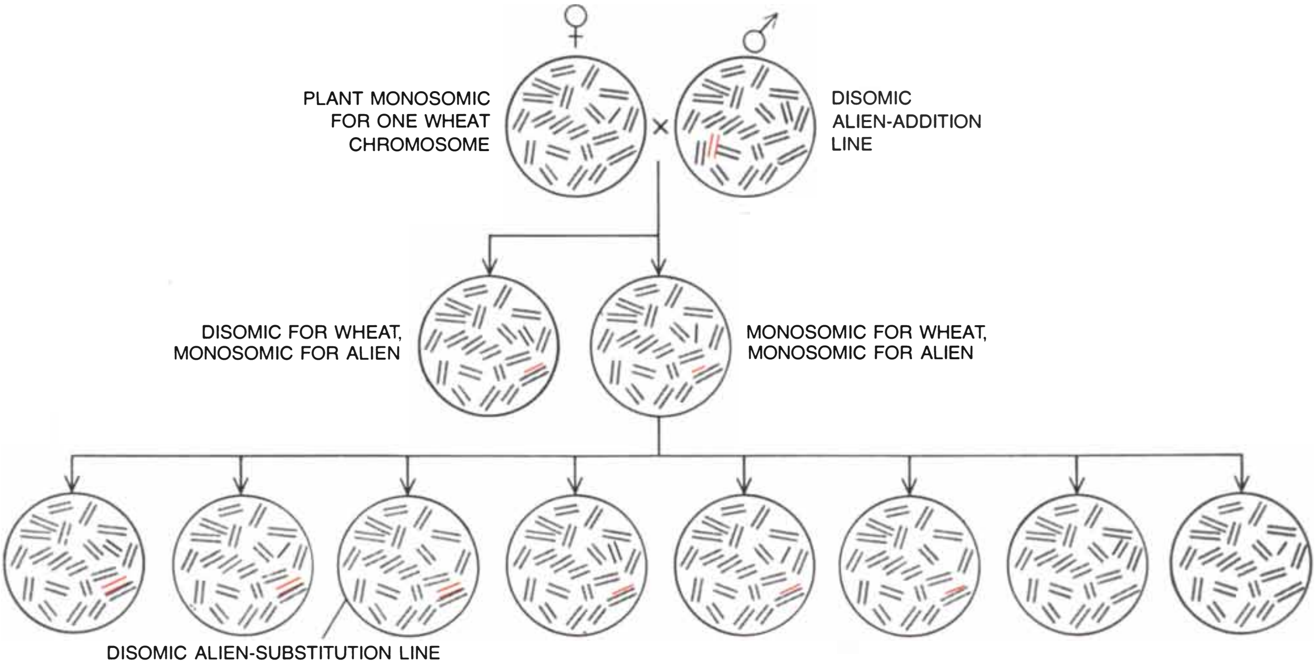
The genome of a wild species of wheat may contain in addition to desirable genes many undesirable ones with deleterious effects. Moreover, the undesirable genes may cancel any advantage the synthetic amphiploid has over its cultivated parent. Such unwanted material can only be eliminated, if it can be eliminated at all, after many generations of backcrosses with the cultivated species. Thus it is a great advantage to transfer the useful gene (or genes) from the alien genome with as little extra chromosomal material as possible; indeed, it is often desirable to transfer only one chromosome or even a segment of a chromosome.

Since the different diploid species of the Triticinae subtribe have been derived from a common ancestor, their chromosomes still retain some degree of similarity, even though they have differentiated considerably from one another. Such chromosomes are called homeologous (from the Greek root *homoios,* meaning similar) rather than homologous (from the Greek root *homos,* meaning the same). Since the basic chromosome number in this group is seven, the chromosomes of each genome of the Triticinae fall into seven distinct homeologous groups; each genome contains one chromosome pair of each homeologous group. The chromosomes of hexaploid wheat, for example, can be divided into seven homeologous groups of three pairs each, reflecting the origin of wheat from three related diploids, each with seven pairs of chromosomes. Since the genes on homeologous chromosomes have similar loci, a chromosome from any Triticinae genome can substitute for its homeologue in almost any other Triticinae genome and compensate, at least to some extent, for the homeologue's absence.

One handy method of transferring genetic material from wild relatives to cultivated wheats is to produce a series of alien-addition lines, each of which possesses the full complement of chromosomes of a cultivated wheat and one pair of chromosomes from a wild relative. The production of alien-addition lines requires that a viable hybrid between the wild species and cultivated wheat be obtainable; the rest of the procedure, although time-consuming, is almost certain to succeed [*see illustration on preceding page*].

This approach has been followed in adding individual chromosomes from several species of *Triticum, Agropyrum* and *Secale* to the hexaploid wheat complement and individual chromosomes from *Secale cereale, Agropyrum elongatum, T. umbellulatum* and the *DD* genome of common wheat to the tetraploid wheat complement. When direct crosses between hexaploid wheat and a wild diploid species are difficult or impossible, tetraploid wheat may serve as a bridging species. This method has been used, for example, in transferring individual chromosomes of *Haynaldia villosa* and *T. umbellulatum* to wheat.

Alien-addition lines, in turn, are used in the creation of alien-substitution lines, in each of which a pair of alien chromosomes is substituted for a pair of homeologous wheat chromosomes [*see illustration below*]. Since each alien chromosome has a homeologue in each of the three genomes of common wheat, three different substitution lines are possible for each alien chromosome. If the homeologues of a particular alien chromosome are not already known, the addition line is usually crossed with the seven monosomic lines (those carrying a given chromosome in one dose instead of the normal two) of one genome of the hexaploid. The double-monosomic first-generation progeny (plants with one alien monosome and one wheat monosome) are allowed to self-pollinate, and plants disomic for the wild chromosome and deficient in the wheat chromosome are recovered among the progeny. These substitution plants will appear infrequently and will have poor vigor and low fertility unless the missing wheat chromosome is a homeologue of the alien chromosome. When a substitution for one wheat homeologue has been made, substitution for the homeologue in each of the other two genomes can



ALIEN-SUBSTITUTION LINE is one in which a pair of alien chromosomes is substituted for a pair of homeologous cultivated-wheat chromosomes. The procedure begins with the crossing of a plant from an alien-addition line with a monosomic plant of common wheat (that is, a plant in which a given chromosome, emphasized here in black, is present in only one dose instead of the normal two). The double-monosomic first-generation progeny (plants with one alien monosome and one wheat monosome) are allowed to self-pollinate, and plants disomic for the wild chromosome and deficient in the homeologous wheat chromosome are recovered among the progeny. Each plant has 20 pairs of wheat chromosomes (*gray lines*) that are not involved in the hybridization steps. Male gametes with only 20 chromosomes derived from the selected first-generation plants presumably fail to compete successfully and hence are not shown.

be attempted with a high probability of success.

Once addition lines or substitution lines are created they can be maintained, propagated and supplied to various laboratories and breeding stations for testing under different conditions. These lines are in many cases fully fertile, so that they can be raised in quantity and used in large-scale field analysis. Substitution lines are more stable than addition lines, from which the alien chromosome tends to be lost.

Alien-addition lines and alien-substitution lines are seldom satisfactory for immediate practical use, although there has been one notable exception, namely several commercial European varieties with a pair of rye chromosomes substituted for a wheat pair. The main use of alien additions and substitutions is as starting material for making transfers involving less than an entire chromosome.

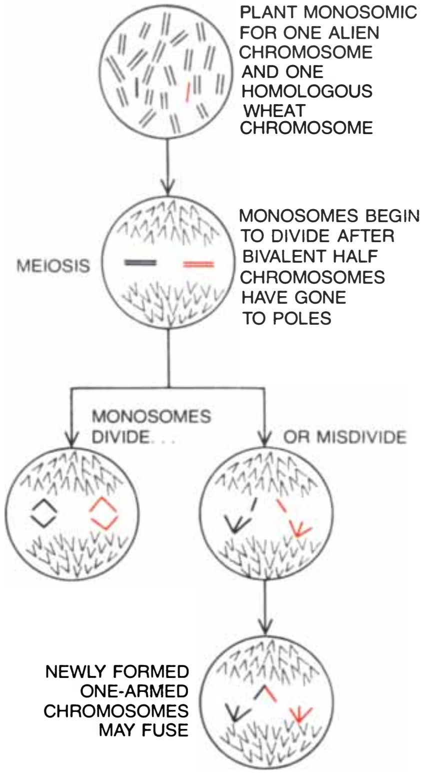
The transfer of only one arm of an alien chromosome instead of the entire chromosome reduces still further the danger of introducing undesirable genes along with the wanted one. Such transfers are made possible by the strong tendency of univalent (unpaired) chromosomes in wheat to misdivide at meiosis to give rise to one-armed chromosomes, which can then unite with other newly formed one-armed chromosomes. Thus from a plant monosomic for an alien chromosome and one of its wheat homeologues a substitution of one arm of the alien chromosome for the corresponding arm of the wheat chromosome can be obtained [*see illustration on this page*]. An entire alien arm, however, is still likely to carry unwanted genes as well as the wanted one. This probability, coupled with the fact that such a substitution is seldom if ever easier to obtain than a substitution involving only a portion of an arm, tends to make the whole-arm substitution an expedient to turn to only if attempts to obtain more desirable (that is, shorter) transfers fail.

Most alien chromosomes that do not ordinarily pair with their wheat homeologues can be induced to do so by altering the genetic environment. There are a number of genes in wheat that promote chromosome pairing and several that inhibit it. One gene in particular, the "pairing homeologous," or *Ph*, gene on the long arm of chromosome No. 5B, has a more decisive effect than any other. In its presence homologues pair regularly, but homeologues do not do so at all, whereas in its absence homeologues also pair (although less frequently). By simply deleting the *Ph* gene or neutralizing its action, then, most Triticinae chromosomes can be induced to pair with their wheat homeologues. Such induced homeologous pairing is usually the method of choice for transferring genes from alien chromosomes to those of wheat.

Several useful genes have already been transferred from wild relatives to common wheat through induced homeologous pairing and crossing-over. At the Plant Breeding Institute in Cambridge, England, a gene conferring resistance to stripe rust was transferred from *T. comosum*. At the Cereal Genetics Research Unit at the University of Missouri two genes conferring resistance to leaf rust were transferred from two different chromosomes of *Agropyrum elongatum*, and a gene conferring extraordinary resistance to many virulent races of stem rust was transferred from *T. speltoides*.

The likelihood of the successful transfer of an alien gene through induced homeologous pairing depends on the particular alien chromosome concerned. Not only must the chromosome be able to pair homeologously with one or more wheat chromosomes; it must also ordinarily pair in the right region or regions. Some alien chromosomes, such as those of most wild species of *Triticum* and those of *Agropyrum elongatum* and *A. intermedium*, are capable of pairing rather freely with their wheat homeologues, but others, such as those of rye and barley, have somewhat doubtful pairing ability. Even if some distantly related chromosomes pair, they may cross over with their homeologues unequally, leading to an undesirable duplication or a deficiency of certain segments.

Where an alien gene cannot be transferred through induced homeologous pairing the experimenter may want to try to induce the transfer by the use of ionizing radiation. Indeed, this technique was developed and widely employed before methods for inducing homeologous pairing became available. It is a much more laborious process, because the transferred alien segment, unless it is very short, must replace a segment of approximately the same length from a homeologous wheat chromosome. Although there seems to be some tendency for radiation-induced translocations to involve homeologous segments, such translocations are still greatly outnumbered by those involving nonhomeologues. Most of the translocations produced are quite unsatisfactory, because they involve a deficiency of some important wheat genes or a duplication of genes carried by the alien segment. Therefore many translocations involving the desired gene must be obtained and evaluated if there is to be a reasonable chance of getting even one translocation that has no associated deleterious effects. Nevertheless, by means of radiation potentially useful disease-resistance genes have been transferred to common wheat from *T. umbellulatum, A. elongatum, A. intermedium* and *Secale cereale*.



SUBSTITUTION OF ONE ARM of an alien chromosome for an arm of a homeologous wheat chromosome can result from the simultaneous misdivision of two univalent chromosomes at meiosis: the type of cell division in which the number of chromosomes is reduced in the production of reproductive cells. The misdivision is followed by the fusion of the ends of the one-armed chromosomes that happen to lie near each other. The right kind of misdivision occurs rarely in the same cell.

In spite of the known and anticipated difficulties of transferring genes to wheat from its more distant relatives, attempts to make such transfers are decidedly worthwhile. The greater the distance over which transfers can be made, the greater the possibility of introducing useful characteristics not present in the cultivated wheats.

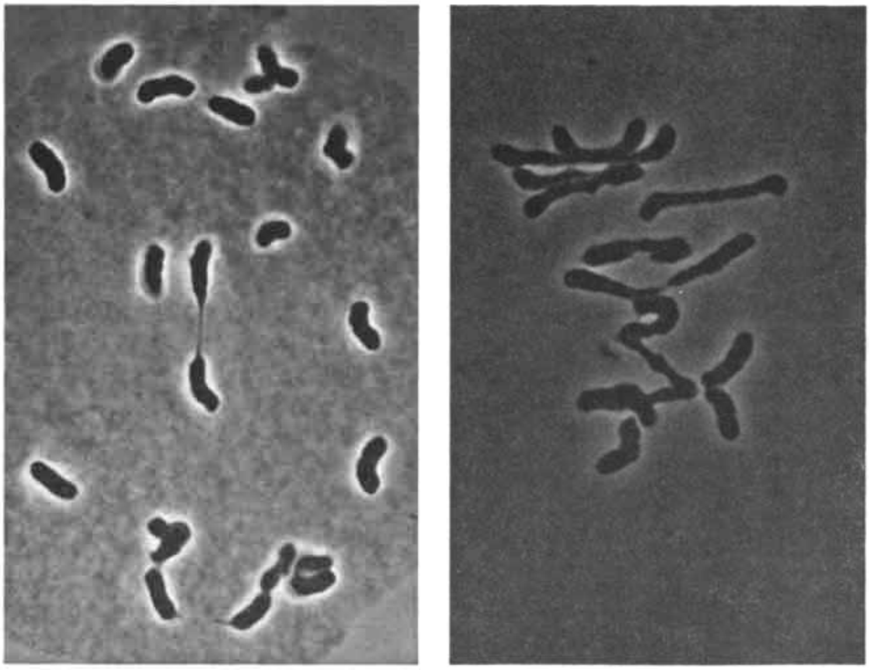
Wild relatives of wheat also have characteristics, for example male sterility, for which the genes are not in the nucleus of the cells but in the surrounding cytoplasm. Such characteristics can be transferred to cultivated wheats and perhaps utilized in the production of hybrid wheat. Many lines of common and durum wheat containing alien cytoplasms have been produced by crossing the wild species as a female with cultivated wheat and then backcrossing the first-generation hybrid and subsequent generations, again as females, to the cultivated species.

The wild species of the Triticinae show different degrees of cytogenetic affinity and phylogenetic relatedness to the cultivated common and durum wheats and can be classified accordingly [*see illustration on pages 106 and 107*]. In this classification the various wild relatives are divided into two main groups: those whose genomes are homologous with the genomes of the cultivated wheats and those whose genomes are homeologous with the genomes of those wheats. The first group consists of tetraploid wild wheat, the diploid donors of the *AA* and *DD* genomes of cultivated wheats, and polyploids sharing one genome with the cultivated wheats. Genetic material can be transferred within this group through pairing and crossing-over between homologous chromosomes. In most cases, however, it is necessary to overcome sterility barriers that stem from different ploidy levels or the dissimilarity of the genomes.

The second group of wild relatives includes species that are somewhat more distant from the cultivated wheats. The genomes of these species are homeologous rather than homologous with the genomes of the cultivated wheats, indicating that their chromosomes have differentiated from those of wheat to a greater extent. Because of this reduced similarity the transfer of desirable genetic material can be achieved only by inducing homeologous pairing and crossing-over or by inducing translocations. Three kinds of species can be recognized in this group of wild relatives: closely related, less closely related and distantly related.

In summary, it is obvious that in the near future the rich gene pools of the wild relatives of wheat will be exploited more aggressively for the improvement of the cultivated wheats. In order to achieve more efficient and comprehensive utilization of these gene pools more information is needed on the taxonomic, cytogenetic and evolutionary relations among the various wild species, as well as between wild and cultivated forms, on the geographical distribution and ecological specialization of the different species, on the genetic structure of natural populations and on the evolutionary mechanisms of building up genetic variability.

In addition as many samples as possible will have to be collected throughout the distribution areas of the various wild relatives, particularly in the centers of diversity of those species. This will facilitate the study of the nature and ranges of genetic variation of each species, as well as a thorough evaluation of the economic potential of the various collections. The investigation of the crossability relations between the wild relatives and the cultivated wheats should also be pursued vigorously, and more detailed studies of the genetic systems controlling pairing and crossing-over in the Triticeae should be initiated. Finally, simple cytogenetic procedures must be developed for producing and recovering desirable transfers of genetic material from the wild chromosomes to their cultivated wheat homeologues.



CHROMOSOME PAIRING is strongly influenced by the presence or absence of a single gene (the "pairing homeologous," or *Ph*, gene) on one of the wheat chromosomes. The suppression of homeologous pairing by the *Ph* gene is evident in these photographs of the chromosomes of hybrids obtained by crossing *T. turgidum durum* (durum or macaroni wheat) with *Secale cereale* (rye). With the *Ph* gene present (*left*) 19 chromosomes are unpaired and only two are paired; with the *Ph* gene absent (*right*) 18 chromosomes are paired and three are unpaired.

WHAT'S BETTER THAN SPEED READING?

SPEED LEARNING

(SPEED PLUS COMPREHENSION)

Speed Learning is replacing speed reading. It's easy to learn... lasts a lifetime... applies to everything you read... and is the only accredited course with the option of college or continuing education credits.

Do you have too much to read and too little time to read it? Do you mentally pronounce each word as you read? Do you frequently have to go back and re-read words or whole paragraphs you just finished reading? Do you have trouble concentrating? Do you quickly forget most of what you read?

If you answer "yes" to any of these questions — then here at last is the practical help you've been waiting for. Whether you read for business or pleasure, school or college, you will build exceptional skills from this major breakthrough in effective reading, created by Dr. Russell Stauffer at the University of Delaware.

Not just "speed reading" — but speed reading-thinking-understanding-remembering-and-learning

The new *Speed Learning Program* shows you step-by-proven-step how to increase your reading skill and speed, so you understand more, remember more and use more of everything you read. The typical remark made by the 75,000 slow readers who completed the *Speed Learning Program* was: "Why didn't someone teach me this a long time ago?" They were no longer held back by the lack of skills and poor reading habits. They could read almost as fast as they could think.

What makes Speed Learning so successful?

The new *Speed Learning Program* does not offer you a rehash of the usual eye-exercises, timing devices, costly gadgets you've probably heard about in connection with speed reading courses or even tried and found ineffective.

In just a few spare minutes a day of easy reading and exciting listening, you discover an entirely new way to read and think — a radical departure from anything you have ever seen or heard about. Research shows that reading is 95% *thinking* and only 5% eye movement. Yet most of today's speed reading programs spend their time teaching you rapid eye movement (5% of the problem) and ignore the most important part (95%) *thinking*. In brief, *Speed Learning* gives you what speed reading *can't*.

Imagine the new freedom you'll have when you learn how to dash through all types of reading material *at least* twice as fast as you do now, and with greater comprehension. Think of being able to get on top of the avalanche of newspapers, magazines and correspondence you have to read . . . finishing a stimulating book and retaining facts and details more clearly and with greater accuracy than ever before.

Listen-and-learn at your own pace

This is a practical, easy-to-learn program that will work for you — no matter how slow a reader you think you are now. The *Speed Learning Program* is scientifically planned to get you started quickly . . . to help you in spare minutes a day. It brings you a "teacher-on-cassettes" who guides you, instructs, encourages you, explaining material as you read. Interesting items taken from *Time* Magazine, *Business Week*, *Wall Street Journal*, *Family Circle*, *N.Y. Times* and many others, make the program stimulating, easy and fun . . . and so much more effective.

Executives, students, professional people, men and women in all walks of life from 15 to 70 have benefited from this program. Speed Learning is a fully accredited course . . . costing only 1/5 the price of less effective speed reading classroom courses. Now you can examine the same, easy, practical and proven methods at home . . . in spare time . . . without risking a penny.

Examine Speed Learning FREE for 15 days

You will be thrilled at how quickly this program will begin to develop new thinking and reading skills. After listening to just one cassette and reading the preface you will quickly see how you can achieve increases in both the speed at which you read and in the amount you understand and remember.

You must be delighted with what you see or you pay nothing. Examine this remarkable program for 15 days. If, at the end of that time you are not convinced that you would like to master Speed Learning, simply return the program and owe nothing. See the coupon for low price and convenient credit terms.

Note: Many companies and government agencies have tuition assistance plans for employees providing full or partial payment for college credit programs.

In most cases, the entire cost of your Speed Learning Program is Tax Deductible.

COLLEGE CREDITS

You may obtain *2 full semester hour credits* for course completion, wherever you reside. Credits offered through Whittier College (California). Details included in your program.

CONTINUING EDUCATION UNITS

National Management Association, the world's largest association of professional managers, awards 3.0 CEU's for course completion. CEU's can be applied toward the certificate in Management Studies.

PROFESSIONAL SOCIETIES

Speed Learning is offered internationally to members of professional associations such as: American Chemical Society, Foundation for Accounting Education, Institute of Electrical and Electronics Engineers and dozens more. Consult your Education Director for information.

BUSINESS, INDUSTRY, GOVERNMENT

Many companies and government agencies offer Speed Learning as a wholly-paid or tuition reimbursement program. Consult your Training or Personnel Director for details.

31S AB

learn INCORPORATED 113 Gaither Drive, Mt. Laurel, NJ 08054

YES! Please rush me the materials checked below:

☐ Please send the Speed Learning program @ $89.95 plus $4 postage and handling.
☐ Please send the Speed Learning Medical Edition @ $99.95 plus $4 postage and handling.
☐ Please send the Junior Speed Learning program (ages 11 to 16) @ $79.95 plus $4 postage and handling.

NJ residents add 5% sales tax.

Check method of payment below:

☐ Check or money order enclosed
☐ Charge my credit card under the regular payment terms
☐ Visa ☐ Master Card Interbank No.__________ ☐ American Express
Card No.__________ Exp. Date__________

I understand that if after 15 days I am not delighted in every way, that I may return the materials in their original condition for a full refund. No questions asked.

Name __________
Address __________
City__________ State__________ Zip__________
Signature__________

If you don't already own a cassette player, you may order this Deluxe Cassette Recorder for only $49.95. (Includes handling and delivery.)
Check here to order ☐

Outside USA add $10 per item plus $4 surface mail—Airmail extra

The Mind-Body Problem

Could calculating machines have pains, Martians have expectations and disembodied spirits have thoughts? The modern functionalist approach to psychology raises the logical possibility that they could

by Jerry A. Fodor

Modern philosophy of science has been devoted largely to the formal and systematic description of the successful practices of working scientists. The philosopher does not try to dictate how scientific inquiry and argument ought to be conducted. Instead he tries to enumerate the principles and practices that have contributed to good science. The philosopher has devoted the most attention to analyzing the methodological peculiarities of the physical sciences. The analysis has helped to clarify the nature of confirmation, the logical structure of scientific theories, the formal properties of statements that express laws and the question of whether theoretical entities actually exist.

It is only rather recently that philosophers have become seriously interested in the methodological tenets of psychology. Psychological explanations of behavior refer liberally to the mind and to states, operations and processes of the mind. The philosophical difficulty comes in stating in unambiguous language what such references imply.

Traditional philosophies of mind can be divided into two broad categories: dualist theories and materialist theories. In the dualist approach the mind is a nonphysical substance. In materialist theories the mental is not distinct from the physical; indeed, all mental states, properties, processes and operations are in principle identical with physical states, properties, processes and operations. Some materialists, known as behaviorists, maintain that all talk of mental causes can be eliminated from the language of psychology in favor of talk of environmental stimuli and behavioral responses. Other materialists, the identity theorists, contend that there are mental causes and that they are identical with neurophysiological events in the brain.

In the past 15 years a philosophy of mind called functionalism that is neither dualist nor materialist has emerged from philosophical reflection on developments in artificial intelligence, computational theory, linguistics, cybernetics and psychology. All these fields, which are collectively known as the cognitive sciences, have in common a certain level of abstraction and a concern with systems that process information. Functionalism, which seeks to provide a philosophical account of this level of abstraction, recognizes the possibility that systems as diverse as human beings, calculating machines and disembodied spirits could all have mental states. In the functionalist view the psychology of a system depends not on the stuff it is made of (living cells, metal or spiritual energy) but on how the stuff is put together. Functionalism is a difficult concept, and one way of coming to grips with it is to review the deficiencies of the dualist and materialist philosophies of mind it aims to displace.

The chief drawback of dualism is its failure to account adequately for mental causation. If the mind is nonphysical, it has no position in physical space. How, then, can a mental cause give rise to a behavioral effect that has a position in space? To put it another way, how can the nonphysical give rise to the physical without violating the laws of the conservation of mass, of energy and of momentum?

The dualist might respond that the problem of how an immaterial substance can cause physical events is not much obscurer than the problem of how one physical event can cause another. Yet there is an important difference: there are many clear cases of physical causation but not one clear case of nonphysical causation. Physical interaction is something philosophers, like all other people, have to live with. Nonphysical interaction, however, may be no more than an artifact of the immaterialist construal of the mental. Most philosophers now agree that no argument has successfully demonstrated why mind-body causation should not be regarded as a species of physical causation.

Dualism is also incompatible with the practices of working psychologists. The psychologist frequently applies the experimental methods of the physical sciences to the study of the mind. If mental processes were different in kind from physical processes, there would be no reason to expect these methods to work in the realm of the mental. In order to justify their experimental methods many psychologists urgently sought an alternative to dualism.

In the 1920's John B. Watson of Johns Hopkins University made the radical suggestion that behavior does not have mental causes. He regarded the behavior of an organism as its observable responses to stimuli, which he took to be the causes of its behavior. Over the next 30 years psychologists such as B. F. Skinner of Harvard University developed Watson's ideas into an elaborate world view in which the role of psychology was to catalogue the laws that determine causal relations between stimuli and responses. In this "radical behaviorist" view the problem of explaining the nature of the mind-body interaction vanishes; there is no such interaction.

Radical behaviorism has always worn an air of paradox. For better or worse, the idea of mental causation is deeply ingrained in our everyday language and in our ways of understanding our fellow men and ourselves. For example, people commonly attribute behavior to beliefs, to knowledge and to expectations. Brown puts gas in his tank because he believes the car will not run without it. Jones writes not "acheive" but "achieve" because he knows the rule about putting *i* before *e*. Even when a behavioral response is closely tied to an environmental stimulus, mental processes often intervene. Smith carries an umbrella because the sky is cloudy, but the weather is only part of the story. There are apparently also mental links in the causal chain: observation and expectation. The clouds affect Smith's behavior only because he observes them and because they induce in him an expectation of rain.

The radical behaviorist is unmoved by appeals to such cases. He is prepared to dismiss references to mental causes, however plausible they may seem, as the residue of outworn creeds. The radical behaviorist predicts that as psychologists come to understand more about the relations between stimuli and responses they will find it increasingly possible to explain behavior without postulating mental causes.

The strongest argument against behaviorism is that psychology has not turned out this way; the opposite has happened. As psychology has matured, the framework of mental states and processes that is apparently needed to account for experimental observations has grown all the more elaborate. Particularly in the case of human behavior psychological theories satisfying the methodological tenets of radical behaviorism have proved largely sterile, as would be expected if the postulated mental processes are real and causally effective.

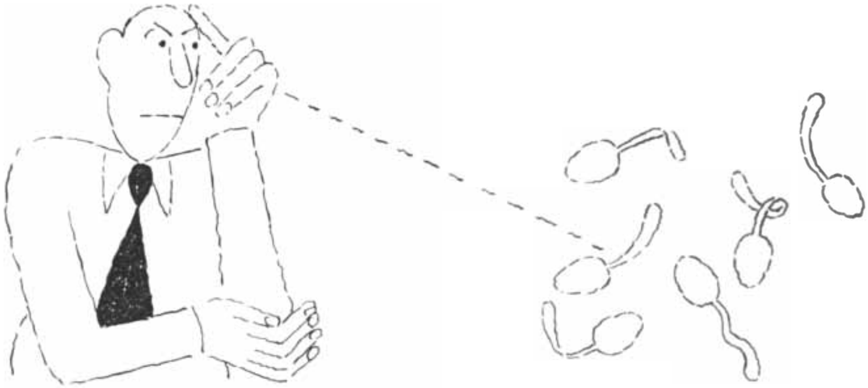
Nevertheless, many philosophers were initially drawn to radical behaviorism because, paradoxes and all, it seemed better than dualism. Since a psychology committed to immaterial substances was unacceptable, philosophers turned to radical behaviorism because it seemed to be the only alternative materialist philosophy of mind. The choice, as they saw it, was between radical behaviorism and ghosts.

By the early 1960's philosophers began to have doubts that dualism and radical behaviorism exhausted the possible approaches to the philosophy of mind. Since the two theories seemed unattractive, the right strategy might be to develop a materialist philosophy of mind that nonetheless allowed for mental causes. Two such philosophies emerged, one called logical behaviorism and the other called the central-state identity theory.

Logical behaviorism is a semantic theory about what mental terms mean. The basic idea is that attributing a mental state (say thirst) to an organism is the same as saying that the organism is disposed to behave in a particular way (for example to drink if there is water available). On this view every mental ascription is equivalent in meaning to an if-then statement (called a behavioral hypothetical) that expresses a behavioral disposition. For example, "Smith is thirsty" might be taken to be equivalent to the dispositional statement "If there were water available, then Smith would drink some." By definition a behavioral hypothetical includes no mental terms. The if-clause of the hypothetical speaks only of stimuli and the then-clause speaks only of behavioral responses. Since stimuli and responses are physical events, logical behaviorism is a species of materialism.

The strength of logical behaviorism is that by translating mental language into the language of stimuli and responses it provides an interpretation of psychological explanations in which behavioral effects are attributed to mental causes. Mental causation is simply the manifestation of a behavioral disposition. More precisely, mental causation is what happens when an organism has a behavioral disposition and the if-clause of the behavioral hypothetical expressing the disposition happens to be true. For example, the causal statement "Smith drank some water because he was thirsty" might be taken to mean "If there were water available, then Smith would drink some, and there was water available."

I have somewhat oversimplified logical behaviorism by assuming that each mental ascription can be translated by a unique behavioral hypothetical. Actually the logical behaviorist often maintains that it takes an open-ended set (perhaps an infinite set) of behavioral hypotheticals to spell out the behavioral disposition expressed by a mental term.



DUALISM is the philosophy of mind that regards the mind as a nonphysical substance. It divides everything there is in the world into two distinct categories: the mental and the physical. The chief difficulty with dualism is its failure to account adequately for the causal interaction of the mental and the physical. It is not evident how a nonphysical mind could give rise to any physical effects without violating the laws of conservation of mass, energy and momentum.

RADICAL BEHAVIORISM is the philosophy of mind that denies the existence of the mind and mental states, properties, processes and operations. The radical behaviorist believes behavior does not have mental causes. He considers the behavior of an organism to be its responses to stimuli. The role of psychology is to catalogue the relations between stimuli and responses.

The mental ascription "Smith is thirsty" might also be satisfied by the hypothetical "If there were orange juice available, then Smith would drink some" and by a host of other hypotheticals. In any event the logical behaviorist does not usually maintain he can actually enumerate all the hypotheticals that correspond to a behavioral disposition expressing a given mental term. He only insists that in principle the meaning of any mental term can be conveyed by behavioral hypotheticals.

The way the logical behaviorist has interpreted a mental term such as thirsty is modeled after the way many philosophers have interpreted a physical disposition such as fragility. The physical disposition "The glass is fragile" is often taken to mean something like "If the glass were struck, then it would break." By the same token the logical behaviorist's analysis of mental causation is similar to the received analysis of one kind of physical causation. The causal statement "The glass broke because it was fragile" is taken to mean something like "If the glass were struck, then it would break, and the glass was struck."

By equating mental terms with behavioral dispositions the logical behaviorist has put mental terms on a par with the nonbehavioral dispositions of the physical sciences. That is a promising move, because the analysis of nonbehavioral dispositions is on relatively solid philosophical ground. An explanation attributing the breaking of a glass to its fragility is surely something even the staunchest materialist can accept. By arguing that mental terms are synonymous with dispositional terms, the logical behaviorist has provided something the radical behaviorist could not: a materialist account of mental causation.

Nevertheless, the analogy between mental causation as construed by the logical behaviorist and physical causation goes only so far. The logical behaviorist treats the manifestation of a disposition as the sole form of mental causation, whereas the physical sciences recognize additional kinds of causation. There is the kind of causation where one physical event causes another, as when the breaking of a glass is attributed to its having been struck. In fact, explanations that involve event-event causation are presumably more basic than dispositional explanations, because the manifestation of a disposition (the breaking of a fragile glass) always involves event-event causation and not vice versa. In the realm of the mental many examples of event-event causation involve one mental state's causing another, and for this kind of causation logical behaviorism provides no analysis. As a result the logical behaviorist is committed to the tacit and implausible assumption that psychology requires a less robust notion of causation than the physical sciences require.

Event-event causation actually seems to be quite common in the realm of the mental. Mental causes typically give rise to behavioral effects by virtue of their interaction with other mental causes. For example, having a headache causes a disposition to take aspirin only if one also has the desire to get rid of the headache, the belief that aspirin exists, the belief that taking aspirin reduces headaches and so on. Since mental states interact in generating behavior, it will be necessary to find a construal of psychological explanations that posits mental processes: causal sequences of mental events. It is this construal that logical behaviorism fails to provide.

Such considerations bring out a fundamental way in which logical behaviorism is quite similar to radical behaviorism. It is true that the logical behaviorist, unlike the radical behaviorist, acknowledges the existence of mental states. Yet since the underlying tenet of logical behaviorism is that references to mental states can be translated out of psychological explanations by employing behavioral hypotheticals, all talk of mental states and processes is in a sense heuristic. The only facts to which the behaviorist is actually committed are facts about relations between stimuli and responses. In this respect logical behaviorism is just radical behaviorism in a semantic form. Although the former theory offers a construal of mental causation, the construal is Pickwickian. What does not really exist cannot cause anything, and the logical behaviorist, like the radical behaviorist, believes deep down that mental causes do not exist.

An alternative materialist theory of the mind to logical behaviorism is the central-state identity theory. According to this theory, mental events, states and processes are identical with neurophysiological events in the brain, and the property of being in a certain mental state (such as having a headache or believing it will rain) is identical with the property of being in a certain neurophysiological state. On this basis it is easy to make sense of the idea that a behavioral effect might sometimes have a chain of mental causes; that will be the case whenever a behavioral effect is contingent on the appropriate sequence of neurophysiological events.

The central-state identity theory acknowledges that it is possible for mental

LOGICAL BEHAVIORISM is a semantic thesis about what mental terms mean. The logical behaviorist maintains that mental terms express behavioral dispositions. Consider the mental state of being thirsty. The logical behaviorist maintains that the sentence "Smith is thirsty" might be taken as equivalent in meaning to the dispositional statement "If there were water available, then Smith would drink some." The strength of logical behaviorism is that it provides an account of mental causation: the realization of a behavioral disposition. For example, the causal statement "Smith drank some water because he was thirsty" might be taken to mean "If there were water available, then Smith would drink some, and there was water available."

CENTRAL-STATE IDENTITY THEORY is the philosophy of mind that equates mental events, states and processes with neurophysiological events. Property of being in a given mental state is identical with the property of being in a given neurophysiological state.

causes to interact causally without ever giving rise to any behavioral effect, as when a person thinks for a while about what he ought to do and then decides to do nothing. If mental processes are neurophysiological, they must have the causal properties of neurophysiological processes. Since neurophysiological processes are presumably physical processes, the central-state identity theory ensures that the concept of mental causation is as rich as the concept of physical causation.

The central-state identity theory provides a satisfactory account of what the mental terms in psychological explanations refer to, and so it is favored by psychologists who are dissatisfied with behaviorism. The behaviorist maintains that mental terms refer to nothing or that they refer to the parameters of stimulus-response relations. Either way the existence of mental entities is only illusory. The identity theorist, on the other hand, argues that mental terms refer to neurophysiological states. Thus he can take seriously the project of explaining behavior by appealing to its mental causes.

The chief advantage of the identity theory is that it takes the explanatory constructs of psychology at face value, which is surely something a philosophy of mind ought to do if it can. The identity theory shows how the mentalistic explanations of psychology could be not mere heuristics but literal accounts of the causal history of behavior. Moreover, since the identity theory is not a semantic thesis, it is immune to many arguments that cast in doubt logical behaviorism. A drawback of logical behaviorism is that the observation "John has a headache" does not seem to mean the same thing as a statement of the form "John is disposed to behave in such and such a way." The identity theorist, however, can live with the fact that "John has a headache" and "John is in such and such a brain state" are not synonymous. The assertion of the identity theorist is not that these sentences mean the same thing but only that they are rendered true (or false) by the same neurophysiological phenomena.

The identity theory can be held either as a doctrine about mental particulars (John's current pain or Bill's fear of animals) or as a doctrine about mental universals, or properties (having a pain or being afraid of animals). The two doctrines, called respectively token physicalism and type physicalism, differ in strength and plausibility. Token physicalism maintains only that all the mental particulars that happen to exist are neurophysiological, whereas type physicalism makes the more sweeping assertion that all the mental particulars there could possibly be are neurophysiological. Token physicalism does not rule out the logical possibility of machines and disembodied spirits having mental properties. Type physicalism dismisses this possibility because neither machines nor disembodied spirits have neurons.

Type physicalism is not a plausible doctrine about mental properties even if token physicalism is right about mental particulars. The problem with type physicalism is that the psychological constitution of a system seems to depend not on its hardware, or physical composition, but on its software, or program. Why should the philosopher dismiss the possibility that silicon-based Martians have pains, assuming that the silicon is properly organized? And why should the philosopher rule out the possibility of machines having beliefs, assuming that the machines are correctly programmed? If it is logically possible that Martians and machines could have mental properties, then mental properties and neurophysiological processes cannot be identical, however much they may prove to be coextensive.

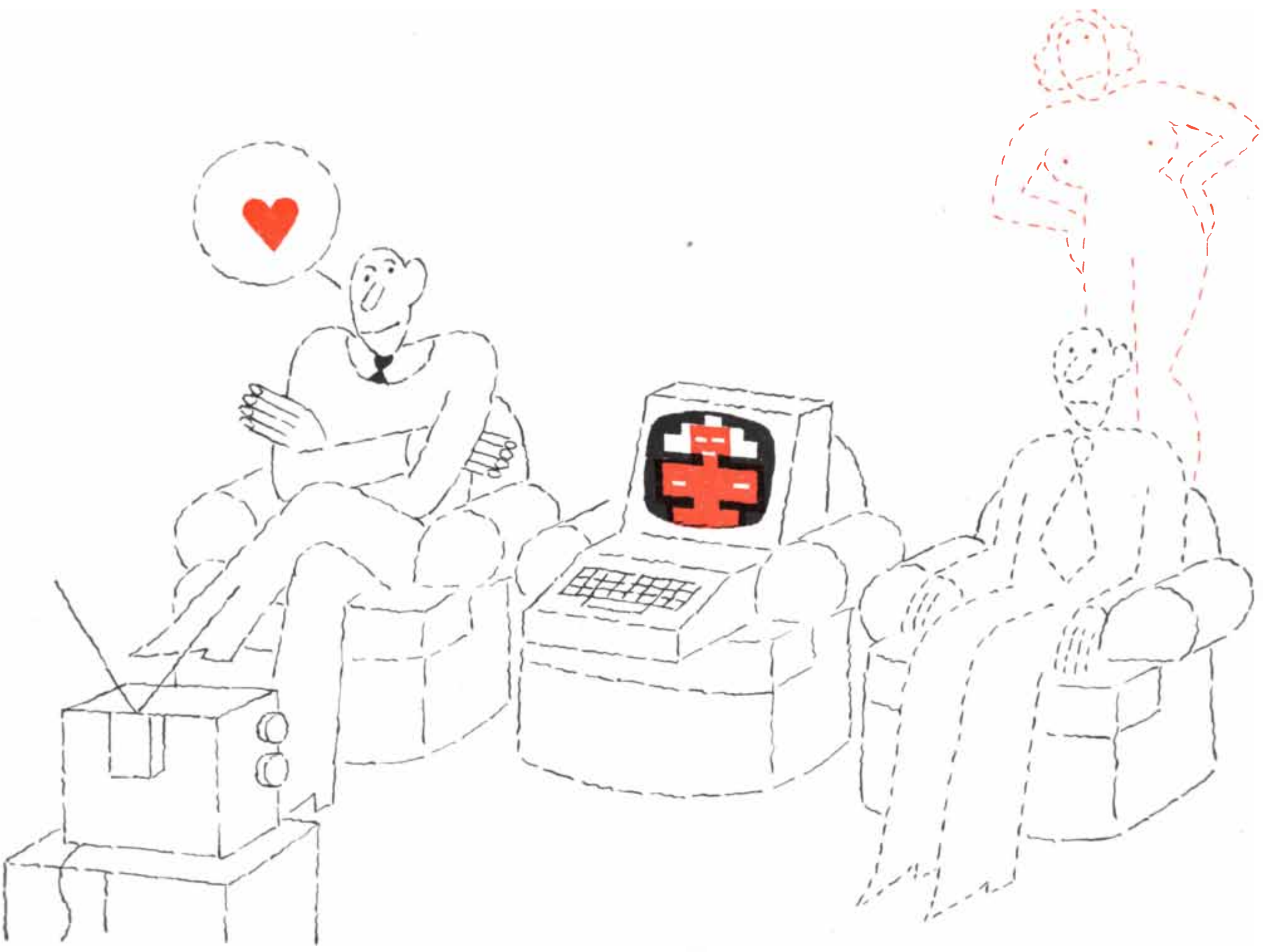
What it all comes down to is that there seems to be a level of abstraction at which the generalizations of psychology are most naturally pitched. This level of abstraction cuts across differences in the physical composition of the systems to which psychological generalizations apply. In the cognitive sciences, at least, the natural domain for psychological theorizing seems to be all systems that process information. The problem with type physicalism is that there are possible information-processing systems with the same psychological constitution as human beings but not the same physical organization. In principle all kinds of physically different things could have human software.

This situation calls for a relational account of mental properties that abstracts them from the physical structure of their bearers. In spite of the objections to logical behaviorism that I presented above, logical behaviorism was at least on the right track in offering a relational interpretation of mental properties: to have a headache is to be disposed to exhibit a certain pattern of relations between the stimuli one encounters and the responses one exhibits. If that is what having a headache is, however, there is no reason in principle why only heads that are physically similar to ours can ache. Indeed, according to logical behaviorism, it is a necessary truth that any system that has our stimulus-response contingencies also has our headaches.

All of this emerged 10 or 15 years ago as a nasty dilemma for the materialist program in the philosophy of mind. On the one hand the identity theorist (and not the logical behaviorist) had got right the causal character of the interactions of mind and body. On the other the logical behaviorist (and not the identity theorist) had got right the relational character of mental properties. Functionalism has apparently been able to resolve the dilemma. By stressing the distinction computer science draws between hardware and software the functionalist can make sense of both the causal and the relational character of the mental.

The intuition underlying functionalism is that what determines the psychological type to which a mental particular belongs is the causal role of the particular in the mental life of the organism. Functional individuation is differentiation with respect to causal role. A headache, for example, is identified with the type of mental state that among other things causes a disposition for taking aspirin in people who believe aspirin relieves a headache, causes a desire to rid oneself of the pain one is feeling, often causes someone who speaks English to say such things as "I have a headache" and is brought on by overwork, eyestrain and tension. This list is presumably not complete. More will be known about the nature of a headache as psychological and physiological research discovers more about its causal role.

Functionalism construes the concept of causal role in such a way that a mental state can be defined by its causal relations to other mental states. In this respect functionalism is completely different from logical behaviorism. Another major difference is that functionalism is not a reductionist thesis. It does not foresee, even in principle,



FUNCTIONALISM is the philosophy of mind based on the distinction that computer science draws between a system's hardware, or physical composition, and its software, or program. The psychology of a system such as a human being, a machine or a disembodied spirit does not depend on the stuff the system is made of (neurons, diodes or spiritual energy) but on how that stuff is organized. Functionalism does not rule out the possibility, however remote it may be, of mechanical and ethereal systems having mental states and processes.

the elimination of mentalistic concepts from the explanatory apparatus of psychological theories.

The difference between functionalism and logical behaviorism is brought out by the fact that functionalism is fully compatible with token physicalism. The functionalist would not be disturbed if brain events turn out to be the only things with the functional properties that define mental states. Indeed, most functionalists fully expect it will turn out that way.

Since functionalism recognizes that mental particulars may be physical, it is compatible with the idea that mental causation is a species of physical causation. In other words, functionalism tolerates the materialist solution to the mind-body problem provided by the central-state identity theory. It is possible for the functionalist to assert both that mental properties are typically defined in terms of their relations and that interactions of mind and body are typically causal in however robust a notion of causality is required by psychological explanations. The logical behaviorist can endorse only the first assertion and the type physicalist only the second. As a result functionalism seems to capture the best features of the materialist alternatives to dualism. It is no wonder that functionalism has become increasingly popular.

Machines provide good examples of two concepts that are central to functionalism: the concept that mental states are interdefined and the concept that they can be realized by many systems. The illustration on the next page contrasts a behavioristic Coke machine with a mentalistic one. Both machines dispense a Coke for 10 cents. (The price has not been affected by inflation.) The states of the machines are defined by reference to their causal roles, but only the machine on the left would satisfy the behaviorist. Its single state (*S0*) is completely specified in terms of stimuli and responses. *S0* is the state a machine is in if, and only if, given a dime as the input, it dispenses a Coke as the output.

The machine on the right in the illustration has interdefined states (*S1* and *S2*), which are characteristic of functionalism. *S1* is the state a machine is in if, and only if, (1) given a nickel, it dispenses nothing and proceeds to *S2*, and (2) given a dime, it dispenses a Coke and stays in *S1*. *S2* is the state a machine is in if, and only if, (1) given a nickel, it dispenses a Coke and proceeds to *S1*, and (2) given a dime, it dispenses a Coke and a nickel and proceeds to *S1*. What *S1* and *S2* jointly amount to is the machine's dispensing a Coke if it is given a dime, dispensing a Coke and a nickel if it is given a dime and a nickel and waiting to be given a second nickel if it has been given a first one.

Since *S1* and *S2* are each defined by hypothetical statements, they can be viewed as dispositions. Nevertheless, they are not behavioral dispositions because the consequences an input has for a machine in *S1* or *S2* are not specified solely in terms of the output of the machine. Rather, the consequences also involve the machine's internal states.

Nothing about the way I have described the behavioristic and mentalistic Coke machines puts constraints on what they could be made of. Any system whose states bore the proper relations to inputs, outputs and other states could be one of these machines. No doubt it is reasonable to expect such a system to be constructed out of such things as wheels, levers and diodes (token physicalism for Coke machines). Similarly, it is reasonable to expect that our minds may prove to be neurophysiological (token physicalism for human beings).

Nevertheless, the software description of a Coke machine does not logically require wheels, levers and diodes for its concrete realization. By the same token, the software description of the mind does not logically require neurons. As far as functionalism is concerned a Coke machine with states *S1* and *S2* could be made of ectoplasm, if there is such stuff and if its states have the right causal properties. Functionalism allows for the possibility of disembodied Coke machines in exactly the same way and to the same extent that it allows for the possibility of disembodied minds.

To say that *S1* and *S2* are interdefined and realizable by different kinds of hardware is not, of course, to say that a Coke machine has a mind. Although interdefinition and functional specification are typical features of mental states, they are clearly not sufficient for mentality. What more is required is a question to which I shall return below.

Some philosophers are suspicious of functionalism because it seems too easy. Since functionalism licenses the individuation of states by reference to their causal role, it appears to allow a trivial explanation of any observed event *E*, that is, it appears to postulate an *E*-causer. For example, what makes the valves in a machine open? Why, the operation of a valve opener. And what is a valve opener? Why, anything that has the functionally defined property of causing valves to open.

In psychology this kind of question-begging often takes the form of theories that in effect postulate homunculi with the selfsame intellectual capacities the theorist set out to explain. Such is the case when visual perception is explained by simply postulating psychological mechanisms that process visual information. The behaviorist has often charged the mentalist, sometimes justifiably, of mongering this kind of question-begging pseudo explanation. The charge will have to be met if functionally defined mental states are to have a serious role in psychological theories.

The burden of the accusation is not untruth but triviality. There can be no doubt that it is a valve opener that opens valves, and it is likely that visual perception is mediated by the processing of visual information. The charge is that such putative functional explanations are mere platitudes. The functionalist can meet this objection by allowing functionally defined theoretical constructs only where mechanisms exist that can carry out the function and only where he has some notion of what such mechanisms might be like. One way of imposing this requirement is to identify the mental processes that psychology postulates with the operations of the restricted class of possible computers called Turing machines.

A Turing machine can be informally characterized as a mechanism with a finite number of program states. The inputs and outputs of the machine are written on a tape that is divided into squares each of which includes a symbol from a finite alphabet. The machine scans the tape one square at a time. It can erase the symbol on a scanned square and print a new one in its place. The machine can execute only the elementary mechanical operations of scanning, erasing, printing, moving the tape and changing state.

The program states of the Turing machine are defined solely in terms of the input symbols on the tape, the output symbols on the tape, the elementary operations and the other states of the program. Each program state is therefore functionally defined by the part it plays in the overall operation of the machine. Since the functional role of a state depends on the relation of the state to other states as well as to inputs and outputs, the relational character of the mental is captured by the Turing-machine version of functionalism. Since the definition of a program state never refers to the physical structure of the system running the program, the Turing-machine version of functionalism also captures the idea that the character of a mental state is independent of its physical realization. A human being, a roomful of people, a computer and a disembodied spirit would all be a Turing machine if they operated according to a Turing-machine program.

The proposal is to restrict the functional definition of psychological states to those that can be expressed in terms of the program states of Turing machines. If this restriction can be enforced, it provides a guarantee that psychological theories will be compatible with the demands of mechanisms. Since Turing machines are very simple devices, they are in principle quite easy to build. Consequently by formulating a psychological explanation as a Turing-machine program the psychologist ensures that the explanation is mechanistic, even though the hardware realizing the mechanism is left open.

There are many kinds of computational mechanisms other than Turing machines, and so the formulation of a functionalist psychological theory in Turing-machine notation provides only a sufficient condition for the theory's being mechanically realizable. What makes the condition interesting, however, is that the simple Turing machine can perform many complex tasks. Although the elementary operations of the Turing machine are restricted, iterations of the operations enable the machine to carry out any well-defined computation on discrete symbols.

An important tendency in the cognitive sciences is to treat the mind chiefly as a device that manipulates symbols. If a mental process can be functionally defined as an operation on symbols, there is a Turing machine capable of carrying out the computation and a variety of mechanisms for realizing the Turing machine. Where the manipulation of symbols is important the Turing machine provides a connection between functional explanation and mechanistic explanation.

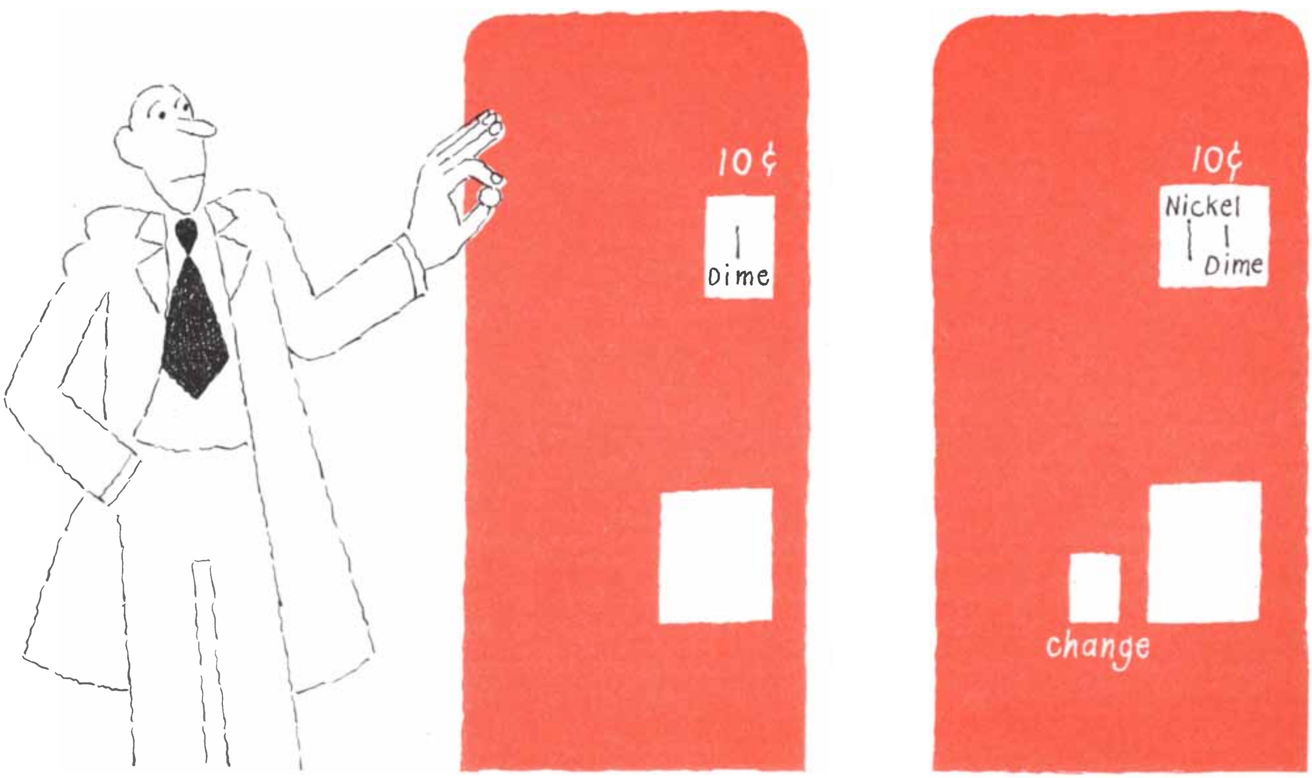
The reduction of a psychological theory to a program for a Turing machine is a way of exorcising the homunculi. The reduction ensures that no operations have been postulated except those that could be performed by a familiar mechanism. Of course, the working psychologist usually cannot specify the reduction for each functionally individuated process in every theory he is prepared to take seriously. In practice the argument usually goes in the opposite direction; if the postulation of a mental operation is essential to some cherished psychological explanation, the theorist tends to assume that there must be a program for a Turing machine that will carry out that operation.

The "black boxes" that are common in flow charts drawn by psychologists often serve to indicate postulated mental processes for which Turing reductions are wanting. Even so, the possibility in principle of such reductions serves as a methodological constraint on psychological theorizing by determining what functional definitions are to be allowed and what it would be like to know that everything has been explained that could possibly need explanation.

Such is the origin, the provenance and the promise of contemporary functionalism. How much has it actually paid off? This question is not easy to answer because much of what is now happening in the philosophy of mind and the cognitive sciences is directed at exploring the

	STATE S0
DIME INPUT	DISPENSES A COKE

	STATE S1	STATE S2
NICKEL INPUT	GIVES NO OUTPUT AND GOES TO S2	DISPENSES A COKE AND GOES TO S1
DIME INPUT	DISPENSES A COKE AND STAYS IN S1	DISPENSES A COKE AND A NICKEL AND GOES TO S1



TWO COKE MACHINES bring out the difference between behaviorism (the doctrine that there are no mental causes) and mentalism (the doctrine that there are mental causes). Both machines dispense a Coke for 10 cents and have states that are defined by reference to their causal role. The machine at the left is a behavioristic one: its single state (*S0*) is defined solely in terms of the input and the output. The machine at the right is a mentalistic one: its two states (*S1, S2*) must be defined not only in terms of the input and the output but also in terms of each other. To put it another way, the output of the Coke machine depends on the state the machine is in as well as on the input. The functionalist philosopher maintains that mental states are interdefined, like the internal states of the mentalistic Coke machine.

scope and limits of the functionalist explanations of behavior. I shall, however, give a brief overview.

An obvious objection to functionalism as a theory of the mind is that the functionalist definition is not limited to mental states and processes. Catalysts, Coke machines, valve openers, pencil sharpeners, mousetraps and ministers of finance are all in one way or another concepts that are functionally defined, but none is a mental concept such as pain, belief and desire. What, then, characterizes the mental? And can it be captured in a functionalist framework?

The traditional view in the philosophy of mind has it that mental states are distinguished by their having what are called either qualitative content or intentional content. I shall discuss qualitative content first.

It is not easy to say what qualitative content is; indeed, according to some theories, it is not even possible to say what it is because it can be known not by description but only by direct experience. I shall nonetheless attempt to describe it. Try to imagine looking at a blank wall through a red filter. Now change the filter to a green one and leave everything else exactly the way it was. Something about the character of your experience changes when the filter does, and it is this kind of thing that philosophers call qualitative content. I am not entirely comfortable about introducing qualitative content in this way, but it is a subject with which many philosophers are not comfortable.

The reason qualitative content is a problem for functionalism is straightforward. Functionalism is committed to defining mental states in terms of their causes and effects. It seems, however, as if two mental states could have all the same causal relations and yet could differ in their qualitative content. Let me illustrate this with the classic puzzle of the inverted spectrum.

It seems possible to imagine two observers who are alike in all relevant psychological respects except that experiences having the qualitative content of red for one observer would have the qualitative content of green for the other. Nothing about their behavior need reveal the difference because both of them see ripe tomatoes and flaming sunsets as being similar in color and both of them call that color "red." Moreover, the causal connection between their (qualitatively distinct) experiences and their other mental states could also be identical. Perhaps they both think of Little Red Riding Hood when they see ripe tomatoes, feel depressed when they see the color green and so on. It seems as if anything that could be packed into the notion of the causal role of their experiences could be shared by them, and yet the qualitative content of the experiences could be as different as you like. If this is possible, then the functionalist account does not work for mental states that have qualitative content. If one person is having a green experience while another person is having a red one, then surely they must be in different mental states.

The example of the inverted spectrum is more than a verbal puzzle. Having qualitative content is supposed to be a chief factor in what makes a mental state conscious. Many psychologists who are inclined to accept the functionalist framework are nonetheless worried about the failure of functionalism to reveal much about the nature of consciousness. Functionalists have made a few ingenious attempts to talk themselves and their colleagues out of this worry, but they have not, in my view, done so with much success. (For example, perhaps one is wrong in thinking one can imagine what an inverted spectrum would be like.) As matters stand, the problem of qualitative content poses a serious threat to the assertion that functionalism can provide a general theory of the mental.

Functionalism has fared much better with the intentional content of mental states. Indeed, it is here that the major achievements of recent cognitive science are found. To say that a mental state has intentional content is to say that it has certain semantic properties. For example, for Enrico to believe Galileo was Italian apparently involves a three-way relation between Enrico, a belief and a proposition that is the content of the belief (namely the proposition that Galileo was Italian). In particular it is an essential property of Enrico's belief that it is about Galileo (and not about, say, Newton) and that it is true if, and only if, Galileo was indeed Italian. Philosophers are divided on how these considerations fit together, but it is widely agreed that beliefs involve semantic properties such as expressing a proposition, being true or false and being about one thing rather than another.

It is important to understand the semantic properties of beliefs because theories in the cognitive sciences are largely about the beliefs organisms have. Theories of learning and perception, for example, are chiefly accounts of how the host of beliefs an organism has are determined by the character of its experiences and its genetic endowment. The functionalist account of mental states does not by itself provide the required insights. Mousetraps are functionally defined, yet mousetraps do not express propositions and they are not true or false.

There is at least one kind of thing other than a mental state that has intentional content: a symbol. Like thoughts, symbols seem to be about things. If someone says "Galileo was Italian," his utterance, like Enrico's belief, expresses a proposition about Galileo that is true or false depending on Galileo's homeland. This parallel between the symbolic and the mental underlies the traditional quest for a unified treatment of language and mind. Cognitive science is now trying to provide such a treatment.

The basic concept is simple but striking. Assume that there are such things as mental symbols (mental representations) and that mental symbols have semantic properties. On this view having a belief involves being related to a mental symbol, and the belief inherits its semantic properties from the mental symbol that figures in the relation. Mental processes (thinking, perceiving, learning and so on) involve causal interactions among relational states such as having a belief. The semantic properties of the words and sentences we utter are in turn inherited from the semantic properties of the mental states that language expresses.

Associating the semantic properties of mental states with those of mental symbols is fully compatible with the computer metaphor, because it is natural to think of the computer as a mechanism that manipulates symbols. A computation is a causal chain of computer states and the links in the chain are operations on semantically interpreted formulas in a machine code. To think of a system (such as the nervous system) as a computer is to raise questions about the nature of the code in which it computes and the semantic properties of the symbols in the code. In fact, the analogy between minds and computers actually implies the postulation of mental symbols. There is no computation without representation.

The representational account of the mind, however, predates considerably the invention of the computing machine. It is a throwback to classical epistemology, which is a tradition that includes philosophers as diverse as John Locke, David Hume, George Berkeley, René Descartes, Immanuel Kant, John Stuart Mill and William James.

Hume, for one, developed a representational theory of the mind that included five points. First, there exist "Ideas," which are a species of mental symbol. Second, having a belief involves entertaining an Idea. Third, mental processes are causal associations of Ideas. Fourth, Ideas are like pictures. And fifth, Ideas have their semantic properties by virtue of what they resemble: the Idea of John is about John because it looks like him.

Contemporary cognitive psychologists do not accept the details of Hume's theory, although they endorse much of its spirit. Theories of computation provide a far richer account of mental processes than the mere association of Ideas. And only a few psychologists still think


Consumer Orientation
No. 10 in a Series
Subject:
The Dynamic Response of Tires
10
Porsche 924 Turbo
ay LATERAL ACCELERATION (g)
CAR A.
CAR B.
A.
B.
WET
An automobile propels itself and changes direction through reaction to the ground.
Most of the external forces that act upon it are generated through and limited by the friction of its tires.
The tires themselves generate friction by mechanical gripping of irregularities, high-speed deformation, and molecular adhesion between the rubber and ground contact-areas.
Because tires are so important, at Porsche we do more than merely specify their type and size. We work with tire manufacturers in their design and testing.
Consider our hydroplaning tests. Hydroplaning occurs when a vehicle's tires cannot displace water from the tire-ground contact area fast enough to allow them to make contact with the ground. Instead, the vehicle's tires actually ride up on a film of water. And, as a result, traction, directional stability, and control are greatly reduced.
Because hydroplaning can never be totally eliminated—on the skid pad at Weissach, we test for both water dispersal and vehicle directional control on recovery (see diagram at left).
Car A, traveling at 45 mph, disperses the 0.25 inches of water in its path and maintains a constant lateral acceleration.
Car B, also traveling at 45 mph, encounters 0.50 inches of water and hydroplanes to the outer edge. Lateral acceleration diminishes during hydroplaning, but returns on recovery.
What we learn from our hydroplaning tests, we incorporate into the tires, the suspension, the total car. Because in over 30 years of competition driving, we've learned it's the sum of an infinite number of details that separates winning from all else.
Test drive the Porsche 924 Turbo today. For your nearest dealer, call toll-free: (800) 447-4700. In Illinois, (800) 322-4400.
PORSCHE + AUDI
NOTHING EVEN COMES CLOSE


How we plated a connector so it wouldn't interfere with your profits.

that imagery is the chief vehicle of mental representation. Nevertheless, the most significant break with Hume's theory lies in the abandoning of resemblance as an explanation of the semantic properties of mental representations.

Many philosophers, starting with Berkeley, have argued that there is something seriously wrong with the suggestion that the semantic relation between a thought and what the thought is about could be one of resemblance. Consider the thought that John is tall. Clearly the thought is true only of the state of affairs consisting of John's being tall. A theory of the semantic properties of a thought should therefore explain how this particular thought is related to this particular state of affairs. According to the resemblance theory, entertaining the thought involves having a mental image that shows John to be tall. To put it another way, the relation between the thought that John is tall and his being tall is like the relation between a tall man and his portrait.

The difficulty with the resemblance theory is that any portrait showing John to be tall must also show him to be many other things: clothed or naked, lying, standing or sitting, having a head or not having one, and so on. A portrait of a tall man who is sitting down resembles a man's being seated as much as it resembles a man's being tall. On the resemblance theory it is not clear what distinguishes thoughts about John's height from thoughts about his posture.

The resemblance theory turns out to encounter paradoxes at every turn. The possibility of construing beliefs as involving relations to semantically interpreted mental representations clearly depends on having an acceptable account of where the semantic properties of the mental representations come from. If resemblance will not provide this account, what will?

The current idea is that the semantic properties of a mental representation are determined by aspects of its functional role. In other words, a sufficient condition for having semantic properties can be specified in causal terms. This is the connection between functionalism and the representational theory of the mind. Modern cognitive psychology rests largely on the hope that these two doctrines can be made to support each other.

No philosopher is now prepared to say exactly how the functional role of a mental representation determines its semantic properties. Nevertheless, the functionalist recognizes three types of causal relation among psychological states involving mental representations, and they might serve to fix the semantic properties of mental representations. The three types are causal relations among mental states and stimuli, mental states and responses and some mental states and other ones.

Consider the belief that John is tall. Presumably the following facts, which correspond respectively to the three types of causal relation, are relevant to determining the semantic properties of the mental representation involved in the belief. First, the belief is a normal effect of certain stimulations, such as seeing John in circumstances that reveal his height. Second, the belief is the normal cause of certain behavioral effects, such as uttering "John is tall." Third, the belief is a normal cause of certain other beliefs and a normal effect of certain other beliefs. For example, anyone who believes John is tall is very likely also to believe someone is tall. Having the first belief is normally causally sufficient for having the second belief. And anyone who believes everyone in the room is tall and also believes John is in the room will very likely believe John is tall. The third belief is a normal effect of the first two. In short, the functionalist maintains that the proposition expressed by a given mental representation depends on the causal properties of the mental states in which that mental representation figures.

The concept that the semantic properties of mental representations are determined by aspects of their functional role is at the center of current work in the cognitive sciences. Nevertheless, the concept may not be true. Many philosophers who are unsympathetic to the cognitive turn in modern psychology doubt its truth, and many psychologists would probably reject it in the bald and unelaborated way that I have sketched it. Yet even in its skeletal form, there is this much to be said in its favor: It legitimizes the notion of mental representation, which has become increasingly important to theorizing in every branch of the cognitive sciences. Recent advances in formulating and testing hypotheses about the character of mental representations in fields ranging from phonetics to computer vision suggest that the concept of mental representation is fundamental to empirical theories of the mind.

The behaviorist has rejected the appeal to mental representation because it runs counter to his view of the explanatory mechanisms that can figure in psychological theories. Nevertheless, the science of mental representation is now flourishing. The history of science reveals that when a successful theory comes into conflict with a methodological scruple, it is generally the scruple that gives way. Accordingly the functionalist has relaxed the behaviorist constraints on psychological explanations. There is probably no better way to decide what is methodologically permissible in science than by investigating what successful science requires.

Gels

They are mostly fluid, given form by a network of polymer strands. A balance of forces maintains this state of affairs; disturbing it infinitesimally can bring on a phase transition and collapse the gel

by Toyoichi Tanaka

A gel is a form of matter intermediate between a solid and a liquid. It consists of polymers, or long-chain molecules, cross-linked to create a tangled network and immersed in a liquid medium. The properties of the gel depend strongly on the interaction of these two components. The liquid prevents the polymer network from collapsing into a compact mass; the network prevents the liquid from flowing away. Depending on chemical composition and other factors, gels vary in consistency from viscous fluids to fairly rigid solids, but typically they are soft and resilient or, in a word, jellylike.

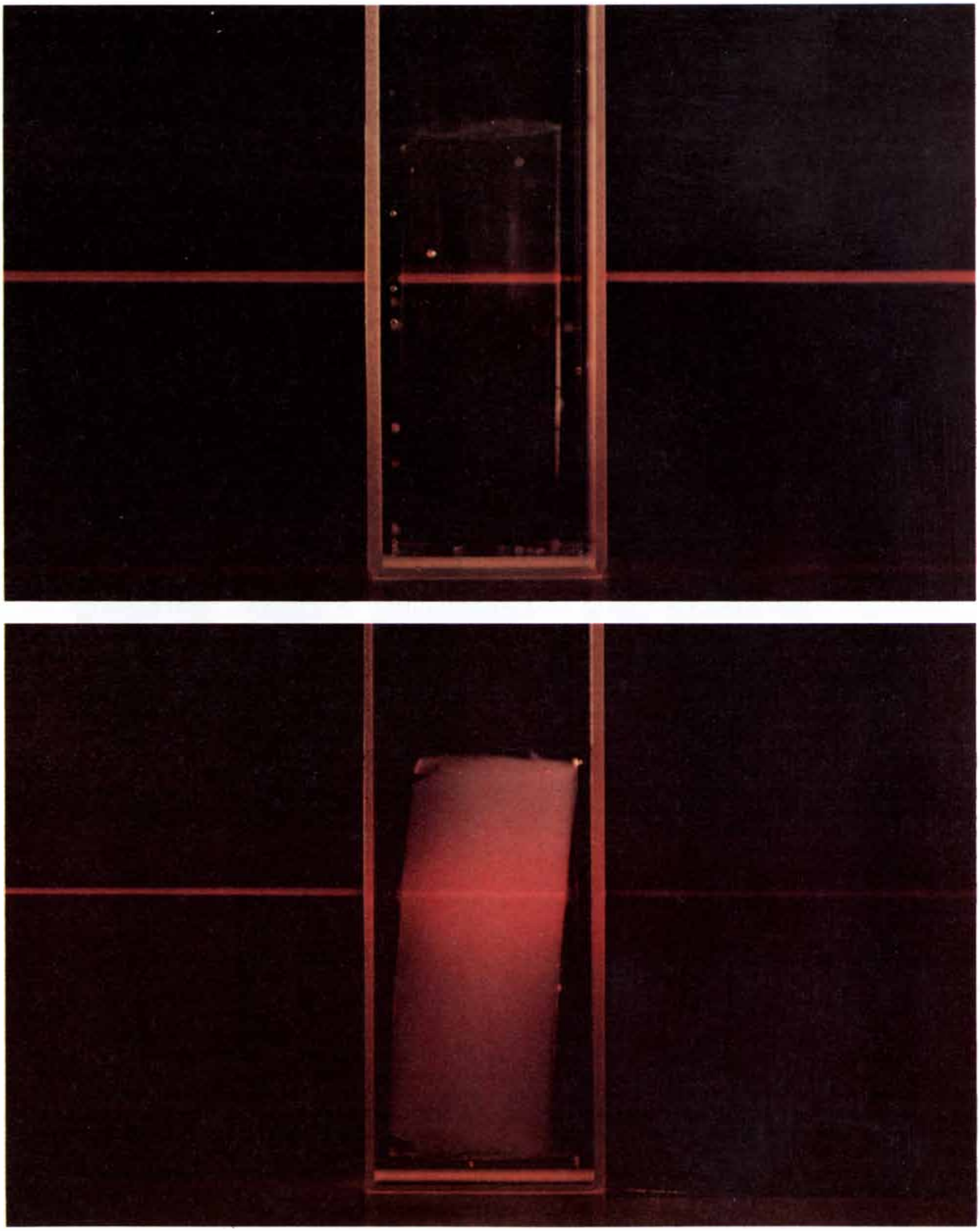
The physical chemistry of various gels has been studied intensively since the 1940's; early contributions of notable importance were made by Paul J. Flory of Stanford University. In 1973 my colleagues and I at the Massachusetts Institute of Technology applied a new technique to the investigation of gels. From measurements of laser light scattered by a gel we were able to determine some of the gel's essential properties, namely the elasticity of the polymer network and the viscous interaction of the network with the fluid. By this method we have recently discovered an entire class of unsuspected phenomena in gels. Drastic changes in the state of the gel can be brought about by small changes in the external conditions. For example, when the temperature is lowered, the polymer network loses its elasticity and therefore becomes increasingly compressible. When a certain critical temperature is reached, the elasticity falls to zero and the compressibility becomes infinite. At the same time the effective size of the pores in the polymer network increases, and at the critical temperature the pores reach macroscopic size. The gel can also swell or shrink by a factor of as much as several hundred when the temperature is varied. Under some conditions the swelling or shrinking is discontinuous, so that an infinitesimal change in temperature can cause a large change in volume. Moreover, temperature is not the only factor that can give rise to such transformations; they can also be brought about by altering the composition, the pH or the ionic strength of the solvent in which the gel is immersed or by imposing an electric field across the gel.

The varied responses of the gel to changes in external conditions can now be understood in the context of phase transitions and critical phenomena. Just as many substances can exist as a liquid or as a vapor under different circumstances, so can a gel sometimes have two phases, which are distinguished by different configurations of the polymer network. The discontinuous change in the volume and other properties of the gel constitutes an abrupt transition between the phases, analogous to the boiling of a liquid. At higher temperatures (and under various other conditions) the two phases of the gel can no longer be distinguished; in a similar way the distinction between liquid and vapor disappears at high temperature and pressure. The set of conditions where the phases of the gel first separate is the critical point of the gel, and there the polymer network exhibits large-scale fluctuations in density, pore size and other properties. The physical mechanisms underlying these changes of state are peculiar to the structure of a gel and can be understood only through an analysis of the forces acting on the polymer network. What may be most interesting, however, is that near the critical point a gel can be described in a universal mathematical language that seems to apply to all critical phenomena, even in the most disparate physical systems.

Perhaps the most familiar gel is the dessert Jello, where the network is made up of polymers derived from animal protein. The network constitutes only about 3 percent of the volume of Jello; the rest is colored, flavored and sweetened water. There are a great many other natural and manmade gels. The vitreous humor that fills the interior of the eye is a gel, and so is the material of the cornea; the synovial fluid that lubricates the joints of the skeleton is another gel. In such biological gels the liquid component allows for the free diffusion of oxygen, nutrients and other molecules, whereas the polymer network provides a structural framework that holds the liquid in place. Gels are also important intermediates in the manufacture of polymers such as rubber, plastics, glues and membranes. In chemistry and biochemistry gels are employed in the analytic methods of chromatography and electrophoresis, where molecules are separated according to the speed with which they percolate through the pores of the gel.

At M.I.T. our experiments have been carried out with gels that have a polymer matrix of cross-linked polyacrylamide. The preparation of the gel begins with two monomers: acrylamide, which is a small organic molecule that terminates in an aminocarbonyl ($-CONH_2$) group, and bisacrylamide, which consists of two acrylamide monomers that are linked through their aminocarbonyl groups. The monomers are dissolved in water, then two more substances are added to initiate a chain reaction of polymerization. The initiators are ammonium persulfate and an organic molecule called tetramethyl ethylene diamine, or TEMED.

The first step in the polymerization is a reaction between ammonium persulfate and TEMED in which the TEMED molecule is left with an unpaired valence electron. The activated TEMED molecule can combine with an acrylamide or bisacrylamide monomer; in the process the unpaired electron is transferred to the acrylamide unit, so that it in turn becomes reactive. Another monomer can therefore be attached and activated in the same way. The polymer can continue growing indefinitely (or until the supply of monomers is exhausted), with the active center being continually shifted to the free end of the chain. If the solution included only acrylamide monomers, the chain would always be straight, or unbranched. A bisacrylamide molecule, however, can be incorpo-



SCATTERED LASER LIGHT marks a change in the microscopic structure of a polyacrylamide gel. In the upper photograph the gel is at room temperature and is highly transparent, so that the light from a helium-neon laser is mainly transmitted, with only a little scattering. In the lower photograph the same gel has been cooled to about zero degrees Celsius, where it develops a milky opacity that strongly scatters the light. Cooling the gel brings it near a set of conditions called the critical point, where large-scale fluctuations arise in the local density of the polyacrylamide matrix. The fluctuations alter the refractive index of the gel and thereby scatter light, a phenomenon known as critical opalescence. Measurements of the scattered laser light reveal details about the structure of the gel, including the elasticity of the polymer network and the size of the pores in it. The photographs were made by Fritz Goro in the laboratory of the author.

rated into two chains simultaneously and forms a permanent link between them. As a result the polyacrylamide grows into a complex web of interconnected loops and branches.

The polymerization takes about 30 minutes. After a few hours the gel is removed from the glassware and soaked in water to wash out any unreacted monomers or initiators. The gel is colorless and highly transparent; it is soft and elastic, with a slippery or slimy surface.

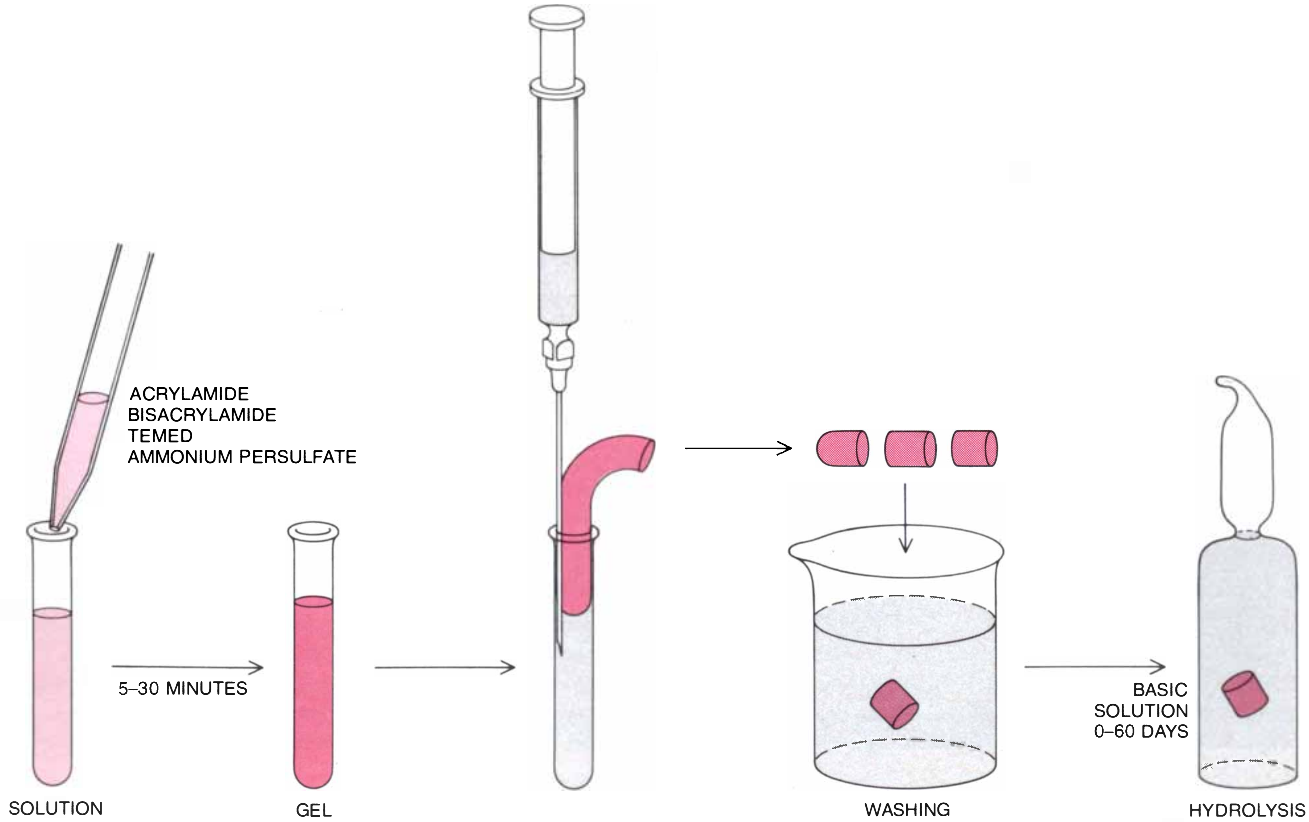
A final step in the preparation is a chemical modification of the polymer network that has a major influence on the nature of the phase transitions observed in the gel. In the native polymer every second carbon atom has an aminocarbonyl, or $-CONH_2$, side chain; in the final step some of the side chains are converted by hydrolysis into carboxyl groups, $-COOH$. Hydrolysis is the splitting of a chemical group by the agency of a water molecule; it is brought about by immersing the gel for a period of days or weeks in a basic solution, one with a *p*H of about 12. The number of side chains hydrolyzed depends on the hydrolysis time. If the gel is not hydrolyzed at all, of course, none of the groups are converted. The maximum hydrolysis, which is attained after about 60 days of immersion, transforms roughly a fourth of the side chains.

The carboxyl group is the functional unit of an organic acid. In solution some of the groups spontaneously ionize to yield H^+ and COO^- ions. In a highly ionized polyacrylamide gel the fraction of the carboxyl groups ionized at any moment is about 25 percent. The positively charged hydrogen ions enter solution in the interstitial fluid, leaving the polymer network with a negative charge. The interaction of these separated electric charges can have a determining effect on the physical properties of the gel.

The phase transitions observed in polyacrylamide gels can be illustrated by the following series of experiments. A batch of identical gels is prepared and specimens are hydrolyzed in a basic solution for various periods ranging up to 60 days. In this state the gels are swollen with water; indeed, a fully hydrolyzed gel has 30 times its original volume. The gels are then put in mixtures of acetone and water and observed for several days. Depending on the hydrolysis time and the proportions of acetone and water, some of the gels are found to have shrunk.

Consider first a gel for which the hydrolysis step has been omitted entirely (zero hydrolysis time). If a specimen is put in a solution with a low concentration of acetone, the gel remains swollen. If the proportion of acetone is somewhat greater (say 20 percent), the gel shrinks slightly. When the concentration of acetone reaches 60 percent, the gel is quite noticeably contracted, perhaps by a factor of 10. By testing many specimens in this way with many water-acetone mixtures, it is possible to construct a graph showing the change in the volume of the gel as a function of acetone concentration. For the gel without hydrolysis the graph is a smooth curve. The relation between volume and acetone concentration is not one of strict proportionality (in which case the graph would be a straight line), but at least the curve has no sharp turns or other discontinuities.

Suppose the same sequence of trials is now carried out with other gels that have been exposed to progressively longer hydrolysis. For a gel that has been hydrolyzed for two days the swelling curve is still smooth, but it has an inflection point, where the downward-tending curve briefly becomes horizontal. Again the gel shrinks somewhat as



PREPARATION OF A POLYACRYLAMIDE GEL begins with monomers, or repeating units, of the polymer, acrylamide and bisacrylamide, and with two substances that serve to initiate the formation of the polymer: tetramethyl ethylene diamine (TEMED) and ammonium persulfate. When the starting materials are dissolved in water, a cross-linked network of polyacrylamide forms in a matter of minutes. The polymer is dispersed throughout the water, so that the gel fills the vessel in which it is made. The gel is removed from the glassware and washed. A final step called hydrolysis has an important influence on the nature of the phase transitions observed in the gel.

the amount of acetone in the solvent is raised from zero to about 40 percent. Near the inflection point, however, which comes at an acetone concentration of 42 percent, the gel becomes highly sensitive to changes in the acetone-water ratio. A small change in the acetone concentration results in a large change in volume. In going from 40 percent to 45 percent acetone, for example, the gel might shrink to a tenth its former volume. Further increases in acetone concentration bring only a slight additional contraction.

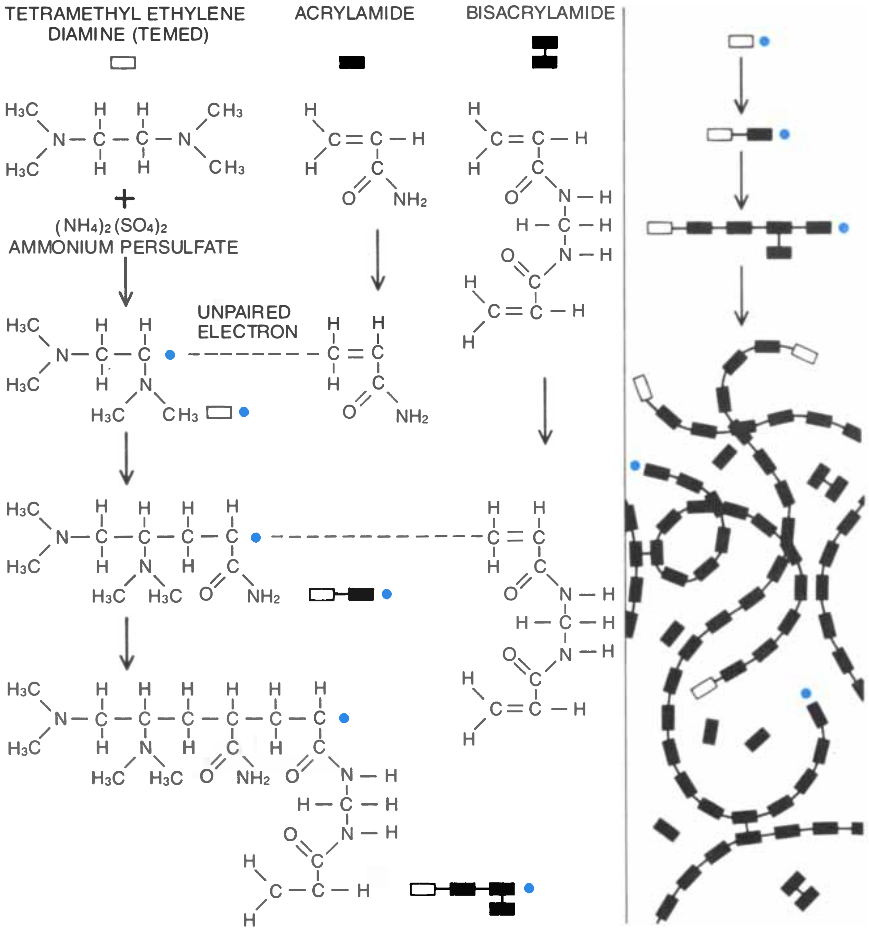
With still longer hydrolysis the sensitivity of the gel to changes in the acetone concentration becomes infinite. The swelling curve then has a discontinuity representing a discrete phase transition. After four days of hydrolysis, for example, the gel shrinks only gradually until the acetone concentration reaches about 40 percent. At 42 percent, however, the gel collapses abruptly. There is a finite reduction in volume (again by a factor of 10 or more) in response to an infinitesimal change in concentration. No matter how small the increment in concentration is, the same shrinkage is observed. Hence the graph has a "step" that cannot be fitted smoothly into the curve.

The abrupt collapse of the gel becomes still more dramatic after longer hydrolysis. A gel that has been hydrolyzed for 60 days can shrink by a factor of more than 350 when the acetone concentration passes from low to high. Almost all of this contraction takes place in a discrete, or discontinuous, transition. The graph, or the swelling curve, approaches the limiting form of a square step.

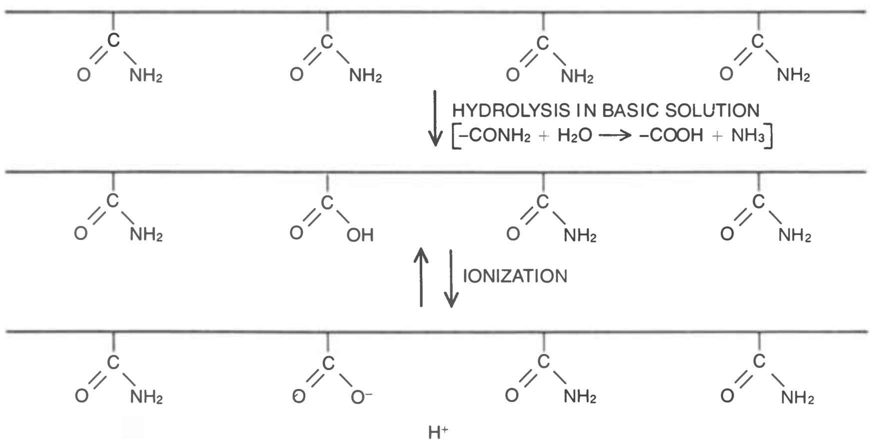
The shrinking of a gel on exposure to a high concentration of acetone is fully reversible. If a collapsed gel is put in a solvent with less acetone, the gel expands again, and if it is put in pure water, it continues expanding until its former bulk is restored. The same discontinuity can be observed in the reverse transition: a fully hydrolyzed gel swells abruptly when the acetone concentration falls below the transition value.

It is important to understand the dynamics of these events. The changes in gel volume are not rapid or sudden. With a specimen a centimeter long it takes about 30 minutes for a new solvent mixture to thoroughly permeate the gel by diffusion. The shrinking or swelling that follows a change in acetone concentration can take much longer to reach completion, perhaps several days. The transitions are abrupt only in the sense that they can be provoked by an arbitrarily small change in the composition of the solvent.

Throughout this discussion it has been tacitly assumed that the concentration of acetone is the only parameter



POLYMERIZATION OF THE GEL proceeds by way of a chain reaction. The first step is the activation of TEMED by ammonium persulfate, which leaves the TEMED molecule with a reactive unpaired electron. The TEMED can combine with acrylamide, which is thereby activated in turn. As the chain of acrylamide units grows, the active site shifts to the free end. Bisacrylamide, which consists of two acrylamide units joined through their $-CONH_2$ groups, can be incorporated into two growing chains. Hence the presence of bisacrylamide leads to the formation of cross-links between chains. With an abundance of cross-links the polymer of a gel has a topologically complex configuration, with loops, branches and interconnections.



HYDROLYSIS OF THE GEL alters some of the $-CONH_2$ functional groups that appear as side chains on the polymer. The hydrolysis is done in a basic solution (one with a *p*H of about 12), where a $-CONH_2$ group can be converted into a carboxyl group ($-COOH$). The fraction of the side chains converted depends on how long the hydrolysis is allowed to continue; the maximum conversion, about 25 percent, is approached after 60 days. The carboxyl group functions as an organic acid, and some groups spontaneously ionize. Positively charged hydrogen ions (H^+) enter solution, and remaining $-COO^-$ groups give polymer strands a negative charge.

being varied. In particular it has been assumed that the experiments are carried out at a fixed temperature. All the same phenomena can be observed by keeping the acetone-water mixture constant and varying the temperature. At high temperature the gel expands and at low temperature it shrinks. Moreover, if the degree of hydrolysis is great enough, the transition is discontinuous. If a gel that has been hydrolyzed for four days is put in a solvent with 42 percent acetone, the transition is observed at room temperature, or about 22 degrees Celsius. Of course, there is nothing special to distinguish the particular conditions of 42 percent acetone and 22 degrees C.; these conditions are a coupled pair that jointly determine the transition point, just as 100 degrees C. and a pressure of one atmosphere jointly determine the boiling point of water. If the acetone concentration were increased, the transition temperature would rise in compensation.

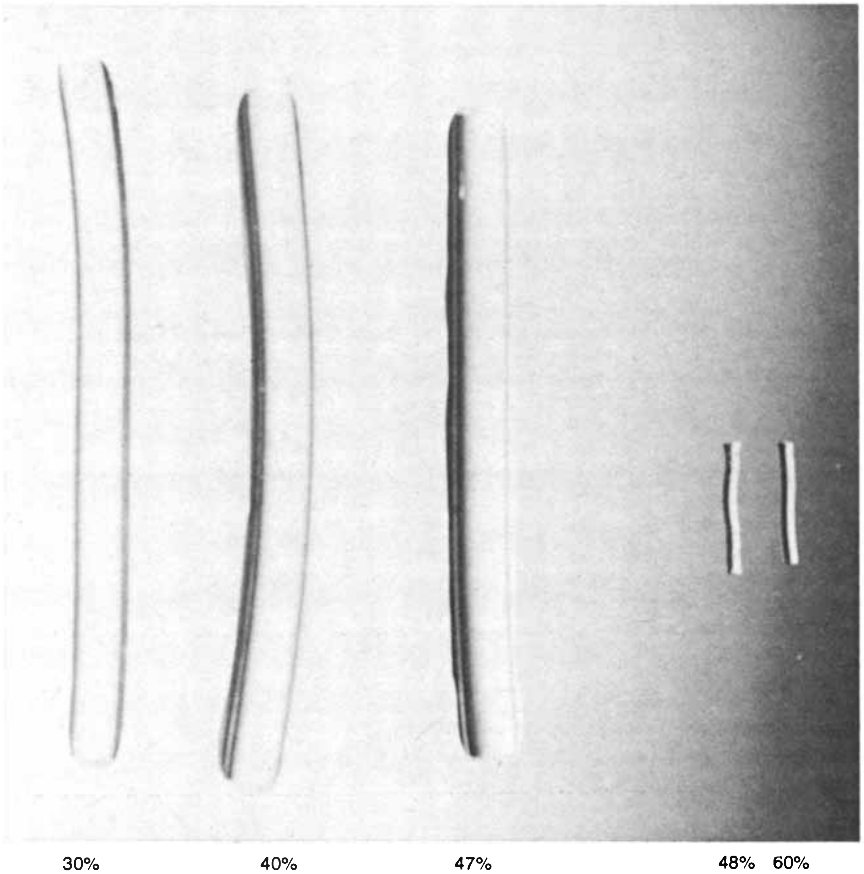
What causes these changes in the volume of a gel when it is warmed or cooled or moved from one solvent to another? Why do the magnitude of the change and the abruptness of the transition depend on the degree of hydrolysis? In order to answer these questions it is necessary to examine all the forces, or pressures, that act to expand or shrink the polymer network. Three such forces have been identified: the rubber elasticity, the polymer-polymer affinity and the hydrogen-ion pressure. The total pressure acting on the gel is the sum of these three components. I shall call this sum the osmotic pressure of the gel, because it determines whether the gel tends to take up fluid or to expel it.

The property of a gel called the rubber elasticity derives from the elasticity of the individual polymer strands, or in other words from the resistance the strands offer to either stretching or compression. (A similar mechanism accounts for the elasticity of rubber, hence the name.) A single linear strand of the polymer can be represented as a chain of rigid but freely jointed segments, which are in constant motion because of their thermal energy. Suppose such a chain is held with its ends fixed. If the ends are far apart, so that the chain is extended almost to its full length, the thermal agitation of the segments gives rise to a force that on the average tends to pull the ends inward. If the ends are held close together, with the length of the chain crumpled between them, the random motions of the segments tend to push the ends apart. Between these extremes there must be some end-to-end distance where the average force is zero and the strand is neither in tension nor in compression. It can be shown mathematically that this equilibrium is attained at an intermediate degree of extension, namely when the ends are separated by a distance equal to the square root of the number of segments multiplied by the length of a segment.

For a strand in any given configuration the strength of the rubber elasticity depends on how actively the polymer segments are moving. Because the motions are thermally induced they are proportional to the absolute temperature; hence the restoring force that tends to extend or contract the chain is also proportional to the temperature. It should be emphasized that the temperature has no effect on the direction of the force, which is determined solely by whether the strand is stretched or compressed compared with the equilibrium length. The temperature influences only the magnitude of the force.

The polymer of a gel is not a single linear strand but a topologically complex network. Moreover, the detailed configuration of the polymer generally is not known, and so the force arising from the rubber elasticity cannot be calculated directly. Nevertheless, certain properties of the force can be deduced. If the gel is swollen, so that most of the polymer strands are stretched beyond their equilibrium length, the rubber elasticity tends to make the gel contract. Conversely, a collapsed gel tends to expand under the influence of the rubber elasticity. In this way a directed force of tension or compression along a single polymer strand gives rise to a pressure in the gel as a whole. By convention the pressure is positive if it tends to expand the gel and negative if it tends to make the gel shrink. Whereas the sign of the pressure depends only on the volume of the gel, its magnitude is determined by the temperature. Raising the temperature strengthens the tendency of a collapsed gel to expand and the tendency of a swollen gel to contract.

The second force acting on the gel, the polymer-polymer affinity, can be traced to an interaction between the polymer strands and the solvent. Such interactions can be either attractive or repulsive, depending mainly on the electrical properties of the molecules. Where the



ABRUPT COLLAPSE of an acrylamide gel is interpreted as a phase transition. All the gels were initially the same size and all were prepared identically and subjected to hydrolysis for 15 days. The specimens were then placed in mixtures of acetone and water with varying concentrations of acetone. It is apparent that solvents with less than 47 percent acetone have little effect on the volume of the gel. In any mixture that exceeds this concentration, however, the gel is reduced to a small fraction of its initial volume. The transition is a discrete one in that a large change in volume results from an arbitrarily small change in the composition of the solvent.

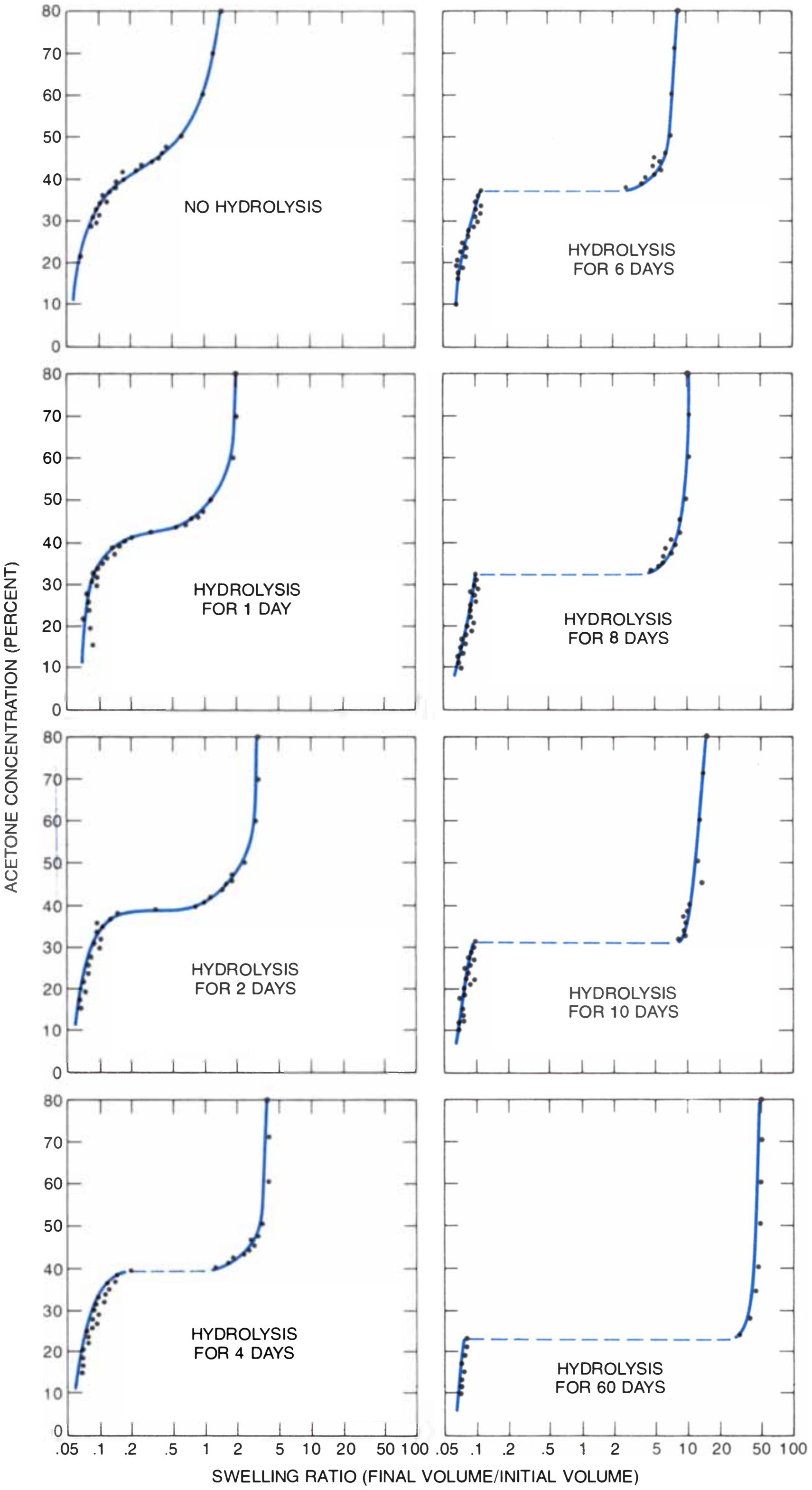
interaction is attractive the polymer can reduce its total energy by surrounding itself with solvent molecules; where the interaction is repulsive the solvent is excluded.

In an acrylamide gel where the solvent is a mixture of water and acetone a polymer strand has a greater affinity for other polymer strands than it has for the solvent molecules. As a result the polymers tend to coagulate and the polymer-polymer affinity creates a negative pressure that in the absence of all other forces would collapse the gel. This negative pressure is independent of the temperature, but it does depend on the composition of the solvent. Acrylamide is less soluble in acetone than it is in water, and so the pressure becomes more negative as the concentration of acetone increases.

The negative pressure arising from the polymer-polymer affinity also depends on the volume of the gel. The affinity is a short-range force: it is fully effective only if two polymer segments are touching. The probability of such direct contact is inversely proportional to the square of the volume. Therefore as the gel contracts, the attractive, negative pressure increases; as the gel expands, the resistance to further expansion abates.

The third and final contribution to the osmotic pressure of the gel, the hydrogen-ion pressure, is associated with the ionization of the polymer network, which releases an abundance of positively charged hydrogen ions (H^+) into the gel fluid. If these ions were the only charges present, they would strongly repel one another; indeed, the force of electrical repulsion would dwarf all the other forces and would quickly destroy the gel. Actually, however, the positive ions are immersed in a sea of negative charges attached to the polymer network, with the result that the gel as a whole maintains exact electrical neutrality. The mutual repulsion of the hydrogen ions is effectively screened by the background of negative charges, and the ions act much as if they were neutral particles.

The hydrogen ions nonetheless do give rise to a pressure. As long as the ions remain within the volume of the gel they move freely, like the molecules of a gas. They cannot leave the gel, however, because of the strong electrical attraction between the ions and the negatively charged polymer. The hydrogen ions therefore act like a gas in a confining vessel, and their random motions generate a positive pressure, just as the moving molecules in a bottle of compressed air give rise to a pressure. The pressure is directly proportional to the absolute temperature, that is, if the volume of the gel is held constant, raising the temperature increases the hydrogen-ion pres-



SWELLING CURVES record the change in the volume of a gel when it is placed in solvents with varying concentrations of acetone. As the acetone content increases, the gel invariably shrinks somewhat, but the extent of the change in volume and the shape of the swelling curve depend strongly on the degree to which the gel polymer is ionized. For a gel without hydrolysis the curve is smooth and continuous and the total change in volume is roughly tenfold. Gels that have been hydrolyzed longer show a larger volume change. Moreover, after two days of hydrolysis the swelling curve develops an inflection point, where the slope is horizontal; with further hydrolysis the inflection point broadens into a discontinuity. At the discontinuity an infinitesimal increase in acetone concentration causes a large reduction in volume; the gel collapses. The collapse is a discrete transition, where the gel separates into a condensed phase and a more rarefied one. After 60 days of hydrolysis the volume changes by a factor of about 350.

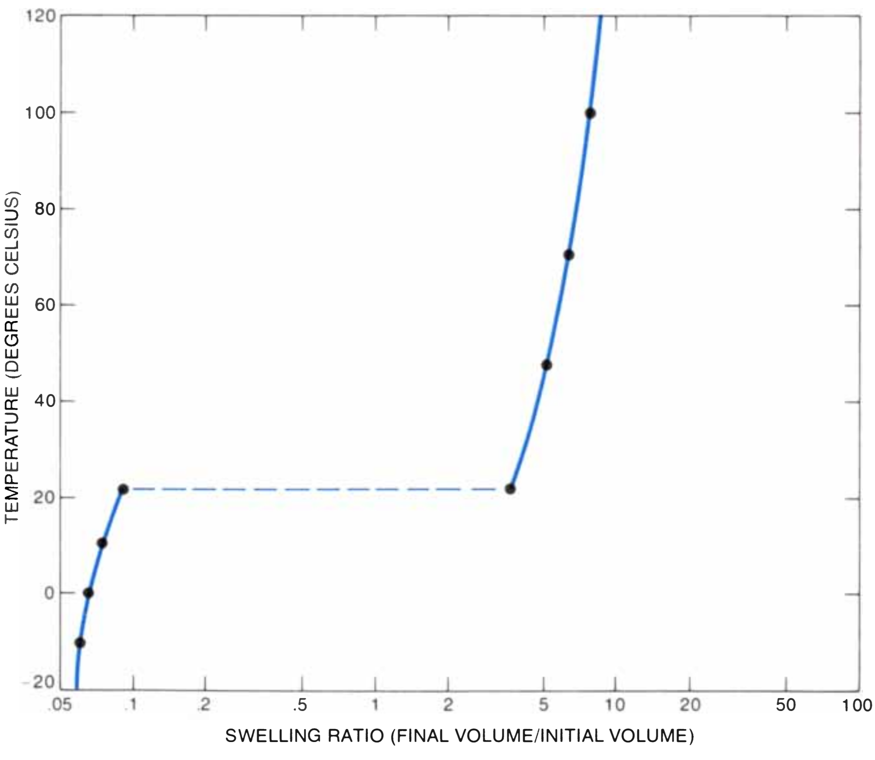
sure. It is inversely proportional to the volume, since expanding the gel dilutes the gas.

The effect of each of the three components of the osmotic pressure is readily analyzed in isolation. What makes the properties of the gel interesting is competition among the components. It is the changing balance of forces that gives rise to phase transitions. Indeed, from a quantitative analysis of the three forces it is possible to construct a phase diagram for the gel.

The diagram is constructed by calculating each of the three component pressures and then adding them to get the total osmotic pressure for many possible combinations of volume and either temperature or acetone concentration. This information is presented in a graph of osmotic pressure v. volume. A series of curves are plotted on the graph, each curve representing all the possible combinations of pressure and volume at a given temperature or a given acetone concentration. The curves are called isotherms, meaning "equal temperature."

An essential principle in interpreting the isotherms is that whenever possible the gel will adjust its volume so that the total osmotic pressure is zero. If the pressure is initially positive, the gel takes up fluid and expands (assuming that excess fluid is available and that there is no constraint on the swelling of the gel). If the pressure is negative, the gel expels fluid and shrinks. The swelling or shrinking continues until an equilibrium is reached, where the various positive and negative pressures exactly cancel one another.

Consider an isotherm that describes the state of the gel at a comparatively high temperature and at a low concentration of acetone. Suppose the gel is initially contracted to a small volume. Because of the low acetone concentration the polymer-polymer affinity, which is always a negative pressure, is small. Because of the high temperature the hydrogen-ion pressure, which is always positive, is large. The rubber elasticity is also positive (because the gel is shrunken) and large (again because the temperature is high). This combination of two large positive pressures and one small negative pressure clearly yields a total osmotic pressure that is strongly positive. The gel therefore tends to expand, but as it does so the osmotic pressure decreases monotonically. The loss of pressure can be attributed to the expansion itself. As the volume grows larger the density of hydrogen ions decreases and so does the pressure they generate. At the same time the polymer network becomes more extended, reducing the rubber elasticity. Ultimately the rubber elasticity falls to zero and then with further swelling becomes negative, as the polymer strands are stretched beyond their equilibrium length. When the three pressures have a sum of zero, the expansion stops. Thus a point representing the state of the gel travels along the isotherm, through various combinations of pressure and volume, until it comes to rest at the intercept of the isotherm and the zero-pressure axis.

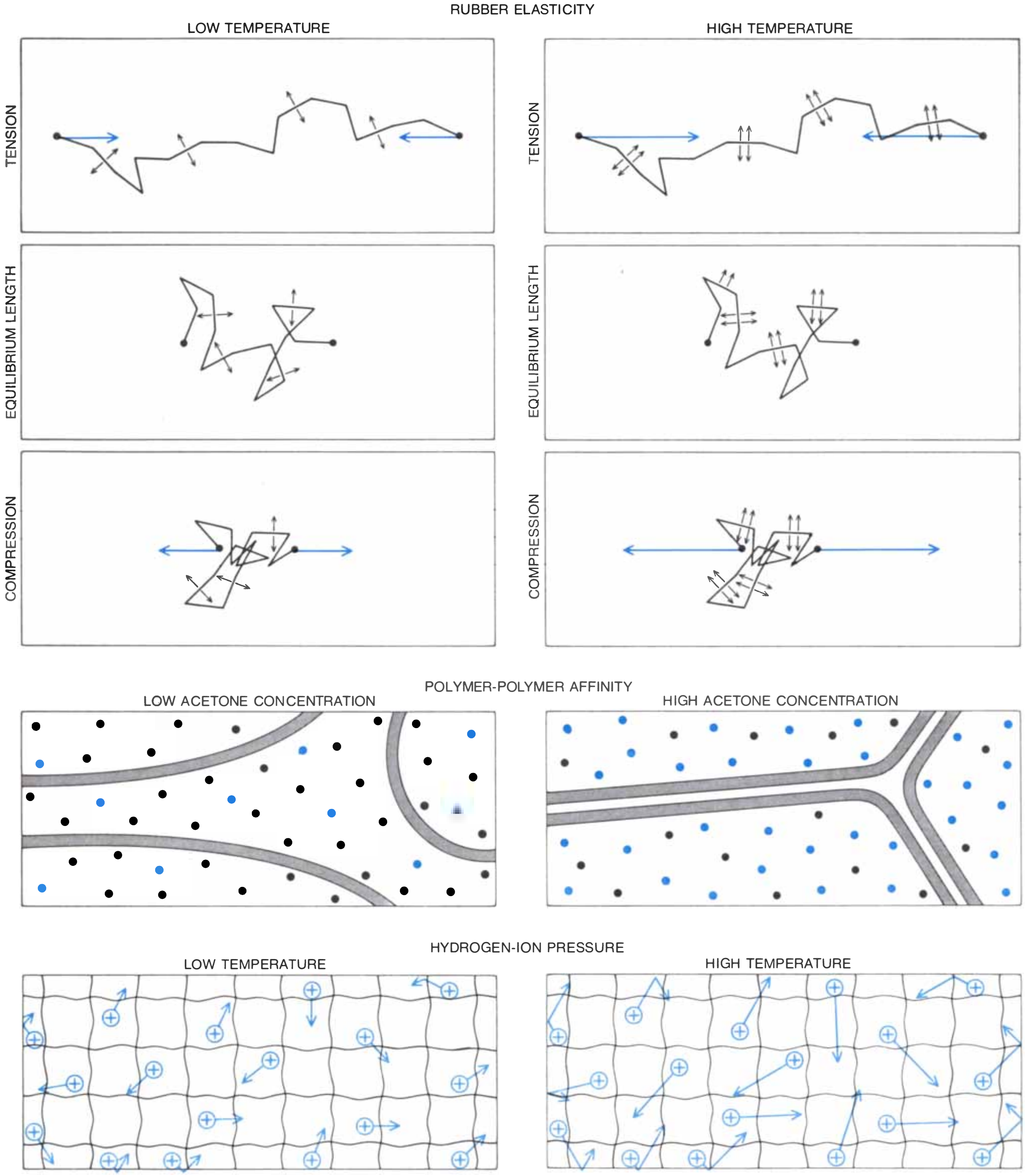


CHANGE OF TEMPERATURE elicits a phase transition when a gel is immersed in an acetone-water mixture of fixed composition. Here the mixture is 42 percent acetone, and the gel collapses when it is cooled to about room temperature: 22 degrees C. Such transitions are fully reversible; when the gel is warmed, it swells at 22 degrees and regains its initial size. Gel was hydrolyzed for eight days; longer hydrolysis would exaggerate discontinuity in swelling curve.

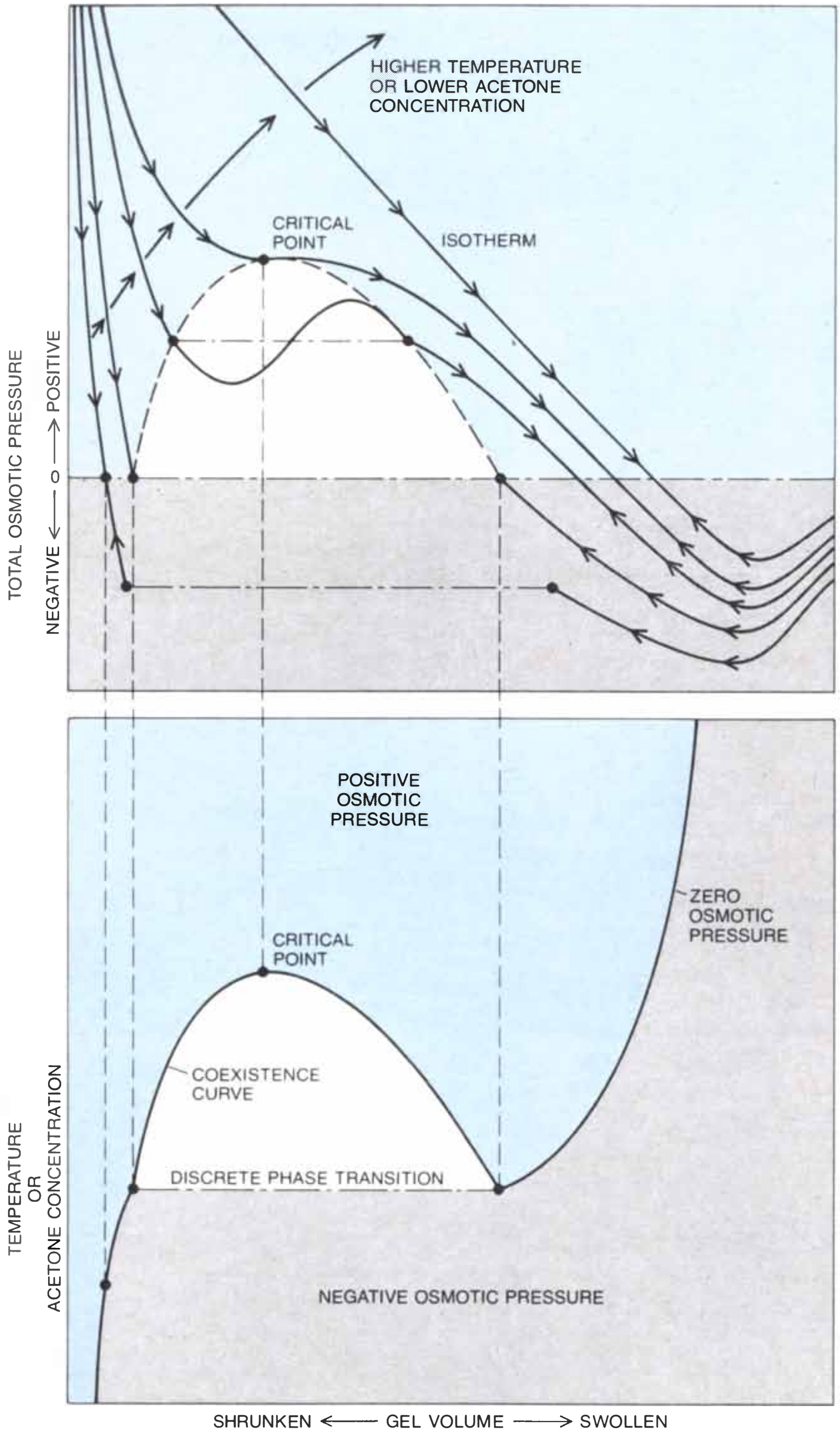
Another isotherm can be generated by changing one of the initial conditions, namely by reducing the temperature or by increasing the acetone concentration. If the temperature is changed, the hydrogen-ion pressure and the rubber elasticity are somewhat less positive; if the solvent composition is changed, the polymer-polymer affinity is somewhat more negative. In either case the net effect is to reduce slightly the positive osmotic pressure in the initial, contracted state. The gel still tends to expand, but the pressure at any given volume is lower than it was at the higher temperature (or lower acetone concentration). If the conditions happen to be exactly those that define the critical point of the gel, the curve has an inflection point, where the pressure remains momentarily constant as the volume continues to increase.

A further reduction in temperature or increase in acetone content now leads to an isotherm with a clearly defined phase transition. In the initial collapsed state the net osmotic pressure is still positive, but it is lower than it was in the preceding cases. The gel therefore again expands, and as it does so the total pressure begins to fall. Before the pressure reaches zero, however, the theoretical isotherm develops a curious wiggle. The pressure reaches a local minimum, then with further expansion of the gel the total osmotic pressure increases for a time before it reaches a local maximum and resumes its downward trend. The region between the local minimum and the local maximum is one of negative compressibility: applying a greater pressure to the polymer network would cause it to expand. This bizarre condition is clearly unstable. Suppose a thermal fluctuation resulted in an infinitesimal swelling of one area of the gel; the expansion would add to the positive osmotic pressure, which would lead to further expansion, and so on. Conversely, if a small region began to shrink spontaneously, the osmotic pressure would be reduced below that of the surrounding regions, and so the contraction would be amplified.

How can the region of negative compressibility be explained? The region is dominated by just one of the three component pressures, namely the polymer-polymer affinity, and it is the properties of this force that cause the strange reverse-slope excursion of the isotherm. Throughout most of the length of the isotherm the rubber elasticity and the hydrogen-ion pressure are



OSMOTIC PRESSURE OF A GEL, which determines whether the gel tends to shrink or to expand, is the sum of three component pressures. The rubber elasticity derives from the resistance of the individual polymer strands to stretching or compression. A strand can be represented by a chain of rigid, jointed segments, each of which is in constant thermal motion. If the chain is stretched almost taut, the random movements of the segments give rise to a tension that pulls the ends of the chain inward. If the chain is compressed into a compact ball, the force is directed outward. At an intermediate length of chain the average force is zero. Because thermal agitation is the ultimate root of the force the magnitude of the rubber elasticity is proportional to the absolute temperature. The polymer-polymer affinity arises from an interaction of the polyacrylamide strands with the interstitial fluid of the gel. The polymer is more soluble in water (*black dots*) than it is in acetone (*colored dots*). When the acetone concentration is high, two strands can reduce their total energy by coagulating and thereby excluding the solvent from the space between them. The polymer-polymer affinity is always a negative pressure, favoring contraction of the gel. Its magnitude increases with the acetone concentration but is independent of the temperature. The third component force is the hydrogen-ion pressure. Within the gel the positive charge of the hydrogen ions is effectively neutralized by the background of negative charges on the polymer network. The ions are therefore able to move freely, like the molecules of a gas, but they are confined to the interior of the gel. Random thermal motions give rise to a pressure that, like a gas pressure, is proportional to the absolute temperature. The hydrogen-ion pressure also depends on the number of ions that are present, and hence on the hydrolysis time.



PHASE DIAGRAM of a gel is constructed by calculating the osmotic pressure for many combinations of volume and either temperature or acetone concentration. In the upper diagram this information is presented as a series of isotherms, which are curves that specify all possible combinations of pressure and volume at a given temperature (or acetone concentration). The highest-temperature isotherm defines a simple relation: as the gel expands, the pressure decreases monotonically. At lower temperatures, however, the isotherms develop first an inflection point with a horizontal slope and then an *S*-shaped wiggle. The reversed slope of the curve in the *S*-shaped region signifies a negative compressibility: as the gel expands, the pressure rises, driving further expansion. Because this condition is unstable the *S* bend must be replaced by a horizontal line, which cuts off equal areas in the two loops of the *S*. Whenever possible the gel will adjust its volume to reduce the osmotic pressure to zero; a point representing the state of the system will therefore travel along an isotherm until it reaches the zero-pressure axis. The lower diagram was derived from the upper, but it gives the volume of the gel as a function of temperature or acetone concentration. Boundary of the gray area is the volume at zero osmotic pressure; it is equivalent to the swelling curve and includes a discrete phase transition. At some positive pressures two phases coexist; at the critical point they are indistinguishable.

positive but are consistently decreasing as the gel expands. The polymer-polymer affinity is negative but is also decreasing as the polymer strands move farther apart. What distinguishes the region of negative compressibility is that here the polymer-polymer affinity becomes less negative faster than the other two pressures become less positive. As a result the net positive pressure increases.

The theoretically derived isotherms give an accurate account of the balance of pressures, but in the region of negative compressibility they do not describe what actually happens in the gel. The real course of events was first deduced by James Clerk Maxwell for liquid-vapor phase transitions in fluids. As a gel expands and loses pressure it never reaches the local minimum in the isotherm; instead the situation becomes unstable as soon as a state is available that has a larger volume and the same pressure. Maxwell showed that the entire *S*-shaped region of the curve should be replaced by a straight horizontal line. The line is to be drawn in such a way that it cuts off loops of equal area for the local minimum and the local maximum. When the gel reaches the Maxwell line, it separates into two domains, one comparatively swollen and one shrunken.

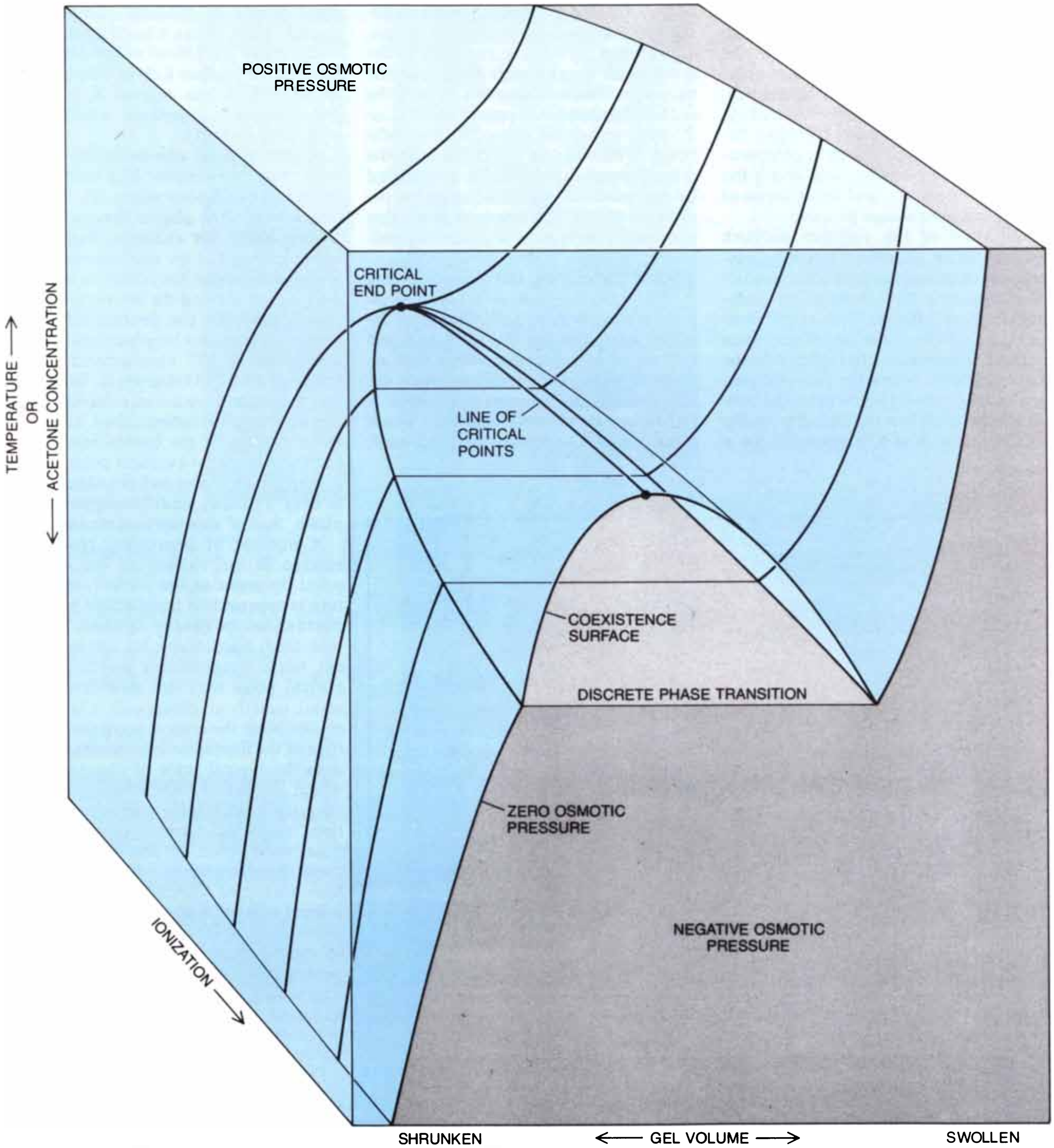
All the information needed for a complete account of the equilibrium properties of the gel is to be found in the isotherms. For example, it is possible to derive from these theoretical curves the swelling curve that was discussed above in an experimental context. The swelling curve gives for any temperature or any acetone concentration the gel volume at which the total osmotic pressure is zero. Hence the swelling curve is defined as the set of intercepts of all possible isotherms and the zero-pressure axis. The curve includes a discrete phase transition at the temperature or acetone concentration where the Maxwell horizontal line falls along the zero-pressure axis.

The complete phase diagram for the gel has three regions. Below the swelling curve the gel has a negative osmotic pressure and is always unstable; it will expel fluid and shrink until it reaches the zero-pressure swelling curve. Another region represents states with a positive osmotic pressure, which are stable only if there is no excess fluid the gel could absorb in order to expand. Between these realms is the domain of dual stability, where some part of the gel shrinks and another part swells. The boundary of this region is the coexistence curve for the two phases. The difference in volume between the two phases is greatest along the swelling curve, and it decreases steadily at higher temperature. Ultimately the coexistence curve reaches a maximum temperature where the two phases are identical, or in other words where they cease to exist as distinguishable phases. That is the critical point of the gel.

One additional factor that was observed to have an important influence on phase transitions in gels has not yet been included in this analysis: the degree of ionization, determined by the hydrolysis time. The discontinuous volume change brought about by the phase transition was found to range from zero (continuous change) to 350-fold depending on the ionization. What is the mechanism underlying this effect?

Examination of the phase diagram suggests a part of the explanation. Evidently the lower the temperature at which the transition takes place (or the higher the concentration of acetone), the larger the change in volume. If the transition temperature is equal to the critical temperature, there is no discrete change in volume but only an inflection point in the swelling curve. As the transition temperature sinks below the critical temperature the Maxwell horizontal line spans an increasing distance between the two branches of the coexistence curve, so that the volume change grows.

It therefore appears that the effect of ionizing the polymer network is to depress the phase-transition temperature below the critical temperature. In this



THREE-DIMENSIONAL PHASE DIAGRAM **includes the influence of ionization on the state of the gel. The front surface of the diagram is equivalent to the lower phase diagram in the illustration on the opposite page, and it corresponds to a fully ionized gel. In a less ionized gel the volume change at the phase transition is smaller, the transition temperature is higher and closer to the critical temperature, and the region in the diagram where the two phases are able to coexist is smaller. Ultimately the transition temperature is equal to the critical temperature, and the discrete phase transition of the gel is abolished. This set of conditions is known as the critical end point.**

way as the gel is cooled or the acetone concentration is increased the phase transition is observed later, but it entails a larger change in volume. The question is, then: How can ionization defer the phase transition to lower temperatures or higher acetone concentrations?

The answer is to be found by returning to an analysis of the three components of the osmotic pressure. If the gel is entirely without ionization, one of these components does not exist: there are no hydrogen ions and hence there is no hydrogen-ion pressure. Under these conditions, as the temperature falls the negative pressure of the polymer-polymer affinity is opposed only by the positive pressure of the rubber elasticity. The negative pressure can therefore become the dominant one at a comparatively high temperature, well above the critical temperature, and the collapse of the gel is a continuous process.

Ionization of the polymer network contributes an additional positive pressure, which augments the rubber elasticity in opposing the effects of the polymer-polymer affinity. The extra pressure is able to keep the gel inflated even at lower temperatures or higher acetone concentrations, where the polymer-polymer affinity would overwhelm the rubber elasticity. When the collapse finally does come, it is at a temperature or a concentration below the critical point, and so the transition is discontinuous.

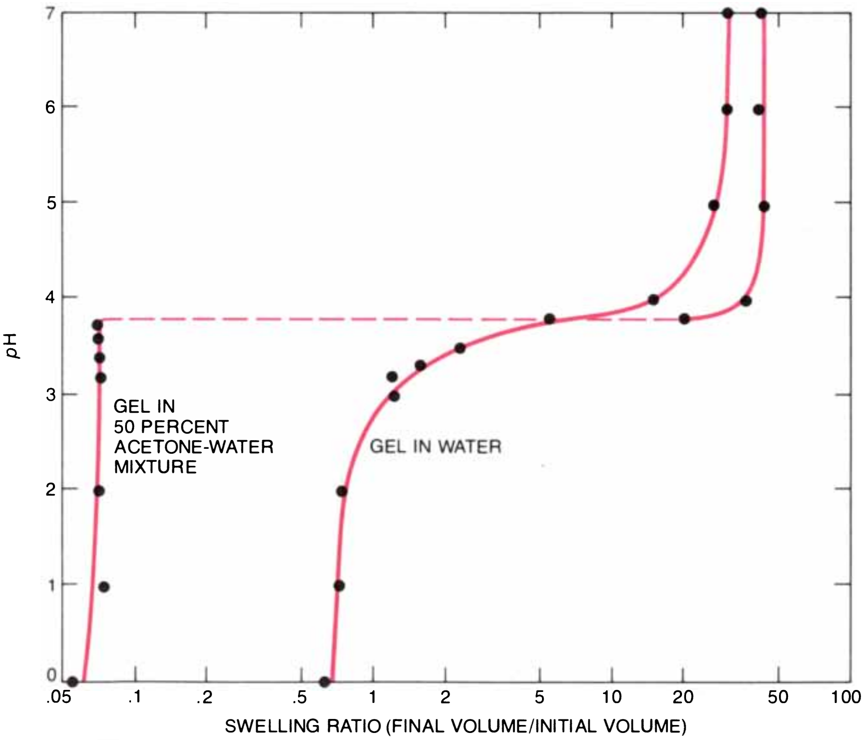
The various influences on the state of the gel can be summarized by constructing a three-dimensional phase diagram in which the state of the gel is defined for various combinations of ionization, volume and temperature (or acetone concentration). The swelling curve and the coexistence curve of the two-dimensional diagram become surfaces in the three-dimensional one and the critical point becomes a line of critical points. In going from a much-ionized gel to a less-ionized one the volume enclosed by the coexistence curve tapers to zero as the transition temperature rises. Where the enclosed volume disappears there is no discrete transition but only a critical point. With still less ionization even the critical point disappears; it is engulfed by the unstable region of negative osmotic pressure. The terminus of the line of critical points is the critical end point.

Phase transitions and critical points are found in a great diversity of physical systems besides gels. The most familiar examples are the solid-to-liquid and liquid-to-vapor transitions that almost all substances undergo when the temperature and pressure are varied. A ferromagnetic material exhibits a phase transition of a somewhat different kind: at high temperature the magnetic moments of the atoms are randomly oriented and the material is unmagnetized, but when the temperature falls to a particular value, the atomic moments begin to line up and the material develops a magnetization. Another phase transition can be observed in a mixture of two fluids, such as aniline and cyclohexane. At high temperature the components are well dispersed and the mixture is homogeneous; at the transition temperature separate phases of different composition appear. Helium has a number of interesting phase transitions at low temperature. At 4.2 degrees Kelvin it becomes a liquid and at two degrees K. another phase begins: a superfluid, which flows with zero viscosity.

A critical point can be found for all such phase transitions. It is defined by the set of conditions where the distinction between the phases disappears. In boiling water, for example, liquid and vapor coexist but are nonetheless clearly distinguishable: they differ in density. As the pressure and the temperature are raised, however, the density difference between the phases becomes smaller. At a pressure of 217 atmospheres and a temperature of 647 degrees K. the liquid and the vapor have the same density and can no longer be distinguished. Here the two branches of the coexistence curve for water merge at a critical point. At all higher temperatures and pressures there is only a single, undifferentiated fluid phase, that of supercritical steam.

A number of interesting properties emerge in the vicinity of the critical point. In water as the critical temperature is approached from above random fluctuations in density become important. Such fluctuations are always present, but at temperatures well above the critical point they are short-lived and small, usually involving only a few molecules. Near the critical point the amplitude of the fluctuations increases, and so does the typical scale of distance over which the fluctuations extend. Moreover, the processes by which the fluctuations "relax," or fade into the background fluid, slow down. At the critical point itself the characteristic length that defines the size of a fluctuation diverges to infinity and the relaxation rate becomes infinitely slow. These critical phenomena can be observed directly. When the density fluctuations approach macroscopic size, they become effective scatterers of light. As a result the water becomes milky and almost opaque, an effect called critical opalescence.

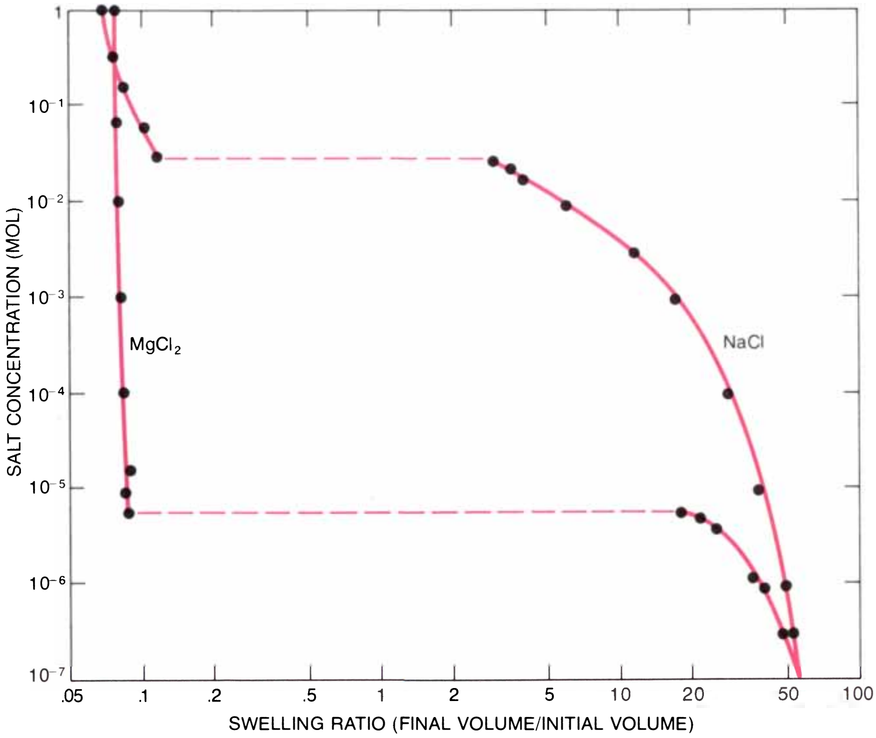
Corresponding critical phenomena are observed in other systems that have a phase transition. In a ferromagnet it is fluctuations in the local direction of magnetization that grow in scale and slow down. In a mixture of fluids the fluctuations are local variations in com-



ACIDITY OF THE SOLVENT is another parameter that can be employed to control the state of a polyacrylamide gel. Adding acid, and thereby reducing the *p*H, causes the gel to collapse. When the solvent is water, the volume change is continuous, but in a 50-percent-acetone solution there is a discrete transition in the state of the gel. The *p*H of the solvent alters the osmotic pressure of the gel by reducing the ionization of the carboxyl groups on the polymer chain.

position. A fundamental and somewhat surprising recent finding is that even in very different physical systems the critical phenomena can often be encompassed by the same mathematical description. Indeed, it turns out that all the systems known to have critical points can be grouped in just a few classes, and all the systems in the same class are mathematically identical. The critical points associated with the liquid-vapor transition, with a binary mixture of fluids and with certain ferromagnets all fall in the same class, and so at an underlying level all three can be represented by the same system of equations. In particular the development of the critical fluctuations follows the same mathematical laws, even though the fluctuations concern density in the first instance, concentration in the second and magnetization in the third.

For now it seems that gels should be assigned to this same class, although the assignment will have to be confirmed by further experiments and theoretical work. Regardless of the classification the essential point remains that all the critical phenomena observed in other systems have their counterparts in a gel, including even the critical opalescence.



DISSOLVED SALTS can also bring on a collapse of the gel. The metal ions of the salt (such as Na^+ and Mg^{++}) shield the negative charges of the polymer, whereas the chloride ions neutralize the hydrogen ions. The result is thus again to reduce the effective ionization of the polymer. The salt concentration needed to induce the phase transition is much smaller with divalent magnesium ions than it is with monovalent sodium ions. Difference is a factor of about 4,000.

The critical fluctuations in gels are local variations in the density of the polymer network. Because the network is flexible and is in constant thermal motion small aggregations of polymer are constantly forming and disintegrating. Small regions devoid of polymer strands also appear and quickly disappear in the same way. Near the critical point the polymer-density fluctuations, like those of other critical systems, grow larger in amplitude and scale, and their rate of relaxation slows. At the critical point these measures diverge toward infinity.

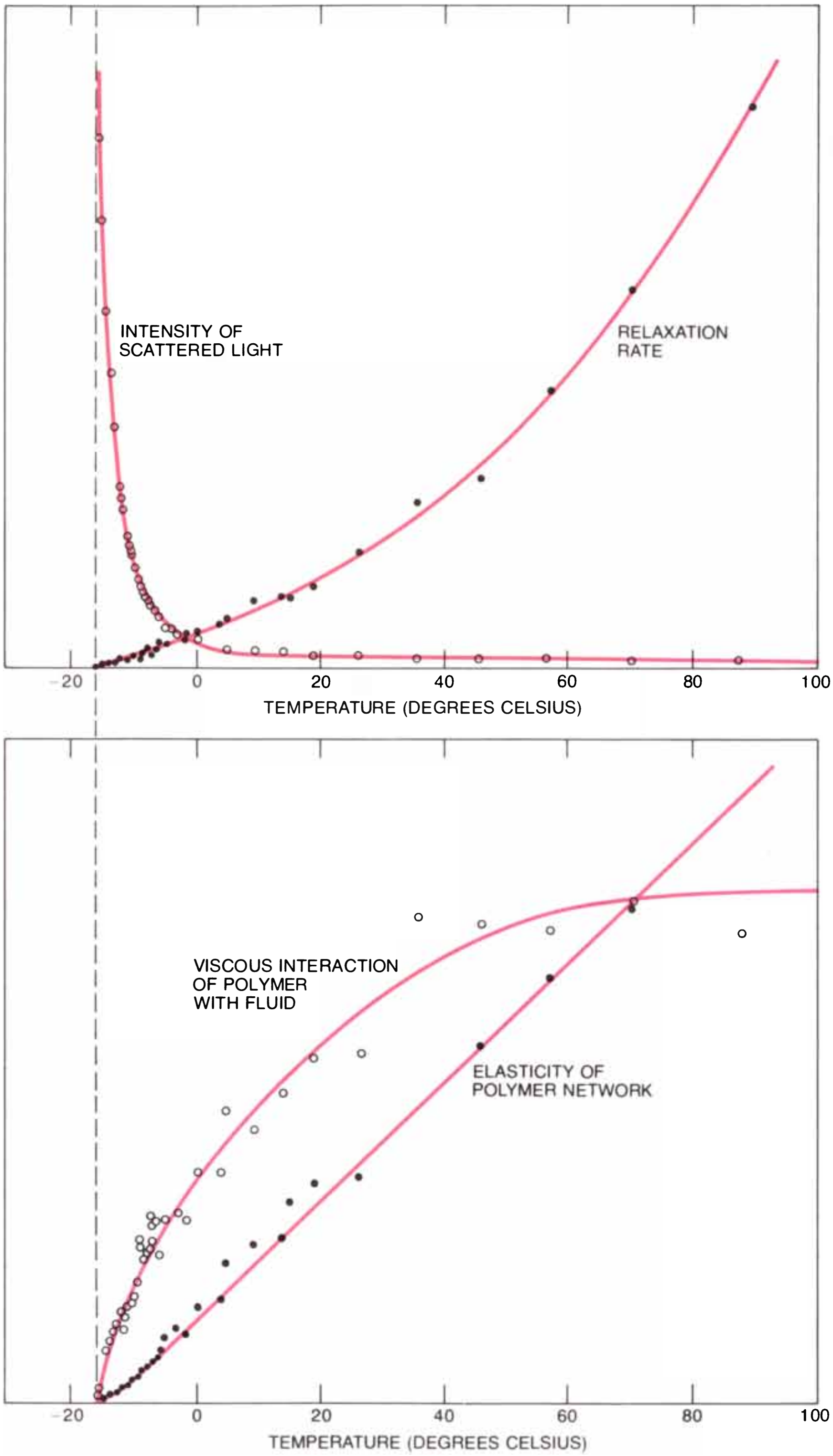
The technique of laser-light scattering offers direct experimental access to the critical fluctuations of gels. As the temperature is lowered toward the critical point the intensity of the scattered light increases. At the critical temperature the gel becomes almost opaque, and so the ratio of scattered light to transmitted light approaches infinity. Random fluctuations in the intensity of the scattered light can also be measured; they indicate the relaxation rate of the corresponding fluctuations in polymer density. The rate becomes indefinitely slow at the critical point.

The development of fluctuations in the polymer matrix is governed by the viscoelastic properties of the gel. Intuition suggests that the amplitude of the fluctuations should be greater at higher temperature; the reason is simply that the ultimate source of the fluctuations is the thermal motion of the polymer. It is also reasonable to expect a connection between the amplitude and the elasticity of the network. Since the elasticity measures the resistance of the network to deformation, the fluctuations should grow larger when the elasticity declines. In fact, it has been demonstrated theoretically that the amplitude of the fluctuations is directly proportional to the absolute temperature and inversely proportional to the elasticity of the polymer network. The intensity of the scattered laser light is thus a direct measure of the ratio of these quantities. The divergence of the scattered-light intensity at a finite critical temperature implies that the elasticity approaches zero. At the critical point the polymer network offers no resistance at all to deformation; it is infinitely compressible.

The relaxation rate can be evaluated in a similar way. The fluctuations will relax quickly if the restoring force, which depends on the elasticity, is large, and if the friction between the network and the interstitial fluid is small. Again, theoretical work has shown that the rate is directly proportional to the elasticity and inversely proportional to the friction force. The friction in turn depends on both the viscosity of the fluid and the average size of the pores in the polymer network. Since the elasticity is known from measurements of the scattering intensity, and since the viscosity can be determined independently, a measurement of the relaxation rate yields information on the pore size in the polymer web. We have found that the pores grow much larger as the critical point is approached.

What is the meaning of an infinite compressibility? How is it possible in a permanently cross-linked network for the pores to grow to many times their normal size?

In general the elasticity of the polymer network as a whole (as distinct from the rubber elasticity of the individual strands) derives from the interaction of the network and the gel fluid. At equilibrium all the components of the osmotic pressure, including any external pressure if there is one, are in balance and the total pressure is zero. If the network is then compressed slightly, a positive pressure is generated. The positive pressure acts as a restoring force that returns the gel to its original volume. At the critical point, however, this mechanism does not operate. There all the forces are not merely in balance; they also remain in balance even after a slight change in volume. If the gel expands slightly, the positive pressure of the rubber elasticity and the hydrogen-ion pressure are reduced but the negative pressure of the polymer-polymer affinity decreases by the same amount in com-



CRITICAL FLUCTUATIONS in the structure of a gel are revealed by measurements of scattered laser light. The upper graph records the intensity of the scattered light and the rate at which the intensity fluctuates as a function of temperature. At high temperature little light is scattered, but as the gel is cooled toward the critical point (here about −17 degrees C.) the scattered intensity rises quickly and diverges toward infinity. At the same time the relaxation rate, which measures the speed with which fluctuations in polymer density form and disappear, slows to zero. The lower graph traces changes in two properties of the gel that can be deduced from the light-scattering measurements. The elasticity of the polymer network indicates the resistance of the network to deformation. At the critical point the elasticity falls to zero, so that the network is infinitely compressible. Viscous interaction between the polymer network and the gel fluid also disappears at the critical point. Interaction is a direct measure of the effective size of the pores in the network, which evidently become very large near the critical point.

pensation. Hence the pressure remains constant and there is no restoring force.

If a change in volume at the critical point causes only a negligible change in pressure, it follows that a small change in pressure can elicit a large change in volume or in the local density of the polymer network. Given this sensitivity even minuscule perturbations in the osmotic pressure can lead to large-amplitude fluctuations in polymer density. These are the critical fluctuations, and they resolve the paradox of large pores in a cross-linked web. Suppose a photograph could be made of the network that would capture its configuration in microscopic detail at some instant. At high temperature the photograph would show a tangled but almost homogeneous net. The average pore size would be determined by the mean distance between cross-links, and there would be little variation around the average. Near the critical point, on the other hand, the photograph would show large fluctuations in polymer density. Some regions would be tightly coagulated, but between them would be open areas with only a sparse distribution of polymer. The average pore size might well be the same as it is at higher temperature, but the average no longer has much significance because there are two sets of pores, one with large pores and one with small. In determining the viscous interaction with the network the effective pore size is that of the dilute, swollen regions, and it is macroscopically large.

If the imaginary photographs of the polymer network could be converted into motion pictures, they would show the relaxation of the density fluctuations. Somewhat above the critical temperature the contracted and the swollen regions would be seen to continually rearrange themselves in response to thermal agitation. Tightly bundled regions would unravel and loose areas would grow dense. As the critical point was approached these rearrangements would slow down, and at the critical point they would become infinitely slow. At a still lower temperature a particular configuration would be permanently frozen in. The coagulated and the swollen regions would then represent the separated phases of the gel.

Temperature and acetone concentration are not the only parameters that can be varied to induce a phase transition in polyacrylamide gels. For example, acetone can be replaced entirely by other solvents, such as alcohol, that also differ from water in their affinity for the polymer. The collapse of the gel can also be brought on at constant temperature and constant solvent composition by changing the *p*H of the solution or by adding a salt to it. Both of these factors alter the effective ionization of the poly-

Learn how to use a computer for $249. And keep the computer.

Introducing the Computer Learning Lab, with the Sinclair ZX80 personal computer.

The Sinclair ZX80 is a powerful, personal computer already in use in tens of thousands of homes and businesses.

Now, for just $249*, the Computer Learning Lab will teach you and your children how to use this complete, expandable computer right in your own home. And when you're done, you get to keep the computer!

The Computer Learning Lab is a self-paced course that teaches the ins and outs of computers. It cuts away jargon and mystique, taking you straight into BASIC—the most common, easy-to-use computer language. And the computer itself does the teaching.

100 LESSONS MAKE LEARNING EASY.

You just take the ZX80 out of the box, connect it to your TV and an ordinary cassette recorder (connectors are provided), and slip in the pre-programmed Learning Lab cassettes. There's nothing extra to purchase.

You'll be working with the computer your very first day!

The cassettes take you through 100 experiments that teach you how to solve problems with the ZX80.

You learn by doing. By actually working with the computer.

And the lessons are designed to be fun and involving. You create your own programs for games, code breaking, interest calculation, and other topics. Then you can apply the principles you've learned to more complex problems.

In fact, you'll be a master of the ZX80. To use in your business, for home budgeting, or just for fun.

We'll also send you a catalog full of ready-to-use programs—from Loan Amortization to Lunar Lander—available for as little as $6.95.

THE ADVANCED DESIGN OF THE ZX80.

The ZX80 is the world's first truly portable computer. It features a touch-sensitive keyboard and a 32-character by 24-line display.

And it performs like a much larger and more expensive computer. Single keystroke entries make typing programs fast and easy. An automatic error detection feature tells you if you make mistakes. And program editing helps you correct them. Yet the complete Computer Learning Lab, including computer, is still several hundred

The Computer Learning Lab is a family learning aid. Children 10 and above will quickly understand the principles of computing—and have fun learning.

dollars less expensive than any comparable computer alone.

The ZX80 is backed by a 30-day money-back guarantee and a 90-day limited warranty with a national service-by-mail facility. Extended service contracts for the ZX80 are available for a minimal charge.

The complete package includes the ZX80, a 128-page guide to computing, a workbook, six program cassettes, and two blank cassettes for storing your own programs.

sinclair

MASTERCHARGE OR VISA ORDERS CALL: (203) 265-9171. We'll refund the cost of your call.
INFORMATION: General and technical (617) 367-1898 or 367-1988. Phones open Mon.—Fri. 8 AM to 8 PM EST.

To: SINCLAIR RESEARCH LTD., 475 MAIN ST., P.O. BOX 3027, WALLINGFORD, CT 06492.

Please send me ________ Computer Learning Lab(s) with the ZX80 personal computer at $249 each (U.S. dollars).

Please send me ________ ZX80 computer(s) only, without the Learning Lab, at $199.95* each.

I enclose check/money order payable to Sinclair Research Ltd. for $______ (Add $5.00 for shipping.)

*For Conn. deliveries, add sales tax.

Name ____________________

Address ____________________

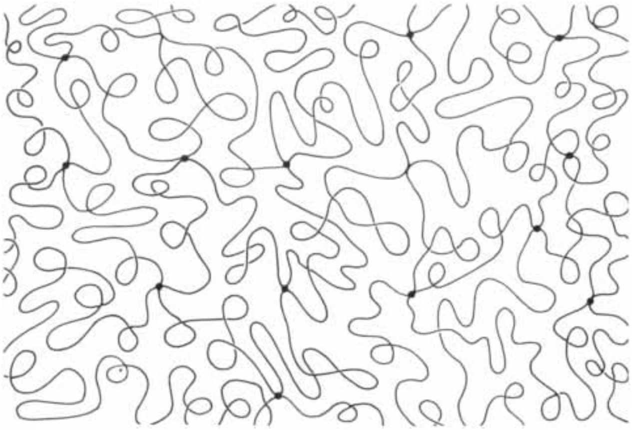
City __________ State ________ Zip ________

Occupation ________________ Age ______

Intended use of ZX80: ____________________

Have you ever used a computer? ☐ Yes ☐ No   Do you own another personal computer? ☐ Yes ☐ No

SA-1-1



CONFIGURATION OF THE POLYMER NETWORK changes drastically as the critical point is approached. At high temperature the distribution of polymer strands is random but on the average uniform. Thermal motions give rise to fluctuations in the density of strands, but the fluctuations are mainly small and short-lived. Near the critical point the fluctuations grow larger in amplitude and scale as the gel separates into domains of high and low density. Below the critical point the fluctuations become frozen in as separated phases.

mer network. Adding acid, which reduces the *p*H or, equivalently, raises the concentration of hydrogen ions, induces some of the ionized carboxyl groups on the polymer to recombine with superabundant hydrogen ions. As a result the hydrogen-ion pressure is reduced and the gel collapses at a higher temperature than it would in a neutral solution.

Adding a salt such as sodium chloride ($NaCl$) or magnesium chloride ($MgCl_2$) achieves the same end by different means. The positively charged metal ions of the salt congregate around the negative charges on the polymer, thereby shielding them. Meanwhile the negative chloride ions neutralize the hydrogen ions in the solution, so that the polymer and the system of dissolved ions are effectively uncoupled. The result is again to reduce the hydrogen-ion component of the osmotic pressure. We have found that the salt concentration needed to induce collapse is 4,000 times less for the divalent, or doubly charged, magnesium ion than it is for the monovalent sodium ion. Surprisingly, the large difference can be explained by the fact that only half as many divalent ions are needed to neutralize the network as monovalent ions. It is interesting to note that divalent ions have an essential role in the contraction of muscle, a process that in some respects seems to resemble the collapse of a gel.

Still another way to induce a phase transition is by applying an electric field across a gel. Typically a gel is placed between two electrodes in a 50 percent mixture of acetone and water. Imposing a small electric field, perhaps half a volt per centimeter, can then lead to the collapse of the gel.

The electric field affects both the negative and the positive charges in the gel. The hydrogen ions flow toward the negative electrode, but they take part in a closed-circuit electrochemical current, so that the number of ions in the gel at any moment is constant. The hydrogen-ion pressure is therefore not affected by the field. The negative charges of the carboxyl groups, on the other hand, are attached to the polymer matrix and cannot flow through the circuit. The polymer is therefore drawn toward the positive electrode. The force is a directed one, but it gives rise to a negative pressure in the gel. A pressure gradient is set up, ranging from strongly negative pressure adjacent to the positive electrode to zero pressure at the other end of the gel.

The pressure set up by the electric field varies smoothly along the length of the gel, but the response of the gel is nonetheless discontinuous. Where the negative pressure is sufficient to bring on a phase transition the gel is substantially shrunken; where the pressure is below the transition threshold the volume of the gel is scarcely changed. Hence the gel is shaped like a wine bottle, with a thick body, a shoulder and a thin neck. Changing the magnitude of the field adjusts the proportions of the body and the neck. When the field is removed entirely, the gel resumes its original form.

A drastic change in volume that can be elicited in response to a considerable variety of small stimuli suggests certain possible applications for ionized gels. For example, a chemical engine might be designed in which the prime mover is a gel. Cycles of expansion and contraction could be powered by changes in temperature, solvent composition, *p*H, ionic strength or electric field. Such a device is particularly appealing for use as an artificial muscle. Because the change in volume is discrete and predictable, a gel might also serve as a memory element or a switch. Arrangements can also be imagined whereby a change in pressure would give rise to a change in color, so that the gels might be employed in the displays of calculators, watches and so on.

An important factor in most such applications is the speed of the phase transition. As I have mentioned, a gel with dimensions of about a centimeter requires several days to reach a new equilibrium volume. The time required, however, is proportional to the square of the linear dimensions of the gel, and so it can be greatly reduced by making the gel small or finely divided. A cylindrical gel one micrometer in diameter, the size of a muscle fiber, would swell or shrink in a few thousandths of a second.

In the eye retinal detachment is caused by a collapse of the vitreous gel; the collapse allows the retina to pull away from the back of the eye. Corneal edema, another common cause of human blindness, is an opacification of the gel of the cornea; it is thought to result from a separation of fluid domains. Further investigation of the physics of phase transitions in gels may improve understanding of these conditions. The ability to determine and control the pore size of a gel may also lead to improvements in the analytic techniques of electrophoresis and chromatography. Apart from these potential applications, a common but curious state of matter has come to be better understood, and the mathematical techniques for describing critical phenomena have been shown to apply to yet another physical system.

THE EXPORT OF EXCELLENCE

A special report on Sweden's technological expertise written by Don Hinrichsen, associate editor of AMBIO, and Victor Kayfetz, a former Stockholm correspondent for Reuters and the London Financial Times.

A century ago Sweden was a tradition-bound, backward country on the northern periphery of Europe where three-quarters of the people lived from agriculture and hundreds of thousands were emigrating to America for lack of opportunities at home.

Today it is a highly industrialized and progressive welfare state with a per capita output exceeding that of the United States. Yet with only 8.3 million people, Sweden is also a leader in a surprising range of technologies.

Natural resources, a deeply-rooted innovative tradition, good labor relations and luck have enabled the Swedes to build their prosperity by exporting nearly half of their manufactured goods. This puts them in a league with the British or West Germans in terms of dependence on trade.

In the 1970's, however, Sweden's rising relative production costs and the spread of technological know-how to newly industrializing countries have led to severe economic strains. While hardly unique to Sweden, these problems came as a shock to a nation accustomed to being held up as a model of orderly growth based on the socially responsible and pragmatic application of increasingly advanced technologies. To retain its competitive edge, Sweden will have to improve the efficiency of its production facilities, marketing, and research and development.

This survey will examine where the country stands today in terms of technological achievements and will focus on areas of current research and development where Sweden is capable of world breakthroughs during the 1980's.

INDUSTRIAL HISTORY

For centuries, Sweden was one of Europe's main sources of copper, silver, iron and forest raw materials, but the Swedes date their industrial revolution from the 1870's.

The first large-scale industrial establishments were the sawmills built along the Gulf of Bothnia on Sweden's northern coast. Production of pulp, paper and wood-based chemicals soon followed.

A modern commercial and special steel industry emerged out of the small, centuries-old and paternalistically run iron foundries of central Sweden. In southern and central Sweden, engineering firms were founded on the basis of Swedish innovations. Some grew into today's multinationals.

The Swedes were making their own versions of the telephone in the 1870's, separators in the following decade, ball bearings and trucks around the turn of the century, cars from the 1920's and military aircraft beginning just before World War II.

After the war, an intact Swedish industrial plant was able to help supply the manufactured goods needed to rebuild Europe. The country strengthened its engineering and shipbuilding industries and diversified further into the pharmaceutical, petrochemical, and electronics fields.

Between 1932 and 1976 the Social Democrats were the dominant political force. They represented the political interests of a powerful labor movement but worked by means of compromise with the non-socialist parties, businessmen, farmers and other interested groups. The Government introduced countercyclical policies and wove a finely-meshed safety net of welfare benefits and services.

The Social Democrats never had a mandate to nationalize industry. In 1976, about 90% remained in private hands and another 5% belonged to cooperatives. They did, however, suceed in establishing full employment as the country's foremost economic policy aim.

Sweden's share of world trade peaked in 1965 at 2.4%. It saw a dramatic drop in the 1975/76 world rècession period, when industrial wages were raised 40% in an effort to catch up with inflation following the oil crisis.

The non-socialist Government in power since 1976, in pursuing a full employment policy, has been forced to support financially, ailing industries (particularly shipbuilding, steel, textiles, electronics and pulp and paper) in an effort to produce an orderly cut back in man power. During 1976-77 it also devalued the krona in an effort to help restore the competitiveness of Swedish exports.

A serious conflict exists over the level of personal taxation where the current marginal rates of 60-80% both discourage work and lead to a 'black economy'. Yet the Social Democratic opposition strongly resists efforts to trim public sector spending which distributes around 65% of the gross national product (GNP), and employs more people than industry. This percentage is the highest of all 24 member countries of the Organization for Economic Cooperation and Development (OECD).

The most successful Swedish companies through these recent economically difficult times have been those which have established production, sales and service units outside their small home country. Indeed a rapid proliferation of such foreign subsidiaries has made Sweden one of the world's most 'multinational' countries in per capita terms.

CURRENT OUTLOOK

When Sweden's Parliament convened in October 1980, Prime Minister Thorbjörn Fälldin announced the first stage of a four-year Government austerity program, aimed at correcting serious imbalances in Sweden's economic situation.

The shortfall in the Government's budget is running at an alarming 11% of Sweden's GNP, far above the OECD average. The austerity program aims at paring it in four stages, to 7% by fiscal 1984/85.

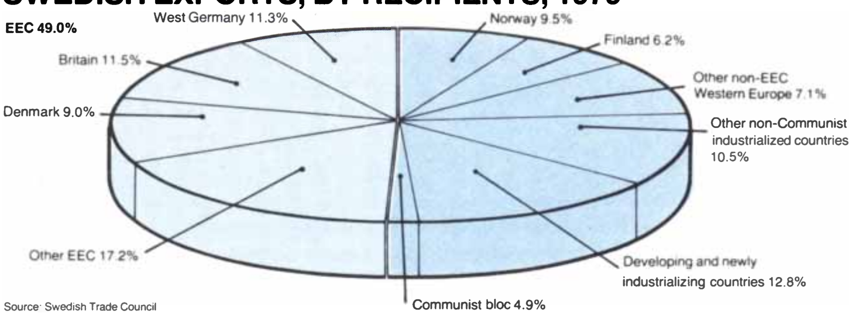
With the public sector pre-empting much of Sweden's borrowing potential both on foreign and domestic credit markets, with interest rates at record levels and the Stockholm Exchange in perpetual doldrums many Swedish companies must try to finance investments from their own resources. Average solvency (asset/debt) ratios of publicly-traded Swedish companies fell from nearly 40% in the late 1960's to around 23% a decade later.

Capital expenditures by industry nonetheless rose in 1980 by some 15% from a historically very low level the preceding year. This was due mainly to the recovery in profits from zero to a barely acceptable level during 1979. Investments seem likely to drop again, since the economic recovery apparently peaked at the end of 1979, after only 18 months.

In the first eight months of 1980, Sweden's exports showed no rise in volume at all compared with the period in 1979.

Late in 1980, official unemployment was at a low 2% and those in relief jobs and retraining programs represented only another 1% of the labor force, but a major rise was expected soon.

SWEDISH EXPORTS, BY RECIPIENTS, 1979



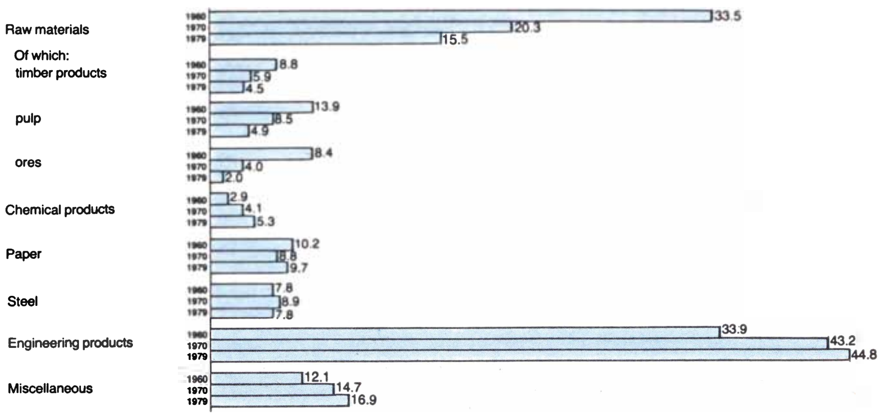
Source: Swedish Trade Council

4.20 Swedish krona equals 1 US dollar

Front cover photo by Angus Forbes

Most industrialists and trade union leaders would concur with the Prime Minister's statement in October that "a solution of Sweden's economic problems presupposes the deliberate expansion of the industrial sector. The transfer of resources from unprofitable to expansive parts of the business sector must be promoted by active industrial, labor market and regional policies."

SWEDISH EXPORTS, BY TYPE OF GOODS %



Source: Swedish Trade Council

R&D POLICY

The Government R&D establishment in Sweden is strictly 'specialized and sectorized'. This means that R&D is not coordinated centrally – there is no one organization with its fingers both on the pursestrings and on the pulse of science and technology. Instead, each Government agency from the Ministries down to the county councils have their own R&D budgets and apportion the pie according to priorities defined by their specific area of activity.

In 1977 a major reorganization took place within the powerful Research Councils, resulting in five councils being pruned down to three: the Council for Research in the Humanities and Social Sciences, the Medical Research Council and the Natural Science Research Council (NFR).

An entirely new organization, the Council for Planning and Coordination of Research (FRN), was created to take a transdisciplinary view, linking together projects in different areas and initiating new research that fails to fit clearly into one sphere of interest. FRN was also given the task of opening up communication channels with the public. Their first substantive endeavor along this line came during the debate prior to last year's nuclear referendum when they provided a series of popular booklets covering the whole spectrum of issues affecting alternative energy and nuclear power development.

There are also well-worn channels of cooperation between different sectors of Swedish industry and research. Underground construction may bring together Atlas Copco, Sandvik, Skanska, Boliden and others. Volvo and ASEA are working closely on the development of essential components for the gas turbine automotive engine. The pharmaceutical companies depend to a great extent on discoveries made in research laboratories at key Swedish universities and institutes.

Five Nobel Prizes are awarded annually in Stockholm. Except for the Economics Prize, created in 1968 by the Bank of Sweden, the awards all stem from a fund set up by the will of Alfred Nobel, the 19th century Swedish entrepreneur whose inventions included dynamite and nitroglycerine.

Since 1901, 363 prizes have been awarded to 459 individuals. By the end of 1980, Americans had captured the most prizes with 138, Sweden had gleaned 23, an impressive number for such a small country.

The spinoff effects of the Nobel Prizes are multiple. As an example, Dr Lars Werkö, executive vice president for R&D at Astra (the pharmaceutical company) points out that important foreign medical researchers readily accept invitations to lecture at Karolinska Medical School in Stockholm because it exposes them to the Nobel Committee or researchers close to it. At the same time the existence of the prizes opens doors for Swedish researchers who want to go abroad to work or to study the research being done by colleagues.

In 1979 total Swedish R&D expenditure amounted to some $1.9 billion. Roughly half of this was financed by industry (including state-owned commercial enterprises and public utilities). R&D expenditures are only 1.9% of GNP, on a par with Japan or France, but slightly below West Germany, Switzerland or the United States. A mere one percent of the world's R&D activities are carried out in Sweden. There is an increasing tendency for Swedish-based multinationals to locate part of the research facilities abroad.

Only 10% of R&D expenditure went to actual research and 90% is poured into the machinery of development.

On the private side, the Wallenberg Foundation awards annual grants of about $5 million. And there are around 700 smaller foundations which supply funds to research and development.

R&D money with Government strings usually goes to develop a specific technology or product. Important sources of public R&D funds include the Industry Fund (under the Ministry for Industry), the Regional Development Funds and the Investment Bank.

When it comes to financing R&D, especially in its infancy, one of the most important organizations is the National Board for Technical Development (STU). In fiscal year 1978/79 STU's total budget was over $95 million, as compared to $15 million in 1968/69 when the organization was first set up. About 10% of the total development efforts in Sweden are financed through STU.

The main areas of STU's involvement include:

- financial support for technical R&D projects and innovations;
- dissemination of technical knowledge, advice and service to industry with special emphasis on small and medium sized enterprises and private inventors;
- long term planning and studies of future technological needs.

STU and the Royal Swedish Academy of Engineering Sciences (IVA) cooperate with the Foreign Ministry in maintaining nine technical attachés in eight countries (including two in the USA). Just as the Trade Council, with its worldwide network of 32 Trade Commissioners, helps Swedish industries market their products abroad, STU keeps its eyes on selected technological developments, making assessments of important factors which may influence the technical/industrial scene in Sweden. It also keeps a 'track record' of Swedish technologies, carefully monitoring the country's technical competence as well as the competitive capability of various key branches of Sweden's export industry.

The impressive growth of STU's budget over the past decade and R&D funds in general, reflects growing governmental and private concern about Sweden's technological future. Nevertheless, there remains an acute lack of risk capital. Much criticism has also been aimed at the way official R&D is carried out. There are simply more projects worthy of support than there are funds to go around. A

number of researchers in the energy field, for example, feel that less money should be spent on the wind energy program and more pumped into the work on alternative fuels like methanol.

Furthermore, the sheer drain on management time due to all the legislation passed each year, and increasingly complex bureaucracy are often cited as reasons why so many smaller businesses in Sweden go bankrupt. Industrialists argue that the numerous economic disincentives to creation of new industrial jobs in Sweden pose a grave threat to hopes that a shift to higher technology can help the country export its way out of chronic payments deficits in the coming decade.

THE MOVE AWAY FROM OIL

Sweden imports about three-quarters of its total energy needs, most of it in the form of increasingly expensive Middle East crude oil. As a result, Swedes consume more imported oil per capita than any other nation.

Sweden's 1978 energy mix consisted of: oil 72%, hydroelectric power 12%, lyes and wood waste 8%, coal and coke 4%,and nuclear power 4%. Nearly 70% of the country's electricity is generated by hydroelectric power plants, while 25% comes from nuclear reactors.

In a referendum on March 23, 1980, Swedes decided to put a 'parenthesis' around the future development of nuclear power. As a result of the vote, Sweden will have six more reactors operating soon – in addition to six already in use – with a combined capacity of 9,440 MW (electric). But all of them are slated to be phased out in 20 to 30 years.

With the future of nuclear power limited and with environmental constraints on further hydroelectric power expansion, Sweden has launched a major effort to develop domestic sources of energy in a planned, nationwide program to decrease dependency on imported oil. The energy R&D budget for the current three-year period (1978-81) has doubled to $185 million.

Some 25% of the total energy R&D program is 'solar' (i.e. renewable). Solar R&D is fairly comprehensive, but strongly focused on biomass, wind energy and space heating.

Some energy planners, especially in industry, believe Sweden must greatly increase its coal imports at least until a large-scale solar-run economy can be implemented. But first the pollution problems related to coal must be solved.

One of the main snags in using solar collectors is how to store the sun's energy from summer to winter. Quite by chance, cosmic ray physicist Ernst-Ake Brunberg and his associates at the Royal Institute of Technology in Stockholm stumbled on a solution to the problem of seasonal storage – a chemical heat pump that is practical to use and relatively cheap to operate.

The system, newly patented under the name Tepidus, utilizes a new kind of salt called Tepidit.

According to Dr Brunberg, Tepidit can be made from simple raw materials; its energy density is high; it takes up, stores and gives back one kWh (kilowatt hour) of heat per kilo of salt, staying solid all the time; and it gives out its heat at a constant temperature.

Kjell Bakken, the man in charge of marketing Tepidit, explains, "The advantage of the method is not only that we can store energy for an indefinite time without loss, but also that the storage capacity per kilo of salt is so high that it will be economical to use heat at 70-80°C wasted from industry, 'charge' the energy containers there and transport them to houses for use."

Physicists Thomas B Johansson and Peter Steen have shown theoretically how Sweden could have an energy system based on domestic renewable sources by the year 2015. In their book *Solar Sweden* , half the country's total energy requirements could be met with fast growing, short-rotation energy forests.

Lars Rey, executive director of the National Board for Energy Source Development (NE) says, "Forestry energy (including logging wastes) tops the list of our energy options. By 1990, 20% of our energy could come from forests."

NE calculates that 3 million hectares, or nearly seven percent of Sweden's total land area, could be used for energy plantations.

Professor Gustaf Sirén of the Agricultural University in Uppsala and his researchers have been cultivating 11 different fast-growing clones of poplars, willows and alders on four experimental stations scattered throughout the country. According to Professor Sirén, "some clones grow four meters a year. Rotation periods of one to three years appear quite feasible."

NE anticipates having 100,000 hectares sprouting high-yielding energy forests by 1990. The harvested trees will be taken to regional processing centers where they will be turned into powder or chips for use in power plants, industrial boilers and eventually homes.

Sweden is located in the belt of westerly winds, and therefore has large areas, particularly along the Baltic coast, where winds blow eight to ten meters per second at heights of 100 meters.

1983-84 will be the evaluation period for the two prototype wind turbines now under construction.

The wind turbine to be erected on the island of Gotland will be a two-bladed horizontal-axis machine with a generating capacity of two megawatts. The other, a three megawatt unit, will be built near the southern tip of Sweden.

NE anticipates having some 100 large-scale wind turbines operating by 1995 if the Government gives them a green light in 1985.

Energy conservation is a constantly recurring theme in much of Swedish R&D. With rising costs on raw materials and labor, one way companies can increase their paper-thin profit margins is to introduce energy-saving technologies or institute more conservation minded practices on the shop floor.

About 40% of Sweden's total energy budget is used by industry. So it is not surprising that the Government is providing loans for all industrial energy-conservation measures which have a pay-off time of over three years.

One of the most spectacular projects comes from the Boliden mining, smelting and chemical company, whose own oil imports account for 1-2% of Swedish energy consumption. Boliden is launching Carbogel, a purified liquid coal process that could slash the country's oil imports by one-third or more in 1990.

Boliden will exploit Carbogel in partnership with Scaniainventor, a Helsingborg-based company which devised the process.

The Carbogel process dovetails nicely with Boliden's existing know-how in separating non-ferrous metals and sulfur from low grade ores.

To make Carbogel, Boliden grinds up coal of any quality and separates off the ash, heavy metals and inorganic sulfur. Then, the company puts in water plus a secret, soap-like additive that disperses the purified coal dust evenly in a permanent suspension resembling heavy fuel oil. Boliden plans to use the sulfur and some of the other residues to help supply its other production activities.

Carbogel can be pumped in pipelines and stored indefinitely, and it cannot explode. Its energy value is 60% that of the same volume of oil, and it burns more cleanly than oil.

The dispersant used in the fuel is a well-known, commercially available chemical, says Boliden's technical director Gunnar Olgård. "The market is so tremendously big that we can't have it for ourselves. We'll be happy if we have just a few percent," he says.

If the plans go through, Boliden will begin within a few years to import millions of tons of American coal annually to a big harbor in Sweden. The company is already building grinding and flotation plants to make Carbogel and will establish a distribution network to supply the fuel to its own factories and to other customers in Scandinavia.

Another method that will allow high-sulfur coals to be burned efficiently with very little atmospheric pollution (especially sulfur dioxide) is pressurized fluidized bed combustion (PFBC). Stal-Laval Turbin AB of Sweden is collaborating with the American Electric Power Company and with West Germany's Deutsche Babcock AG to build a 170 MW combined gas turbine/steam turbine PFBC pilot plant near Brilliant, Ohio by 1986. It will test a new system it has developed in the past few years.

Stal-Laval is also building a much smaller pilot plant in southern Sweden together with the Sydkraft power company. Trial runs will begin early in 1982.

In the new system, limestone or dolomite is placed in the coal-burning vessel to remove 90% of the sulfur.

According to Bengt Wickström, Stal-Laval's coordinator for the project, the low combustion temperatures (800-900°C) and the use of 16-bar pressurization in the vessel reduce the emissions of pollutant gases. Boiler tubes submerged in the fluidized bed provide heat transmission coefficients at least four times those of conventional boilers. The system will use the company's GT 120 gas turbines.

GOING UNDERGROUND

Sweden has been putting things in the ground on a large scale since World War II, when several enormous caverns in Stockholm were blasted out of solid rock, intended as bomb shelters and storage areas for emergency war supplies.

Today, the country uses vast underground areas for hazardous liquids like crude oil and heating fuel, water reservoirs, grain silos, hot water basins fed by industrial waste heat for district heating purposes, telephone exchanges, and even computer centers.

"We have already begun to speak of subterranean urbanization," says Birger Jansson, chief consultant at VBB, a Stockholm-based company which has planned a number of underground projects in Sweden and abroad. "But the human perspective is important. We're not putting humans underground, but rather functions that are not dependent on daylight."

With the onslaught of the energy crisis and the grim possibility of interrupted oil supplies from the turbulent Middle East, the Swedish Government decided that a 90 day supply of oil, corresponding to 10 million cubic meters, must be kept in reserve.

Part of this program is the massive Brofjorden storage system for crude oil, whose second stage will be finished this year. When the third and last stage is built, this Swedish west coast facility alone will hold 40% of the country's strategic oil reserves, or 4 million cubic meters – big enough to swallow up the homes of 30-40,000 people.

One Swedish firm with nearly 60 subsurface storage centers to its credit is the Skanska construction group. They began constructing underground caverns for oil storage in the late 1950's, and have accumulated a wealth of know-how. The company can handle a job from start to finish and has a reputation for its economical and safe designs. They also provide all the environmental control technology necessary to keep the plant operating with a minimum of problems. Skanska says that 85% of their projects are carried out on a turnkey basis.

Atlas Copco, the 'compressed air people' and Sandvik, known for their high quality steel drill bits, have been working together since the 1950's providing know-how and equipment for mining and underground construction.

Says Atlas Copco's managing director Tom Wachtmeister, "We entered hydraulic technology in the early 1970's as a supplement to our compressed-air-powered units and now we are a world leader in this field."

Hydraulics caught on quickly, especially for tunnel drilling since it is faster and requires less energy than compressed air equipment and provides a better working environment.

Atlas Copco's R&D is frequently 10-15 years ahead of the market. For example, they recently developed a new hydraulic water cannon, called the CRAC 200, that splits boulders safely and quickly without explosives.

Magnus Bergman, secretary general of Rockstore 80 – the international symposium on subsurface space held in Stockholm in June 1980 – points out that going underground offers four big advantages: "low construction costs, low operating costs, improved interior climate stability and lower energy consumption."

Boliden, Scandinavia's largest miner and smelter of sulfide ores containing non-ferrous metals, uses ores with extremely low metal content to produce copper, lead, zinc, pyrites, silver and gold, and has thus developed systems for dealing with difficult ores.

Boliden collaborated with Sweden's ASEA in designing a computerized on-stream analysis system for process control during flotation of ores. It utilizes X-rays of slurry samples which are pumped continuously from relevant check points in the ore dressing process. The purpose of the system, which Boliden has tested since 1975 and has now begun to market, is to make it possible to extract larger quantities of valuable metals from the ore while minimizing the consumption of the chemicals (flotation reagents) used to process the slurry.

The on-stream analyzer contains a single lithium drifted silicon energy-dispersive detector, which can be targeted toward each of 16 slurry cells according to a pre-set program. The detector converts the radiation from each test into electrical impulses, which are analyzed in a computer. Information on the process then appears in the form of a print-out or on a visual display unit.

Boliden has received a grant from STU for field tests of a new method of airborne electromagnetic prospecting that will make it possible to spot ore bodies that lie more than 50 meters below the surface or that have poor electrical conductivity. The transmitter in this so-called MEM system is a magnetic di-pole mounted horizontally to the direction of flight. The receiver is an antenna aimed in the direction of flight and either towed behind the plane or located on a second aircraft.

The same principle can also be used in ground-based measurements and in prospecting for water in desert areas, explains Boliden's Kjell Westerberg, one of the project leaders.

SWEDISH STEEL: SMALLER IS BEAUTIFUL

In 1978 Sweden's three largest commercial steel mills – two private and one Government-owned – were combined into the half state-owned steel company, Swedish Steel AB (SSAB). Productivity remains a problem. In 1979 the output per employee was only 120 metric tons (tons) as compared to 400 tons in Japanese mills and 250 tons for the best West German steel plants.

Nevertheless, 1979 was an important year for Sweden's steel makers. Three new processes for producing crude iron were announced: ASEA/Stora Kopparbergs' ELRED; Boliden's INRED; and SKF Steel's Plasmasmelt. All three methods are based on the direct reduction of the ore to crude iron, which makes traditional coking and sintering plants obsolete, thereby increasing production, decreasing costs and saving energy.

ELRED

The interesting twist in this two-stage reduction process is the addition of a new cogeneration power plant. In the first stage, the dressed ore is prereduced in a circulating fluidized bed. The prereduced material is then transported to a direct current electric arc furnace, where the slag is separated.

The gases produced during the extraction of the crude iron are recycled through the power plant that provides electricity to the electric arc furnace. There is even enough power left over to supply other parts of the mill.

INRED

The unique point about this two-stage reduction process is that it takes place in one reactor. First, dressed ore, coal, and oxygen are blown into a flame melting chamber. Exhaust sulfur gases are transformed to steam that drive a turbine which generates electricity used in the smelting and final reduction stages.

Inred is said to be especially suitable for ores which contain several different metals. For example, it permits the extraction of sulfur, iron, and nickel from ferrous nickel sulphides and iron and aluminum from low grade bauxites.

Plasmasmelt

This two stage process uses electric plasma technology. The prereduction takes place in two fluidized beds. The dressed ore is then transported to a plasma burner where it is 'ionized' and quickly converted to crude iron by gases heated to 5,000-7,000° C. Powdered coal or heavy oil is used as a reducing medium. The resulting gas is partly reused for prereduction and partly to run the plasma burner. Only a very small amount of the gas is wasted.

SKF Steel stresses that this process can be added to existing operations without heavy restructuring costs. The company maintains that Plasmasmelt is profitable even if the annual steel production is as low as 250,000 tons.

Industrial sources have said that at least two of these three processes are likely to receive some Government support enabling full-scale pilot plants to be built.

In another steel-related innovation, SKF, the world's largest maker of bearings (SKF Steel is a subsidiary) has developed a new line of advanced wear-resistant roller bearings called 'CC'. They were computer-designed with a slightly different shape to improve roller equilibrium, thereby reducing friction and prolonging equipment life, while also reducing maintenance costs.

FOREST INDUSTRY AND AGRICULTURE

The Swedish forest industry harvests about 70 million cubic meters of timber each year, about equal to the regrowth capacity of the country's commercial woodlands. An estimated 50 million cubic meters of logging wastes are left in the forests, including stumps, limbs, branches, tops, needles and leaves.

NE claims that nearly 28 million cubic meters of logging waste a year could be used to produce energy, while the forest industry reduces that figure to between seven and eleven million cubic meters.

In any event, the Swedish forest industry has been shifting rapidly from the so-called short-wood logging method (where only the main trunk is used) to what is termed whole tree utilization – from stumps to tops.

Stora Kopparberg, an innovator in this field, is one of the oldest companies in the world, dating back to 1347. It collaborated with manufacturers of heavy equipment to design a suitable harvester capable of pulling out stumps as well as shaking out the soil and rocks that come with the roots.

"We were the first to use stumps. We use them as fiber in the pulp and paper process, but of course they can also be used for fuel," says Erik Sundblad, managing director of Stora Kopparberg. At Stora's central laboratory in Falun, researchers are busy developing a new kind of paper with three layers. The middle layer can be made from cheaper mechanical pulp to get a stiffer paper. This also opens up prospects for new types of analytical filter papers and they are even using KemaNobel's 'microspheres' (see section on Chemicals) to see if a new method can be found to make better, thicker packing paper, among other applications.

Agripaper is the brainchild of inventor Gottfried Schmidt, who has developed a method for transforming peat into a new kind of paper on which crops can be germinated in arid or semi-arid climates. His method allows seeds to sprout with 50% less water than normally required. The state owned Örebro paper mill is producing the product and anticipates large sales to Middle East countries.

Alfa Laval, the company founded by Gustav de Laval, the inventor of the centrifugal separator for milk, has introduced a computerized system, called Alfa Feed, which automatically gives milk cows just the right amount of feed. It's all done by a microcomputer which regulates each cow's daily intake, to maximize production and reduce fodder wastage.

ENVIRONMENTAL TECHNOLOGY

Sweden's tough environmental regulations place strict controls on industrial polluters.

Along with Norway, Sweden was instrumental in getting a number of countries to adopt the Convention on Long-Range Transboundary Air Pollution, signed in Geneva, in November 1979. Sweden is also the first country in the world to place a comprehensive ban on cadmium, effective in mid-1982.

Swedes discard about 6.5 million tons of refuse a year (2.5 million tons from households). Of this, 30% is burned in incinerators, while the rest is dumped in land fills. This represents an incredible waste. Now, the resource value of garbage is attracting widespread attention and so too is the technology used to recover it.

One of the world leaders in environmental technology, Svenska Fläktfabriken Flakt) has recently developed a new process for refuse recycling, called RRR (Resources Recovery from Refuse). The first commercial RRR plant, designed to handle 125,000 tons of household refuse a year, is already operating in Wijster, Holland and a second one has just begun shakedown operations at Lövsta, outside Stockholm.

This process uses Flakt's knowledge of air technology to centrally sort, dry and recover household garbage and refuse.

The refuse is shredded and fed into a classifier. Here warm air, shot at high velocity through a slot in the bottom, blows the light fractions (papers, plastics, organic residues) up into a cyclone, while the metals fall to the bottom where a magnet separates the ferrous metals from non-ferrous.

The cyclone collects the light fractions and transports them to another shredder and from there to a second screen, which separates out the organic waste. In the final stage, the remaining paper and plastic is dried and sterilized by a flash drier before final separation in an air classifier.

"We use air to do the work for us. The end result of this system is that oil is saved since much less is incinerated and a lot of material is recovered for recycling," says Per Olof Alfredsson, director of R&D at Flakt.

Another system, called Brini, has been developed by the Malmö-based firm of PLM Environomental Systems. In this process, the paper and plastic (light fractions) are separated out and converted into combustible pellets that have two-thirds the energy value of coal. Furthermore, these pellets have a low sulfur content and can be fired in existing district heating plants and in industrial furnaces that burn solid fuels.

Unlike the Flakt system, PLM's process uses a vibrating bed to separate materials according to their specific weights.

"As a result of our work with Brini, we are now pelletizing straw, wood chips, peat and bark as new fuels for industrial and district heating purposes," says Stig Edner, project manager for PLM Environmental Systems.

A research team headed by Professor Claes Allander at the Royal Institute of Technology in Stockholm is the first in the world to develop a workable method for visualizing the dispersion of invisible gases in enclosed areas.

Most dangerous gases cannot be seen, but they can be photographed with infrared techniques since the gases absorb radiation in the infrared range on their characteristic wavebands.

A special camera designed by AGA Infrared Systems in Stockholm, was used to make a black and white filmstrip of the distribution of laughing gas. When the room was empty the gas simply followed the airstream, but as soon as a person entered, it spread out and headed straight for the victim's nose and mouth.

This system should have immediate practical applications in helping occupational health specialists and testing officers trace dangerous gases in the working environment.

CHEMICALS

With the cadmium ban only 18 months away, KemaNobel, the largest chemical corporation in Sweden, is experimenting with tin to replace cadmium as a stabilizer in plastics. So far the results look promising but inconclusive.

Meanwhile KemaNobel has come up with a new product called 'microspheres' – tiny plastic balls made from polyvinyliden acrylonitrile copolymers and filled with a pressurized gas. Microspheres will have a wide range of applications. When blended with printing inks they produce three-dimensional print. A mix with only 1% microspheres can reduce the fiber content of paper and cardboard by 12%, and when added to cement, the tiny spheres significantly increase its durability in cold climates.

In a bold step, Bofors Nobel Kemi in Karlskoga has built a pilot plant for making a valuable compound called trinitrobenzene (TNB). TNB has a large potential market and a variety of uses including the production of safer explosives that detonate only at high temperatures. Other applications are in deep oil drilling and as a raw material in dyes and drugs.

The catch is that production will take place in one super-quick reaction directly from trinitrotoluene, better know as TNT.

Normally this route is so dangerous that no chemical company has attempted it. But with a long tradition of handling dangerous chemical processes, Nobel Kemi has the necessary expertise.

The world's first industrial process for the manufacture of the amino acid tryptophane (with the aid of microorganisms) has been devised by Svenska Sockerfabriken in cooperation with Bofors Nobel Kemi.

PHARMACEUTICALS

Sweden's most research-intensive industry is pharmaceuticals, where companies are spending about 15% of their turnover on R&D. This is a much higher percentage than in the United States, for example. "We have far too many research projects, an innovative climate, and a rather harmonious collaboration with the academic world," says Erik Danielsson, head of Pharmacia AB, the pharmaceutical division of AB Fortia.

Since the 1940's collaboration with universities and research institutes has helped the Pharmacia companies to create a number of widely-used products based on dextran, a sugar polymer.

These products include a blood plasma substitute, a cell culturing medium used in making vaccine and a gel filtration system used in laboratories around the world to separate molecules by their size. A recent dextran product is Debrisan, sprinkled in the form of small beads into wounds to speed healing by absorbing bacteria-laden fluids.

Mr Danielsson believes that at the moment, the most promising area of R&D for his division is in ophthalmology. Pharmacia attracted great interest last March at a meeting of eye doctors in Los Angeles when it introduced Healon, a purified fraction of sodium hyaluronate made from roosters' combs.

Healon, a jelly-like material, is injected into the eye during surgery. It forms a film over vital parts of the eye, protecting them from damage during the operation. In cataract surgery, this makes it much easier and safer than before to replace the lens of the eye with an artificial substitute.

Pharmacia is collaborating with eye specialists in developing further uses for Healon, such as in glaucoma surgery.

The company is working on new ways of preventing postoperative eye infections, and is also testing a substance that may help eye doctors map out damaged blood vessels in the retina – a frequent cause of blindness – so that they can repair the damage using lasers.

Another field where Fortia, particularly its Pharmacia diagnostics and fine chemicals divisions, has set the pace is in the production of standardized and virtually pure allergen extracts. Allergy patients receive injections of the extract in gradually increasing doses and build up an immunity against the substance that triggers their allergy.

Perhaps the most interesting of these allergen extracts are the Pharmalgen series, used to treat patients who run the risk of violent allergic reactions to the stings of bees, wasps and hornets.

The freeze-dried vaccines, which keep for years, are made from pure insect venom and have shown 95-98% effectiveness in controlled clinical trials where conventional vaccines, which are made from crushed whole insects, proved no more effective than placebos.

Pharmacia also makes other allergen extracts from tree and grass pollens, mites and house dust. It has developed a series of easily-administered blood tests for diagnosing allergies.

In two other areas, the Pharmcia pharmaceutical division is doing promising research which is still in a very early stage. The company is examining the use of so-called SOD enzymes to prevent the body's own defense mechanisms against diseases from releasing a substance that appears to aggravate rheumatism. And it is assessing the potential for treating autoimmune diseases by means of filtering specific chemicals out of the patient's blood.

Sweden's largest pharmaceutical group, Astra, is well known around the world for Xylocaine, the local anesthetic it developed some 30 years ago. Other major Astra products include anti-asthmatic agents, synthetic penicillin and cardiovascular preparations.

Astra appears to have made a world breakthrough in treating a minor but psychologically distressing disease, herpes. A fairly common virus, herpes simplex causes unsightly cold sores around the mouth or inflammations around the genital region.

Astra has been testing a compound called foscarnet sodium (trisodium phosphonoformate), which attacks enzymes in the herpes simplex virus but leaves human cell enzymes unaffected.

Dr Bengt Nilsson, medical director at the Astra subsidiary carrying out this work, says that "We have been able to show that the drug is effective. What remains is to broaden our patient material in a way that will satisfy the documentation requirements of regulatory authorities.

Astra hopes that the progress already made on herpes at its new virus laboratory will lead to the development of drugs to treat influenza.

Meanwhile Astra has broken ground in another field, central nervous system drugs. The aim is to find substances that effectively treat chronic depression but avoid the serious side effects – such as impaired heart action, blurred vision and slowed mental functions– caused by widely-used antidepressants developed in the 1950's.

In 1974 Astra found a compound, zimelidine, which appeared to satisfy its requirements. Unlike older antidepressants which work by heightening the effects of noradrenaline (one of the signal substances that transmit messages between the brain cells), zimelidine acts upon another signal substance called serotonine.

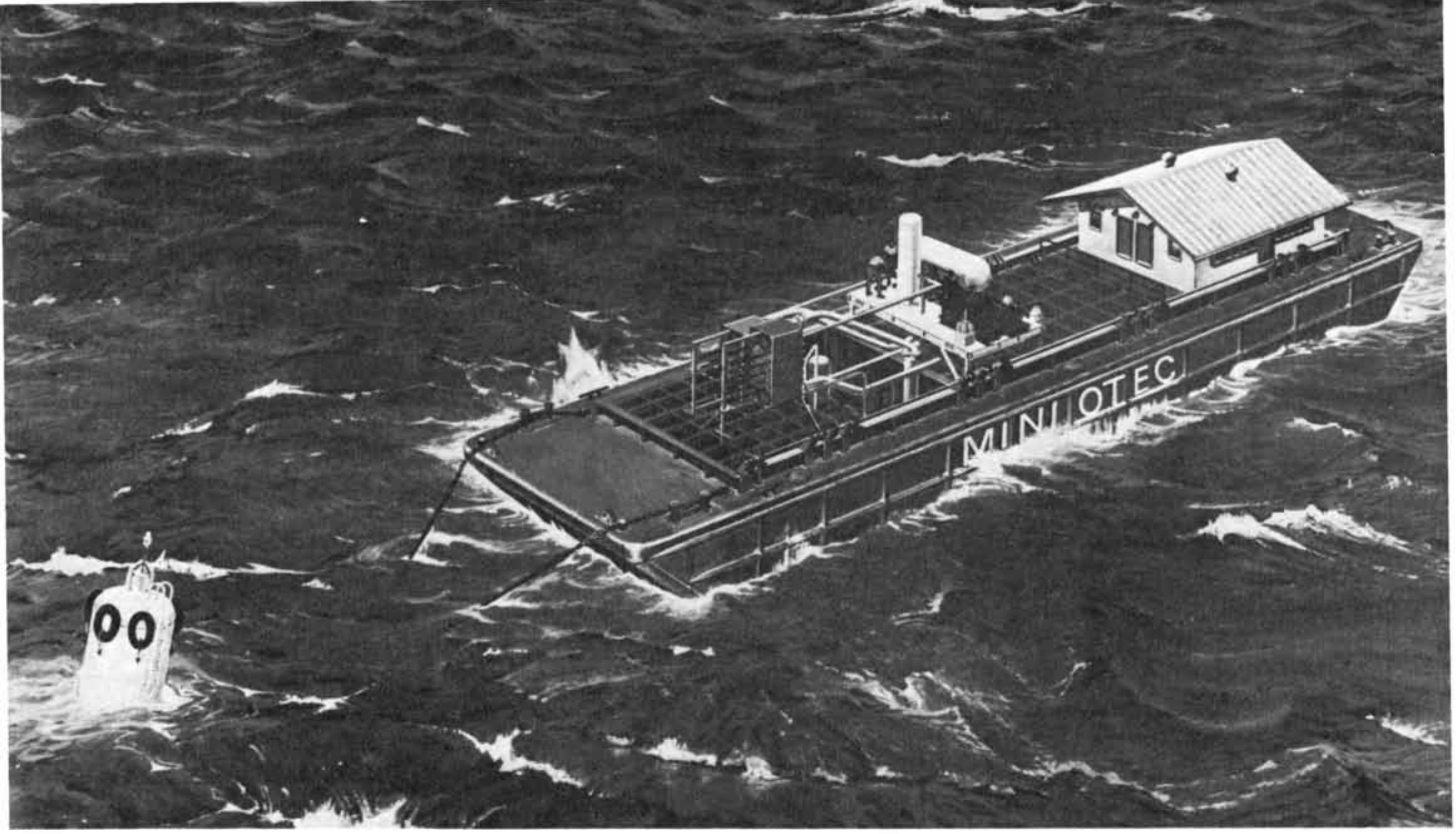
So far clinical trials on over 1,000 patients in Sweden, Britain, the United States Australia and Japan have shown good results, says Dr. Nilsson.

MEDICAL TECHNOLOGY

The export of advanced medical equipment is another Swedish speciality. Foremost in this field is LKB-Produkter in Stockholm, which trimmed back its product range in the 1970's and has emerged as a world leader in selected areas. LKB exports 96% of its production.

The company recently developed a special microtome that can cut large biological samples, including whole organs, into such thin slices that they can be placed directly under a microscope. Cross-sections of whole organs make it possible for cancer researchers to study the spread of tumors.

LKB has recently gone into the field of



Two giant Alfa-Laval titanium plate heat exchangers were the "heart" of the Mini OTEC power plant off the coast of Hawaii.

When Mini OTEC was launched . . .

. . . Alfa-Laval was on board

August 2, 1979. Mini OTEC, launched off Ke-ahole Point from the Island of Hawaii, officially became the world's first successful, self-sustained Ocean Thermal Energy Conversion system at sea. As predicted, it easily produced 50 KW of useable electricity.

Alfa-Laval advance design plate heat exchangers were at the heart of this remarkable pilot project.

By efficiently converting the temperature differences of the solar-heated surface waters and the deep ocean cold waters into turbine-driving energy, Alfa-Laval plate heat exchangers proved the merit of their selection.

Close temperature approach, space saving, low fouling and relative material economy. These features were vital to the success of the entire system and its future as an alternative to other power sources.

Alfa-Laval has helped to pioneer many such important break-throughs in scientific exploration and industrial technology. As the world's largest manufacturer of high speed centrifuges and plate heat exchangers, we are uniquely equipped to provide specialized expertise in liquid processing.

Whether it's separating, heating or cooling liquids in milk, wine, beer or other food plants; purifying oils in factories, power plants, ships or offshore drilling platforms; or clarifying chemicals, pharmaceutical or waste streams, call on us. We'll pioneer with you, too.

For more information, write Alfa-Laval, Inc., Ft. Lee, New Jersey 07024. Outside the United States, write Alfa-Laval AB, Box 500, S-147 00, Tumba, Sweden.

ALFA-LAVAL

analytical luminescence based on substances obtained from firefly lanterns. This method can be used for sensitive measurements of ATP (adenosine triphosphate), which acts as the major energy transfering system in all living cells.

A serious obstacle in the routine use of the firefly reaction for analytical purposes has been that the light was emitted as a brief flash. A new firefly reagent with a stable light emission has now been developed at LKB.

There are many important applications in biochemistry, microbiology and medicine. One example is the measurement of the creatine kinase MB isoenzyme leaking from the heart to the blood after a heart attack.

An extremely compact artificial hand for small children has been developed by the department of neurophysiology at Örebro Regional Hospital, working with engineers at Systemteknik AB near Stockholm. The 210-gram prosthesis is capable of wrist rotation, flexation and other movements, which are triggered by electric signals from the brain that reach the muscles in the stump of the amputated forearm.

The hand, containing an electric motor powered by a six-volt battery, is covered with molded rubber that makes it look very realistic. Provided the stump is suitable, the prosthesis can be fitted to children less than two years old. Children wearing such a myoelectric hand quickly become skilled in using it like a normal one. They can draw, swing on ropes and climb trees. Those collaborating in the project include medical researchers in Gothenburg who developed a similar adult hand in the mid-1970's.

Meanwhile, scientists at the University of Stockholm physics department and at the neuroradiology and neurophysiology departments of Karolinska Hospital have built a positron camera that provides images of regional energy metabolism inside the brain.

Glucose, or grape sugar, is the main energy source of human brain cells. Epileptic foci and tumors have a different glucose metabolism than the rest of the brain, and it is believed that similar changes occur in certain mental diseases and post-traumatic states.

Glucose tagged with radioactive isotopes is injected into the arm. A 95-crystal detector array, located in a ring around the patient's head, then registers radiation emitted from the brain when positrons from the radioactive glucose collide with electrons to form gamma rays. A computer system processes the data and presents it on a console.

The camera is already being used to diagnose malignant tumors, infarctions and hemorrhages, but its applications in metabolic studies are still at an early stage.

ELECTRONICS AND TELECOMMUNICATIONS

For a country that freely admits its shortcomings in keeping up with world computer technology, Sweden has played a major role in developing one specialized area of electronics – the industrial robot.

Today close to 1000 robots are at work in Swedish factories, performing hazardous or monotonous tasks with the blessings of the labor unions. West Germany has 900 such machines at work, North America about 3,000, Japan perhaps 10,000. Sweden is first in the world in terms of robots per head of population.

The best known Swedish robot makers are ASEA and Electrolux. ASEA now manufactures up-market, electronically driven robots capable of following intricate microcomputer programs. It sells 80% of them on foreign markets all over the globe. Electrolux makes cheaper pneumatic robots and exports half of them, mainly to neighboring countries.

Working together with ESAB, the Swedish welding equipment company, ASEA has devised a robot that can perform arc welding. ASEA is now developing a gripper that will allow robots to handle a wider variety of objects.

The company is even offering a robot with 'a sense of feel'. It is programmed to be able to search in two or three dimensions for a corner, or to vary feed rates appropriately during a grinding operation.

By late 1980, some two dozen countries had ordered the L.M. Ericsson AXE digital switching system as part of their national telephone networks.

While continuing to develop accessories for the AXE exchanges, Ericsson is moving increasingly into other areas of telecommunications such as office automation and private communications systems.

By late this year, for example, Ericsson hopes to introduce the Eritex 1000, an advanced terminal designed to fit the public Teletex network that will go into service at that time in Sweden.

Teletex sends a page of text from one subscriber to another in 10 seconds, much faster than telex or facsimile.

Meanwhile Ellemtel, a development company jointly owned by Ericsson and the Swedish Telecommunications Administration, has designed Sweden's first new standard telephone set since the mid-1960's. Ericsson is marketing this all-electronic system internationally and is installing half a million sets in Saudi Arabia alone.

Called the Diavox, it is a very compact pushbutton instrument with good ergonomic features, a new type of electret microphone, and a modular electronic system with tiny circuit boards that can be interchanged if a different set of functions is desired.

Selcom, a high-technology electronics firm near Gothenburg, has developed a very sophisticated device for non-contact measurements of dimensions. Called the Optocator, it uses solid state optoelectronic techniques to provide precision measurement data concerning position, dimensions, contours, vibrations, thickness and so on. Unlike other non-contact measuring systems, the Optocator does not have to be close to the measuring surface. For example, it can measure newly-formed steel billets before they cool off.

A microprocessor continuously monitors measured variables and gives a warning signal should the measurements exceed desired upper or lower tolerance limits.

Facit Data Products, part of the Electrolux Group, is one of the world's largest producers of peripheral data products, holding a 50% U.S. market share in certain mini-computer peripherals

One of their most recent innovations is the 4540 serial matrix printer. The head uses print hammers which are mounted on flexible arms held back by electromagnets. For each impact, the current is shut off, allowing the flexhammer to snap forward. The small, tight head is guaranteed to produce 500 million clear characters, equivalent to about five years' normal use, at a breakneck speed of 250 characters per second.

MARINE AND OFFSHORE

After the bottom fell out of the global tanker market in the mid-1970's, the Government took over all the main Swedish shipyards. Last June, Parliament said that each unit within Swedyards, the State-owned marine and offshore group, must break even by 1985 or face closure.

The big Götaverken Arendal yard in Gothenburg became the new offshore product arm of Swedyards. In 1977 Arendal started off by building North Sea jack-up oil drilling units, semi-submersible drilling and accommodation rigs based on designs licensed from Friede & Goldman of New Orleans. Arendal is the first yard in the world to carry out major modifications in these designs to fulfill rigorous new offshore safety and living standard requirements by Norwegian authorities. An example is 'Treasure Supporter', a twin-pontooned stabilized six-column rig with living quarters for 600 people. Delivered in March 1980, it can even withstand a collision with a supply ship.

According to Arendal's offshore project manager, Hans Rolfsman, one column of the rig cannot break off and allow the whole structure to tip over and sink, as the French-built hotel rig 'Alexander Kielland' did last year in the North Sea.

Arendal is also designing entirely new

types of icebreaking bulk and cargo ships for use in the Arctic, a region expected to grow dramatically in importance as a supplier of oil, coal and other minerals.

Meanwhile the Uddevallavarvet shipyard (UV) north of Gothenburg was first in the world to market so-called Aframax tankers of just under 80,000 dead-weight tons (dwt).

Designed with a shallow draught of 12.2 meters, allowing it access to most American West Coast ports, the Aframax has a comparatively short length (228 meters) compared to its width (42 meters) which gives it good maneuverability. Furthermore and most important, its broad beam gives the vessel a much larger payload than conventional tankers of the same draught and length.

The yard is trying out other designs incorporating a twin-skeg stern with dual engines and rudders, which have advantages both from a maneuverability and a fuel-consumption standpoint.

In the space of five years, UV has cut fuel consumption on its newly-designed ships by 20%. The yard has received STU funds for further energy-saving studies and believes that further modifications in hull shape and machinery systems might save another 10-15%.

Uddevallavarvet has been a leader in designing multigrade ore-bulk-oil (OBO) ships with fast turnaround times in port because of efficient loading and unloading.

TRANSPORTATION

ASEA's latest success in using hot isostatic pressure to make extremely corrosion-resistant ceramic turbine wheels, capable of withstanding temperatures of 1,400° C, "is a technical breakthrough that may open the way to the ceramic automotive gas turbine," says Gunnar Engström, the company's executive vice president for R&D.

As part of its collaboration with the Volvo subsidiary United Turbine (UT) in Malmö, southern Sweden, ASEA used the Quintus press at its high pressure laboratory in Robertsfors to produce complete silicon nitride turbine wheels, with a solid hub and blades that have trailing edges as thin as 0.3mm. Pulverized glass is the transmission medium in the process, which uses a pressure of 2,000 bar and heat of 1,700° C.

The ceramic turbine wheels can resist temperatures several hundred degrees centigrade hotter than the superalloys used in today's turbines. Raising the combustion temperature will allow a decisive improvement in the performance and fuel economy of the gas turbine.

Although a jet engine is nothing but a long single-shaft gas turbine, most efforts to adapt turbines to the automotive industry have failed, because in cars these machines have simply proved too expensive to run.

Now the odds may be tipping in favor of automotive gas turbines in this decade. This is due to the energy crisis, coupled with ASEA's breakthrough in ceramics as well as recent advances in turbine design made by Professor Sven-Olof Kronogård, head of UT. Seeing the economic potential, the Industry Fund has lent Volvo and UT nearly $5 million to build prototype turbines for cars by 1983.

The new KTT (Kronogård Turbine/Transmission) system is the end product of 25 years of R&D. It combines the engine, transmission and differential into a single compact unit. Professor Kronogård's ingenious twist is to add a turbine to the usual two.

The third turbine supplies auxiliary power via planetary gears to the first two, driving accessories like the fuel pump and alternator with 'recycled' energy that would ordinarily be wasted.

The KTT unit is half the size of a normal piston engine with automatic transmission, has less than a fifth as many moving parts, and produces about 140 hp. When fully developed it will be a big fuel saver as efficient as the best diesel engine, Professor Kronogård maintains. It can burn a variety of fuels, requires no harmful additives like lead to improve performance, uses little lubrication oil and is air cooled.

Another company in Malmö, United Stirling, is also working on an engine of the future. Half the company's R & D budget is financed by the U.S. Department of Energy and half by the Swedish Government. The Stirling engine works on continuous combustion, has multi-fuel capability, low toxic and particle emissions and quiet, vibration-free operation. There the similarities with turbines end. The Stirling engine converts heat from a burning air/fuel mixture to mechanical energy through alternating compression and expansion of a confined gas. Residual heat is recycled through a heat exchanger.

Its power, efficiency and fuel economy compare favorably with the diesel engine. Its emissions are 90% less and noise is much lower. Besides automotive use, it works well as a stationary electricity generator and in marine applications.

One big advantage of the Stirling engine is that it has the potential to burn solid fuels like wood chips or even coal. Tests in California's Mojave Desert this spring will determine whether a modified version can also operate on direct solar energy.

Another engine of the more immediate future is the new Saab-Scania system which enables a turbocharged automobile engine to run efficiently on gasoline of different octane numbers without prior adjustments, thereby saving the engine from possible damage. Designated the Automatic Performance Control (APC) or 'thinking engine', the system also reduces fuel consumption by as much as 8% and improves acceleration performance by up to 20%, the company states. APC will be included in the new 1982 model of the Saab 900 Turbo cars, to be introduced during the fall of 1981.

In the APC, an electronic unit analyzes data from a knock sensor mounted on the engine block and a charge pressure sensor on the intake manifold. It instantaneously instructs a solenoid valve to open or close the pressurized air supply to the control wastegate on the turbocharger.

Volvo, which exports over 80% of the cars it builds in Sweden and is now Scandinavia's biggest industrial group, has recently unveiled a prototype called the Volvo Concept Car (VCC).

The VCC uses aluminum and other lightweight materials wherever possible and has passive seat belts, a new rear axle suspension and a retractable front spoiler that reduces air resistance when the car is running at 70 km/h or faster. A TV-like screen on the dashboard provides fuel consumption readings from a trip computer as well as information like tire pressure and electric bulb malfunctions.

Meanwhile Volvo has come up with an innovation concerning buses. It is a flywheel system that absorbs energy from the braking action of a city bus and then uses it in the acceleration phase, cutting fuel consumption by 15-25%. In normal city traffic, a bus driver can shut off the diesel engine and drive the vehicle more than one kilometer using energy from the flywheel. This reduces air pollution, while the hydrostatic system that handles the energy transfer provides an unusually even, quiet acceleration, Volvo says. Since early 1980, two flywheel buses have been tested in regular service on the streets of Stockholm.

The Stockholm Public Transport Authority (SL) has just finished testing a computerized bus traffic control system from Datasaab on 60 of its inner-city coaches covering five routes.

A radio channel automatically broadcasts information to a central computer from each bus – including the location of the bus and the number of passengers being carried.

Transponders installed along the route record the passage of each bus. Aboard the bus, an odometer tells the computer the vehicle's position while it is traveling between transponders. The passenger-counting units use an infrared device and are mounted by the doors of the bus.

Two Swedish electronics companies, telecommunications specialist L.M. Ericsson and its ITT-owned rival Standard Radio and Telephone (SRT), have each developed a fail-safe computerized system to stop trains automatically before collisions or accidents can occur.

By late 1980 both the Ericsson and SRT

systems had been installed on more than 150 Swedish State Railway trains, while Ericsson's had also been ordered by the railroads of Norway, Denmark and Finland.

A display panel tells the driver, among other things, the maximum permissible speed of the train. The system issues several audible warnings if this maximum is being exceeded. If the driver does not act accordingly within a few seconds, it activates the brakes.

AEROSPACE

Sweden's long-standing policy of peacetime nonalignment and neutrality during wars led in the 1930's to the creation of an independent Swedish military aircraft industry.

Prime contractor for a whole series of Swedish defense aircraft has been the aerospace division of Saab-Scania at Linköping.

The latest version of the double-delta winged Viggen jet series is the all-weather fighter JA37 being produced for the Swedish Air Force , one of Western Europe's most efficient air defense organizations, until the late 1980's.

Like the other four types of Viggen, the single-seat JA37 is built to take off and land on a 500 meter runway or strip of ordinary highway. Its main weapons are Sky Flash air-to-air radar missiles but it is also equipped with Bofors air-to-surface rockets and Swiss Oerlikon automatic cannons. The pulse-doppler radar system was developed and built by L.M. Ericsson.

Last year Parliament allocated $48 million for initial development work during 1980-82 on a successor to Viggen.

Beginning around 1990, all Viggen types should be replaced with a single combat aircraft with the working name JAS. It will have to combine fighter, attack and reconnaissance functions but at low construction and operating costs.

Saab-Scania, Volvo, Ericsson and SRA Communications recently formed the 'Industry Group JAS', which will match the Government's development outlay and present a joint proposal on the aircraft by July 1981. The Government is scheduled to announce a final decision on the project during 1982.

Engineers at Saab-Scania say the aim of the exercise is to design an aircraft with Viggen's good performance but using the latest technology to cut fuel consumption in half. Smaller, more efficient engines and non-metallic lightweight materials will account for most of the expected fuel savings.

In the meantime, a company formed in 1979 by Saab-Scania and Sweden's Bofors (the steel, armaments and chemical group) is working on another defense-related aerospace project – the development of a new turbojet anti-ship missile system, the RBS15, to arm Swedish Navy 'Spica' torpedo boats by 1985. The company, Saab-Bofors Missile Corporation (SBMC), will later supply the Swedish Air Force with an airborne anti-ship missile system.

Otherwise the Saab-Scania aerospace division is in the throes of a major effort to shift 50% of its R&D capacity into civilian projects. Early in 1980 it therefore signed an agreement with Fairchild Industries of the United States for the joint design, development, production and marketing of a 34-passenger twin turboprop commuter aircraft with low fuel consumption.

The plane, known as the Saab-Fairchild 340, will have a cruise speed of 500 km/h and a range of 800 nautical miles. Regular production will begin in 1984.

Datasaab, a company owned in equal shares by Saab-Scania and the Swedish Government, has become one of Europe's leading suppliers of military and civilian air traffic control (ATC) systems as well as simulators for training their operators. The company is also a leader in financial terminals and small business computers.

In recent years Datasaab has designed and built complete ATC systems for the main Swedish airports and for the Dutch, Belgian and West German air forces.

Since the 1960's Sweden has been deeply involved in a space project in cooperation with several other Western European countries. Scientists have been using the Esrange base in Swedish Lapland, Europe's only wholly civilian rocket-launching area, to send up rockets as high as 600 km to study the Northern Lights (Aurora Borealis) and the effects of the sun on the outer atmosphere.

Now the Government in Stockholm has authorized Sweden's first space satellite, the Viking. It will be launched from French Guiana in May 1984 on a piggy-back flight aboard the European rocket Ariane.

The Viking's orbit will take it over both poles six times daily and its altitude will vary from 800 to 15,000 km.

It will be the first earth satellite to do extensive measurements of that portion of the magnetosphere where ionized particles are heated up and accelerated into high energy. Several Swedish scientific institutions will analyze the data.

Main contractor will be Saab-Scania, with Boeing providing the platform. The octagonal, 550 kg Viking is expected to have a life span of eight months.

The Swedish Space Corporation is coordinating the ambitious project, which is part of an overall Government strategy to launch Sweden into yet another high-technology export industry.

A LOOK AHEAD

The country's educational system has received a lot of attention in recent years. Critics complain that the quality of technical education has fallen off sharply and indeed statistics show that fewer qualified students pursue technical or scientific educations now than in the 1960's.

Some segments of industry are already feeling the pinch in qualified personnel. The Academy of Engineering Sciences (IVA) has warned that if action is not taken, Sweden could suffer from a shortage of educated manpower and might have trouble maintaining its high-level technological base.

In order to better assess future prospects for industry, IVA (the oldest engineering academy in the world), spent a year and a half and about $1.1 million compiling a major report on the country's future technical capability and industrial competence. Completed in 1979, the study gives a comprehensive overview of the industrial/business sectors and reviews R&D spending. The report urges more support for technical institutions and universities, plus increased support to STU's 'knowledge areas' and concludes with a shopping list of technologies where future R&D should be directed.

The report says Sweden's current prosperity (tenuous as that may be) stems from its ability to develop clever twists on existing technologies. The Swedes create new products and processes, are quick to adopt new production techniques, successfully exploit traditional market niches as well as open up new ones, and accurately target their R&D. Still, if the country's course is not altered to meet the problems that are anticipated, this may not be enough to keep Sweden from falling behind economically.

Professor Gunnar Hambraeus, director of IVA, comments. "Swedish industrialists are very good at finding market niches and wedging into them. Once in a niche it is very difficult to dislodge them. The market philosophy of many companies rests on the bedrock of being the best in one or two carefully selected technological areas."

Looking ahead to the next decade Hambraeus is still optimistic, "We have a solid knowledge base in pulp and paper, steels, chemicals, electronics and energy technologies. These will have to carry us through. But we should also develop other areas like microelectronics and biotechnology.

All things considered he has confidence in Sweden's future innovative capacity, "I'm surprised every year by the number of innovations that do manage to get through the system."

Constructing large and technically advanced projects is always a race. And your perpetual rival is time. Beat it—or it will beat you.

In many large projects one year gained or lost could mean a difference in total investment of up to 10%. Maybe millions of dollars—won or lost.

At Skanska you'll find all the resources a large construction project could possibly need.

We have technical experts to handle all stages of planning and design. We put thoroughly experienced people on the job site. And we run our own fleet of machinery—covering everything from housebuilding to dredging.

And the point is that all these resources are gathered within our company. Along with an open mind to combine them for true efficiency. To safeguard the time and cost limits of your project.

Through our years in international construction we have delivered many large and technically advanced projects in record time. And often ahead of schedule.

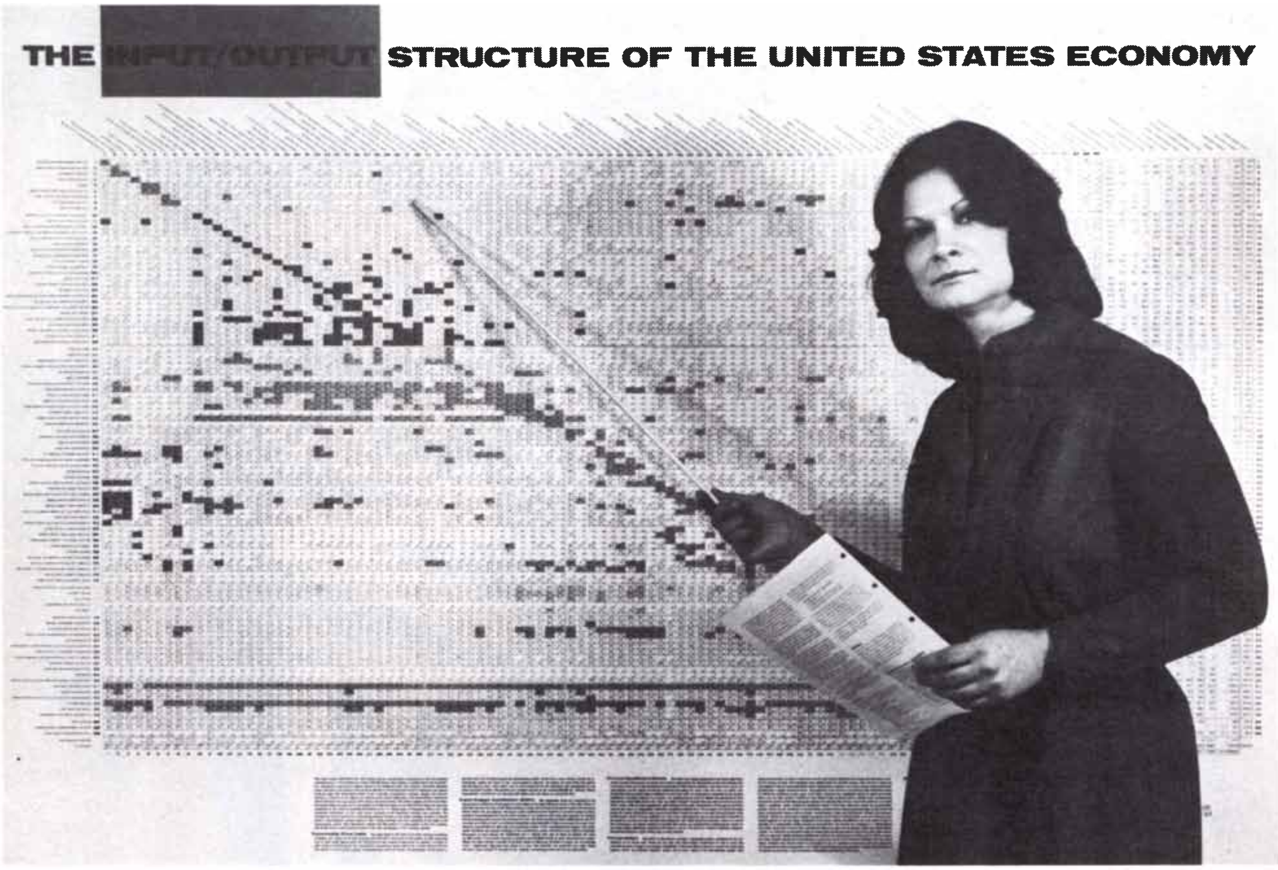
So let us carry your colours.

Skanska is one of Europe's leading contractors with worldwide operations. We build power plants, tunnels, harbours, bridges, hotels, hospitals and other large-scale projects—often on a design-construct or turnkey basis.

Skanska International Division, S-211 02 Malmö, Sweden.
Tel. +46-401 440 00. Telex 32247 Skanska S.

SKANSKA

Civil Engineering and Building Contractors



WHAT MAKES THE U.S. ECONOMY TICK?

The editors of Scientific American have prepared a wall chart displaying for the 1980's the Input/Output Structure of the U.S. Economy based on the latest interindustry study from the U.S. Department of Commerce.

The Scientific American Input/Output wall chart does for economics what the table of elements does for chemistry. It answers at a glance questions about the linkage between the microeconomics of the firm and the macroeconomics of the system; about the web of technological interdependencies that tie industry to industry; about the industry-by-industry direct and indirect consequences of swings in public and private spending; about the impact of change in technology, and about any other topic you can think of. You are rewarded by surprise as well as by confirmation of your hunches. For teaching and practical and theoretical studies, here is a powerful, graphic tool.

In the familiar format of the Scientific American Input/Output wall charts for the 1960's and 1970's, the wall chart for the 1980's measures 65″ × 52″ and is printed in eight colors. Each of the nearly 10,000 cells in the 97-sector interindustry matrix shows (1) the interindustry commodity flow, (2) the direct input/output coefficient and (3) the "inverse" coefficient. Where the direct input/output coefficient exceeds .01, the cell is tinted in the color code of the industrial bloc from which the input comes. This device, combined with triangulation of the matrix, brings the structure of interindustry transactions into graphic visibility.

A supplementary table displays, industry by industry, the capital stock employed; the employment of managerial, technical-professional, white-collar and blue-collar personnel; the energy consumption by major categories of fuel, and environmental stress measured by tons of pollutants.

The editors of SCIENTIFIC AMERICAN are happy to acknowledge the collaboration, in the preparation of this wall chart, of Wassily Leontief, originator of input/output analysis—for which contribution to the intellectual apparatus of economics he received the 1973 Nobel prize—and director of the Institute for Economic Analysis at New York University.

Packaged with the chart is an index showing the BEA and SIC code industries aggregated in each of the 97 sectors.

SCIENTIFIC AMERICAN

415 Madison Avenue, Dept. I/O, New York, N.Y. 10017

YES, send me the new Scientific American 1980's Input/Output wall chart—printed in eight colors on heavy paper measuring 65″ × 52″.

I understand my chart(s) will be shipped prepaid, uncreased*, flat-rolled and in a sturdy mailing tube. (Allow 4-6 weeks for delivery)

I enclose a check for $_____ for _____ chart(s) at $35 each, saving me $5 each in handling charges.

Bill me for $_____ for _____ chart(s) at $40 each.

Name____________________

Company____________________

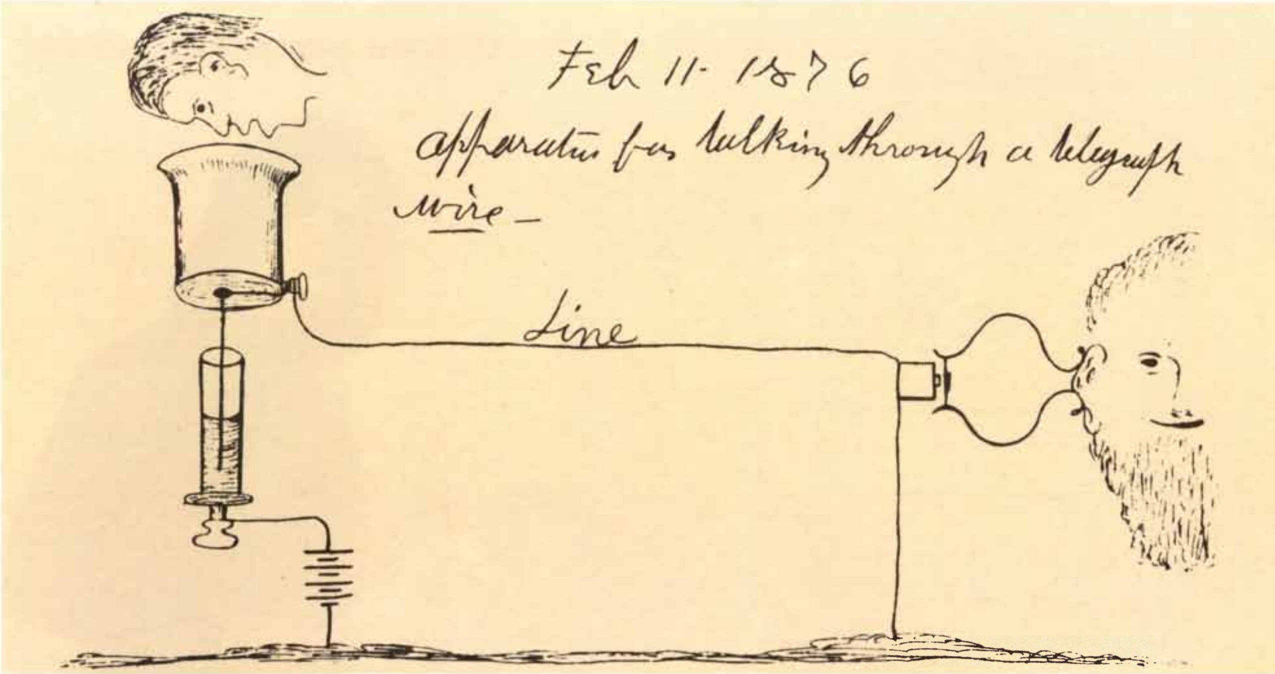
Street Address____________________

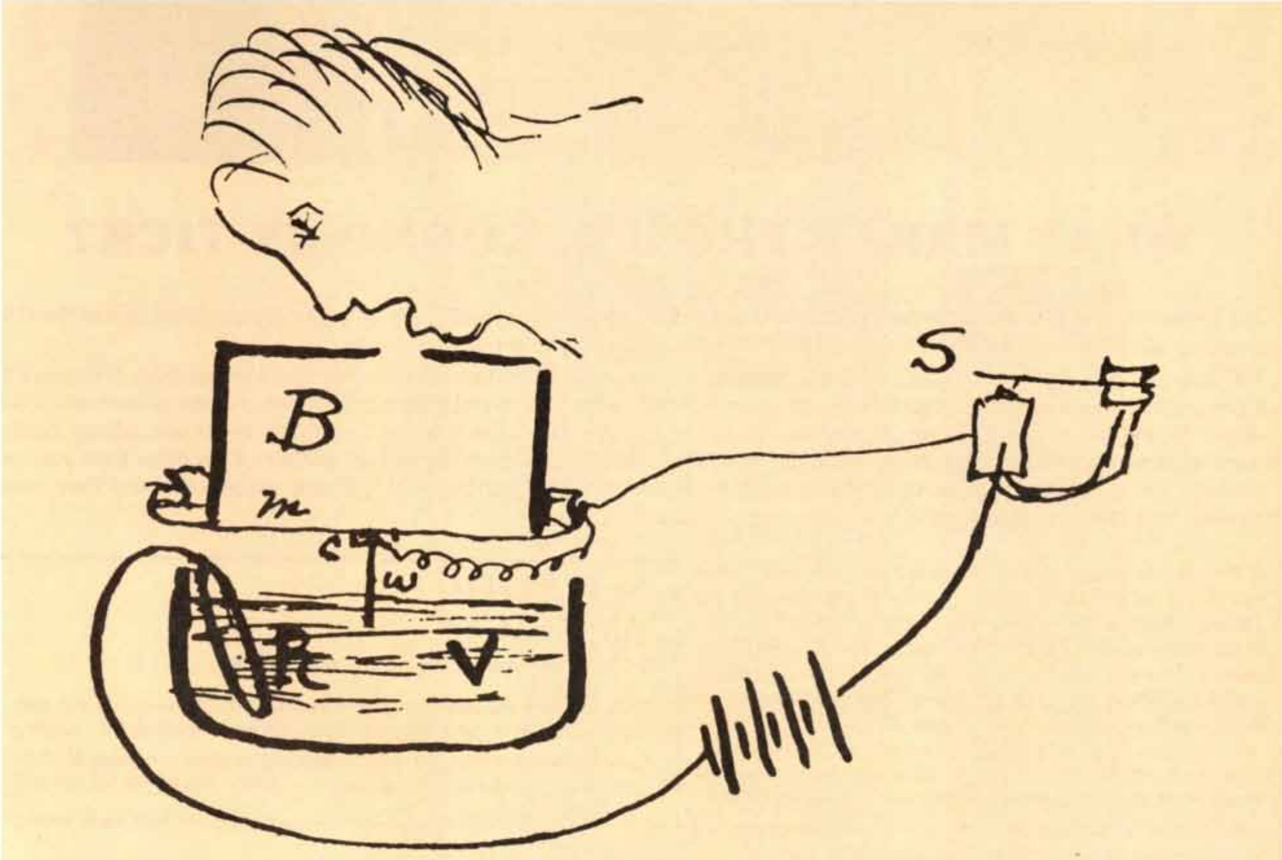
City, State, Zip Code____________________

JFM/81

*Charts shipped international require one fold.

Make check payable to Scientific American. Please add sales tax for California, Illinois, Massachusetts, Michigan, Ohio and New York.





STRIKING PARALLELS between the telephones envisioned by Elisha Gray and Alexander Graham Bell are evident in their respective sketches of the instruments. Both Gray's transmitter (*top*) and Bell's (*bottom*) depended on varying the resistance to the flow of current from a battery. Both variations would be caused by the vertical movement of a needle in a liquid bath; the motion would be due to the response of a diaphragm to the sound waves of the human voice. In Gray's transmitter the variation in resistance would depend on changes in the distance between the tip of the needle and the bottom electrode. In Bell's the variation would depend on the changes in the area of the wedge-shaped needle tip immersed in the bath. The varying current would then pass through an electromagnet (*right*) at the receiving end of the circuit; variations in the magnetic field would cause a second diaphragm (in Gray's scheme) or a metal reed (in Bell's) to vibrate, thereby reproducing the sound waves that actuated the transmitter. Gray made the sketch of his device on February 11, 1876, some two months after he conceived the idea. Bell made his sketch on March 9, 24 days after filing his patent application.

Two Paths to the Telephone

As Alexander Graham Bell was developing the telephone Elisha Gray was doing the same. Bell got the patent, but the episode is nonetheless an instructive example of simultaneous invention

by David A. Hounshell

In one day in 1876—February 14—the U.S. Patent Office received two communications describing the electrical transmission and reception of human speech by means of variations in the resistance of the transmitter. The variable-resistance device was the original telephone. The first description was in the form of a patent application by a 29-year-old amateur inventor whose name became world-famous: Alexander Graham Bell. The second description, which arrived only hours later, came from a 41-year-old professional inventor who had been granted the first of his many electrical patents almost a decade earlier: Elisha Gray.

Who was Elisha Gray? Why is Bell widely if not universally known as the inventor of the telephone and Gray, who envisioned the same device at the same time, known to few except historians of technology? To answer the question it is helpful to have an understanding of not only the technical aspects of this classic example of simultaneous invention but also the social ones. In the history of the telephone the differences between the world of the professional and the world of the amateur appear at almost every turn, as will be made clear by a brief exploration of the two worlds, first Gray's and then Bell's.

Elisha Gray, born in Barnesville, Ohio, in 1835, attended Oberlin College, but he was not graduated owing to ill health. His early interest in the electrical aspects of telegraphy led in 1867 to his first patent, for a self-adjusting telegraph relay. His device attracted the attention of the principal firm in the field, the Western Union Telegraph Company. It also earned Gray enough money for him to secure a partnership in a Cleveland firm that manufactured telegraphic instruments. He and his partner, Enos Barton, transformed the company into the nation's leading maker of electrical apparatus, renamed it the Western Electric Manufacturing Company and in the early 1870's relocated it in Chicago. In 1872 Western Union acquired a one-third interest in the business, and Western Electric eventually became the sole supplier of telegraphic equipment to Western Union.

Telegraphy had grown steadily in the U.S. from its introduction in 1844, and that growth had accelerated in the Civil War years. By the end of the war some 50,000 miles of telegraph wire had been strung over routes totaling nearly 30,000 miles, and a decade later the figures had risen to 250,000 miles of wire over routes of more than 100,000 miles. Until 1872, however, each wire could transmit only one message in one direction at a time. Enlarging the transmission capacity between any two points could be accomplished only by stringing new wires.

This one-way-at-a-time bottleneck was built into the telegraph system because the transmitted signal consisted of intermittent pulses of direct current. In 1872, however, Western Union adopted a "duplex" system of transmission developed by Joseph B. Stearns, a Boston electrician. A modification of the Morse system, the Stearns system allowed two messages to be transmitted simultaneously, one from each end of the wire. The adoption of the duplex system effectively doubled the capacity of the Western Union network. Could multiplex systems be developed, systems that would increase the capacity many times further? One thing was certain: the inventor of a multiplex system could virtually name his own price.

Gray's close relations with Western Union made him aware of the rewards awaiting the successful inventor, and early in 1874 he hit on a solution almost by pure accident. One day he found his nephew, whom he allowed to use his electrical apparatus, amusing himself in the bathroom with two battery-powered circuits. In one circuit the battery caused a reed "electrotome" to vibrate. The vibrations opened and closed the other circuit, which included an induction coil. Gray's nephew had also connected one lead of the second circuit to the zinc bathtub and was closing the circuit and "taking shocks" from the induction coil by rubbing his hand, which held the other lead, across the surface of the tub.

The vibrating reed made an audible hum, and when Gray's nephew rubbed his hand on the tub, a second hum of the same pitch was heard. Gray's interest was piqued. He changed the frequency of the reed and found that the sound made by the rubbing of his hand on the tub changed to match it. The action of the induction coil was to transform the on-off impulses imposed on the circuit by the vibrating reed into a sinusoidal wave of electric current.

Gray was quick to explore the phenomenon of what he named vibratory currents, seeking to find some practical use for their transmission and reception. Although he found no immediate application, he so resolutely believed vibratory currents would have a major usefulness that he resigned as the superintendent of Western Electric, determined to pursue the matter on an independent full-time basis. In order to do so he secured the financial backing of Samuel S. White, a wealthy manufacturer of dentistry equipment in Philadelphia.

With funds at his disposal Gray soon built four experimental devices: two transmitters of frequencies in the audible range and two receivers. One of the transmitters he called a single-tone transmitter; it was essentially a refinement of his nephew's bathtub apparatus. The other he called a two-tone transmitter; it was capable of simultaneously generating sinusoidal waves of two different frequencies. One of the two receivers seems quaint in retrospect; the other was quite conventional.

The first receiver consisted of a violin with its strings removed and a silver plate attached to the soundboard. When one of the leads of the induction-coil circuit was connected to the plate and the hand holding the other lead was rubbed across the metal surface, as it was in the bathtub experiment, the tones generated by either transmitter were reproduced with a richer quality. The other receiver consisted of an electromagnet and a metal diaphragm. When the magnet caused the diaphragm to vibrate according to the frequencies of the current in the induction-coil circuit, the tones of the transmitters were faithfully reproduced. The results suggested to Gray three possible applications.

The most obvious application and probably the easiest to perfect was what today would be called an electric organ; Gray thought of it as a "musical telegraph." He would merely have to build a keyboard consisting of switches that would actuate a series of single-tone transmitters, each tuned to a different musical pitch. It would even be possible to sound chords by pressing two or more keys simultaneously.

A more immediate commercial application was implicit in the ability of a telegraph wire to carry a "composite current," one consisting of two, four or a much larger number of frequencies. Could not each tone be made to carry a telegraph signal? The musical telegraph could function as a multiplex-signal transmitter if a receiver could be devised that was able to segregate the individual tones of the composite current. For this purpose, Gray realized, neither of his receivers would be of any use.

The magnet-and-diaphragm receiver, however, was perfectly suited to the third application Gray had in mind. If many combinations of tones could be carried by wire and the composite signal could then be reproduced electrically, would it not be possible to transmit the sounds of the human voice? Gray may have perceived an irony in this application. Just as the potential multiplex system of telegraphy was in need of a receiver, so the potential composite-current system of telephony was in need of a transmitter. Gray saw no easy way to solve the voice-transmitter problem, although he immediately envisioned a complex synthesizing device consisting of a number of individual tone transmitters, each responding to a different tone of the human voice.

RECEIVING APPARATUS, a stringless violin with a silver plate covering the soundboard, was devised by Gray in 1874 to draw attention to the "vibratory currents" first made audible by the bathtub experiment (*see top illustration on opposite page*). Here one lead from the induction coil connects to the silver plate. The demonstrator (Gray in this instance) holds the other lead in his hand and rubs his fingers along the metal surface. The musical tone thus generated duplicated the tone of his transmitter, a metal reed that vibrated at audio frequencies.

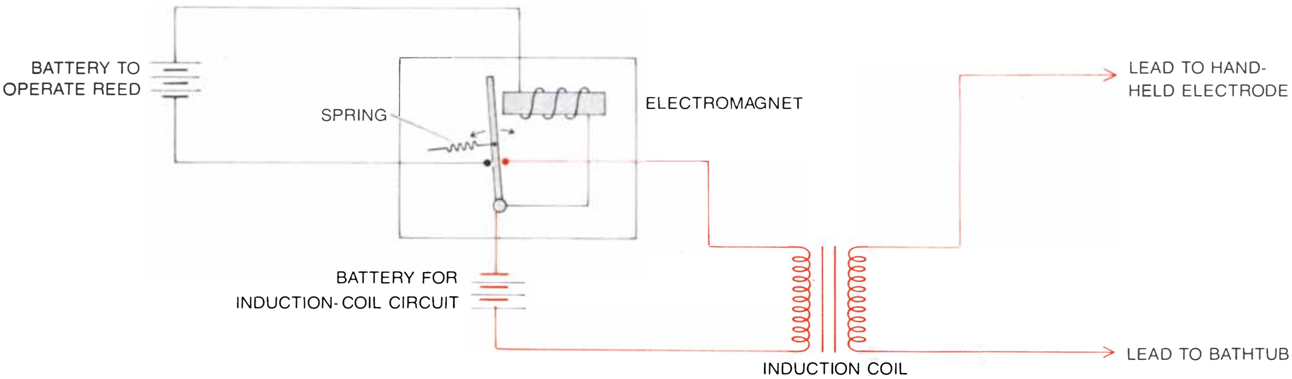
Within a few weeks of the bathtub episode Gray, in a rush of experimental activity, had discovered and in his mind explored the seemingly boundless possibilities of utilizing audible vibratory currents. In May of 1874, confident that he had sufficiently investigated the implications of these currents, he demonstrated his transmitters and receivers to audiences of telegraphy experts in Washington, New York and Boston. Reports of the demonstrations allow the inference that he mentioned all three potential applications of this work: the transmission of music, the transmission of multiple messages and the transmission of the human voice. Gray's tour gave rise to a flurry of debate.

The New York Times reported one Western Union official as saying that Gray had taken "the first step toward doing away with manipulating instruments [that is, telegraph keys] altogether." "In time," the official continued, "the operators will transmit the sound of their own voices over the wire, and talk with one another instead of telegraphing."

The leading journal of the industry, *The Telegrapher,* took the opposite tack. Declaring that the transmission of the human voice was nothing new, the journal cited an account it had published five years earlier describing the Reis telephone. Johann Reis, a German schoolmaster and experimenter, had in 1861 coined the word "telephone" to describe a laboratory device he had built to reproduce music and the voice. *The Telegrapher* noted that the Reis telephone had proved to have "no direct practical application" and remained "a mere scientific . . . curiosity."

The Telegrapher also mentioned what it said was an old joke in telegraphic circles. Voice communications had once been tried on the telegraph line between Philadelphia and New York, the story went, "but had to be given up on account of the Philadelphia operator's breath smelling too strongly of bad whiskey." The journal's view of the Reis telephone was echoed by A. L. Hayes, one of Gray's own patent attorneys and himself an expert in electrical technology. Assuring his client of the novelty and importance of the new discoveries, he went on to say the German instrument was "merely a toy" that could be made to work only with careful handling and that amounted to no more than a scientific oddity. Reis had died that same year, and interest in his "curiosity" had languished.

Discouraged by the negative opinions on voice communication, Gray turned in the summer of 1874 to the development of the remaining applications: the musical telegraph and multiplexing. For the transmission of music



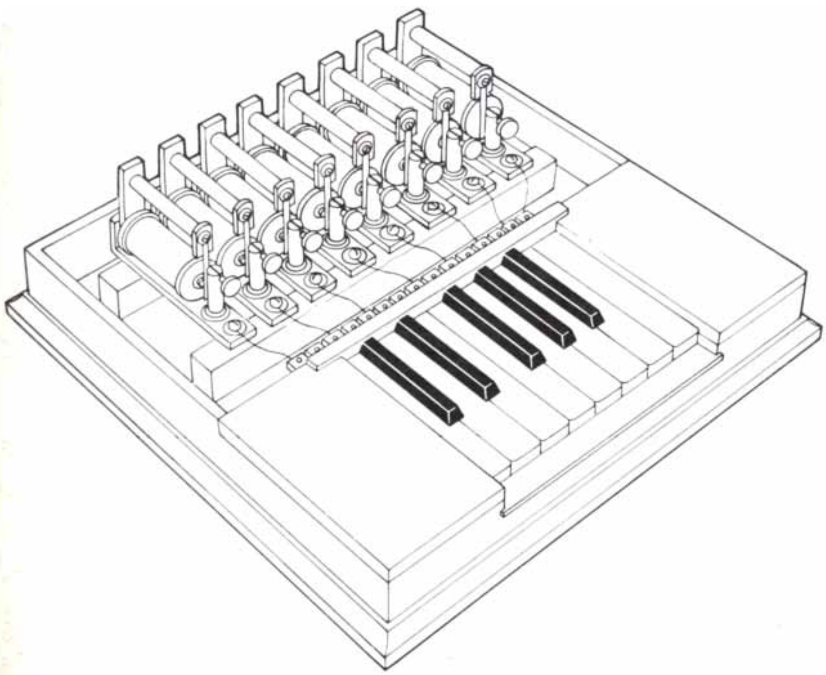
BATHTUB EXPERIMENT, based on a chance discovery by Gray's nephew in 1874, involved the circuits illustrated here. The first circuit, at the left, was closed and opened by the back-and-forth motion of a spring-loaded metal reed that vibrated at a fixed frequency. When the circuit was closed, it energized an electromagnet. The response of the reed to the attraction of the magnet reopened the circuit and the action of the spring promptly reclosed it. A second pair of contacts, one of them fixed to the reed, simultaneously closed and opened a second circuit (*color*). An induction coil in this circuit converted the interrupted current into a continuous sinusoidal current of the same frequency as the tone of the reed. When one of the leads from the induction coil (*right*) was attached to a zinc tub, and the experimenter, holding the other lead, rubbed his hand across the tub's surface, a sound was heard identical in pitch with the tone of the reed.

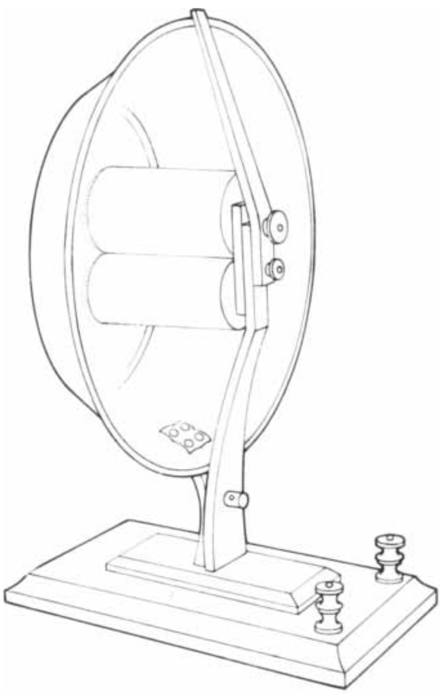
he built an organlike apparatus with an array of single-tone transmitters covering a range of one octave; later it was enlarged to a two-octave apparatus. To improve the tone quality of the receiver Gray made the diaphragm larger by replacing the flat metal plate with a washbasin; thereafter the device was known as the washbasin receiver.

In August and September of 1874 Gray toured England with these devices and others. Among those for whom he demonstrated them was John Tyndall, who succeeded Michael Faraday at the Royal Institution, and J. Latimer Clark, then perhaps the most eminent figure in British telegraphy. Gray also took advantage of the visit to test how well his vibratory currents performed when they were conducted by submarine cables. He concluded that there were no technical obstacles in the way of transforming musical telegraphy into a multiplex-message system, where both he and his backer White knew the greatest financial rewards would lie.

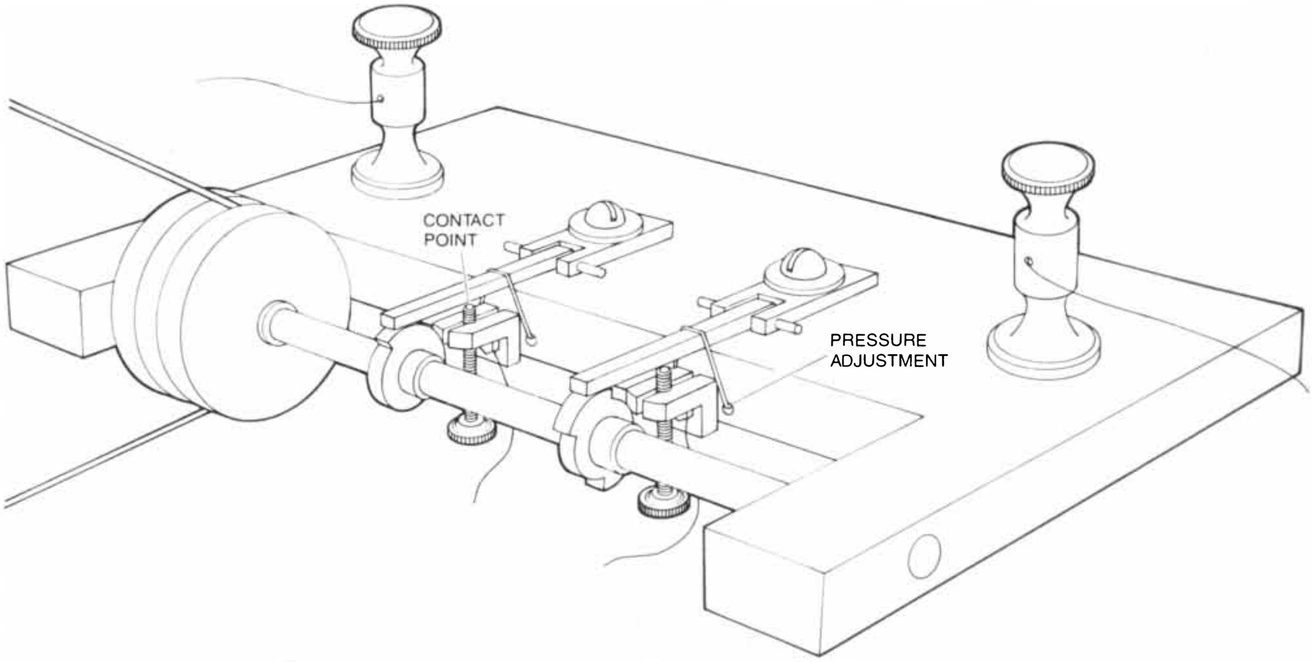
For the rest of 1874 Gray concentrated on multiplexing. In an effort to learn more about audio-frequency currents, as they would be called today, he built what he called a "mechanical" transmitter. The vibrating reed was replaced by two cams, mounted on a shaft rotated at speeds equal to audio frequencies, that opened and closed two sets of contact points. Adjustable "elastic springs" were included for regulating the pressure between the points.

The first time Gray tested his mechanical transmitter, on January 1, 1875, he observed something completely unexpected. With only one set of contact

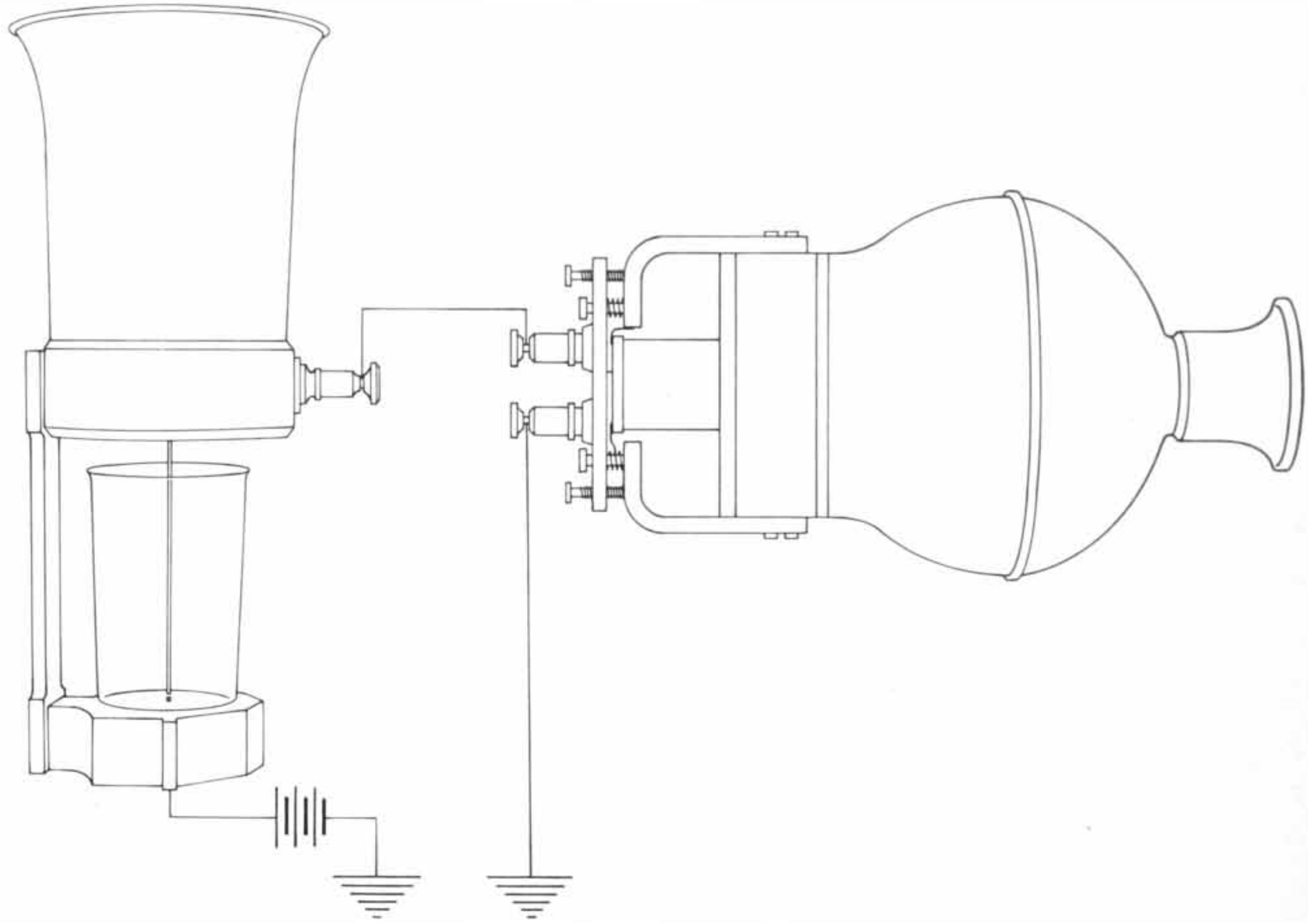




ADVANCED TRANSMITTER AND RECEIVER were two of the devices Gray took to England in 1874. The keyboard device at the left incorporated an array of single-tone transmitters that covered a range of one octave. The upended washbasin at the right formed the diaphragm of a receiving apparatus. Paired electromagnets translated "vibratory currents" into vibrations of the diaphragm.



SUBSTITUTE FOR REEDS, devised by Gray late in 1874, was a belt-driven camshaft. Two cams opened and closed sets of contact points, and springs allowed for adjustment of the pressure between the points. Testing his "mechanical transmitter" in the audio-frequency range on January 1, 1875, Gray found that changing the pressure between points altered the output of interrupted current in such a way that his washbasin receiver emitted various voicelike sounds. He concluded that a simple device could transmit the human voice.



GRAY'S TELEPHONE, in the form presented in his Patent Office caveat application of February 14, 1876, differed little from the apparatus he had sketched three days earlier. The slight changes in current produced by the movement of the transmitter diaphragm at the left resulted in electromagnet-powered vibrations of the receiver diaphragm at the right. Although Gray might have contested Bell's patent application, filed earlier the same day, he accepted the advice of his patent attorney and his financial backer and dropped the matter.

points in operation he found, not unexpectedly, that the washbasin receiver produced tones that varied in frequency according to the speed of the camshaft. When he then adjusted the tension of the spring, however, thereby either lessening or increasing the pressure between the points, he found he "was able to imitate many different [voice] sounds." He at once concluded that the complex synthesizing voice transmitter he had conceived of earlier was unnecessary. Some much simpler device could be made to transmit the human voice.

Even with the prospect of voice communication so much improved, however, Gray did not pursue it. Instead he chose to continue the development of a multiplex telegraph as a potentially far more profitable venture. His strides toward this goal are evident in the number of patent applications he filed early in 1875. He was soon to learn that his work was coming "into interference," as it is said in patent law, with the applications of another inventor. That inventor was Alexander Graham Bell.

By coincidence Bell was also in hot pursuit of a multiplex-telegraph system working within the audio-frequency spectrum. (What Gray called vibratory currents Bell called undulatory currents.) The youthful Bell was an amateur, but he knew that a pot of gold awaited the inventor of a practical multiplex telegraph. Furthermore, although Bell was an amateur inventor, he was a thoroughgoing professional in his own field: elocution and speech therapy. His father, Alexander M. Bell, professor of elocution at the University of Edinburgh, had won international recognition for his system of teaching the deaf to speak, and his son had an intimate knowledge of the physiology of human speech. Indeed, he had already put forward a theory of vowel tones.

One consequence of this work was profound. Alexander J. Ellis, a prominent British phonetician who had learned of Bell's theory, pointed out to him that the same theory had been advanced in a classic work, "On the Sensation of Tone as a Physiological Basis for the Theory of Music," by the German polymath Hermann von Helmholtz. Ellis also called Bell's attention to the electrically driven tuning fork used by Helmholtz in many of his experiments. Bell acquired a French edition of Helmholtz' book and read it during his passage from England to America in 1870.

When Bell settled in Boston in 1871 to teach at the School for Deaf Mutes, he found others who were closely acquainted with Helmholtz' acoustical work, among them Lewis Monroe, a friend of the Bell family's and professor of elocution at the Massachusetts Institute of Technology. Monroe told Bell that experimental apparatus of the Helmholtz type was available at M.I.T. and suggested that he and Bell repeat the experiments someday. He also lent Bell a copy of a recent book on acoustics by Tyndall. Bell was soon in contact with others who were interested in investigating sound, among them Charles R. Cross, assistant to the physicist Edward C. Pickering at M.I.T., and the Boston physician Clarence J. Blake, lecturer on otology at the Harvard Medical School.

The background of Bell's work on the telephone was therefore acoustical. The immediate path to that work, however, was like Gray's the multiplex telegraph. Bell began to pursue such an invention late in 1872, when he read an account in a Boston newspaper of Western Union's adoption of the Stearns duplex system. If Stearns had become rich by devising a system that could transmit only two messages at a time, and those in opposite directions, what wealth awaited the man who invented a system that could do more?

The Helmholtz apparatus gave Bell a starting point. Helmholtz had devised a way of generating an intermittent current at audio frequencies by using the vibrating end of one tine of a tuning fork as an interrupter. He then employed the intermittent current to drive other tuning forks. The method suggested to Bell a way of transmitting multiple messages. If several Helmholtz tuning-fork interrupters were tuned to different frequencies, each interrupter could transmit a separate message. This scheme left Bell facing the same problem Gray had faced: devising a receiver that could sort out the combined messages. Bell, however, was proceeding more on blind faith than Gray. He had yet to learn whether it was even possible to transmit composite tones by wire.

From late in 1872 through most of 1873 Bell worked off and on at his multiplex-telegraph scheme. Unlike Gray, who employed a professional instrument maker, Bell built his own apparatus. Because he had almost no mechanical skill the apparatus was crude. For example, at first he could not rig a Helmholtz interrupter that would work for more than a few seconds. His being an amateur at invention was no mere artificial distinction.

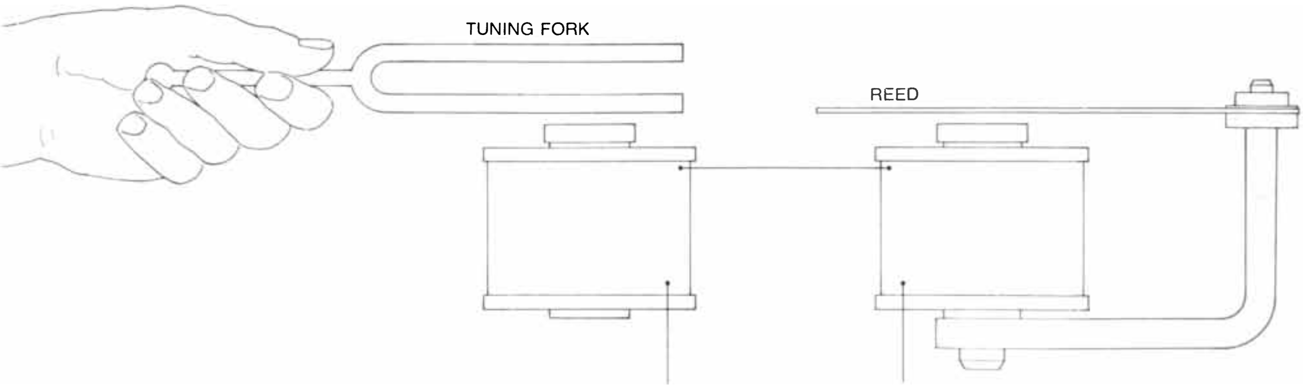
Late in the spring of 1874 Bell learned about Gray's work with vibratory currents. He at once accelerated the pace of his own work. Up to then Bell's efforts had been almost without success, but his faith in his multiplex scheme remained strong. That faith might well have faltered, however, if he had not found himself in partnership with Gardiner G. Hubbard, a telegraphy enthusiast. Hubbard had hired Bell to tutor his daughter, who had been made deaf by scarlet fever. Hubbard viewed Western Union as an enemy of progress in telegraphy, believing the monopoly was stifling innovation. The two men naturally talked about telegraphy, and in October of 1874 Bell disclosed that he was working on a multiplex scheme. Hubbard had predicted even before he knew Bell that "one wire [might eventually] be used for four or possibly eight messages." He now provided Bell with enough money to hire an expert in the field of electrical studies and pay for the services of the expert's instrument maker. The quality of Bell's apparatus soon improved. Perhaps Hubbard's most important contribution to the partnership was his insistence that Bell keep his mind on the main goal: multiplex telegraphy.

By the summer of 1875 Bell's work had convinced him it was possible to transmit speech, the same conclusion that had been reached earlier that year by Gray. Although Gray had set the notion aside in favor of further work on multiplex telegraphy, Bell's keen interest in the voice made him feel that such an achievement would be of the first importance. Like Gray, Bell had initially conceived of a complex voice transmitter, but experiments with tuned steel reeds in transmitters and receivers led him to devise an instrument for both purposes that consisted of a reed attached to the center of a diaphragm. Both the transmitter and the receiver were placed near one of the poles of an electromagnet with a slightly but permanently magnetized core.

With this pair of instruments Bell and his assistant were able, on July 1, 1875, to transmit and receive what Bell described as "vocal sounds." These sounds were not, however, speech. The system required that Bell or his assistant shout into the diaphragm of the transmitting instrument. The diaphragm was agitated by the sound waves, which caused the reed to vibrate. The motion generated a weak "undulatory current" in the transmitting electromagnet, which actuated the receiving electromagnet, causing the receiving diaphragm to reproduce the movements of the transmitting diaphragm.

Bell was captivated by the potential of these instruments. (It is worth noting that the first commercial telephones, which appeared in 1877, were not much more than improved versions of the 1875 magnet-and-diaphragm instruments.) Hubbard, however, was less than interested; like Gray and Gray's backer he kept his sights on the development of the multiplex telegraph.

By this time Gray and Bell were playing cat and mouse with each other. Each suspected that the other was spying on him; each believed his own work was the more advanced, but each wor-



BELL'S EXPERIMENTS, urgently pursued after he had filed his patent application, included some, such as the one illustrated here, that used electromagnets in both the transmitter and the receiver. Bell had tried a similar approach in 1875 but without success. He did not return to the magnetoelectric transmitter-receiver system until after he had successfully transmitted speech with a variable-resistance instrument. Here the vibration of a tuning fork near the electromagnet at the left caused its magnetic field to fluctuate. The fluctuations produced corresponding fluctuations in the current and hence in the field of the electromagnet at the right, making the steel reed above it vibrate with the same frequency as that of the tuning fork. A more advanced state of such a transmitter-receiver was the system Bell exhibited at the Centennial Exhibition in Philadelphia in June, 1876. It was the one used for the first commercial telephone in 1877.

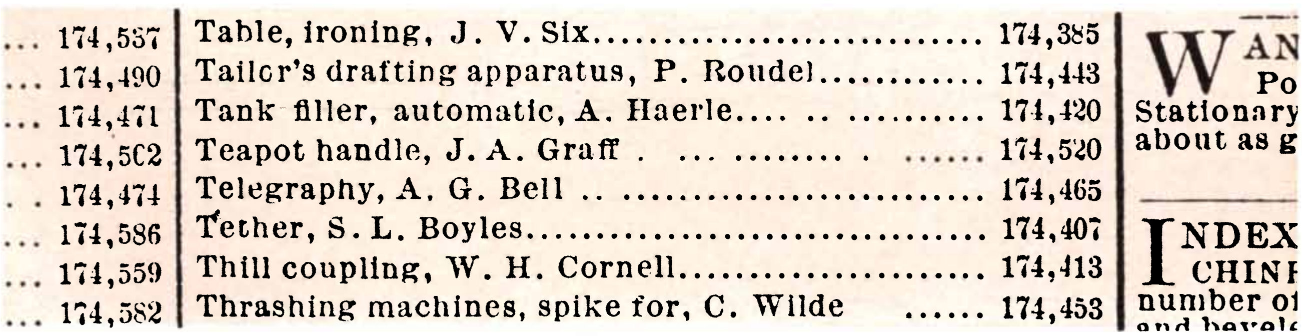
ried that the other might achieve the decisive breakthrough. Gray, however, gradually came to the conclusion that Bell's effort had been derailed. In October of 1875 Gray wrote his patent attorney: "Bell seems to be spending all his energies in [the] talking telegraph. While this is very interesting scientifically it has no commercial value at present, for [the telegraph industry] can do more business over a line by methods already in use than by that system. I don't want at present to spend my time and money for that which will bring no reward."

All the same, when later that month Gray happened to see two boys playing with a homemade toy known as a "lovers' telegraph," he immediately realized how an electric telephone should be constructed. A lovers' telegraph is what would be known today as a tin-can telephone. Two cans with the top removed are connected by a string knotted inside a hole punched in the bottom of each can. When the string is stretched tight, sounds uttered into one of the cans make its bottom vibrate. The vibrations are carried by the taut string to the bottom of the other can, where the sounds are coarsely reproduced.

Gray recognized the electrical analogue of the toy. The electric transmitter would consist of a voice chamber (the can) and a diaphragm (the bottom of the can). If one end of a wire was attached to the diaphragm and the other end was immersed in a liquid with a high electrical resistance, the movement of the wire in response to the vibrations of the diaphragm could be transformed into a vibratory current that faithfully reproduced the various frequencies of speech. Gray already knew that his electromagnet-and-diaphragm receiver could turn the vibratory current back into sound waves.

Although the device seemed to Gray to lack commercial application, he believed it would work and intended to patent it. He already knew that the movement of a wire in a liquid could produce vibrations in an electric current. While he was still a proprietor of Western Electric the company had made and sold liquid rheostats: devices that changed the resistance of a circuit by varying the depth to which a metal rod was submerged in a liquid with a high electrical resistance. Because of his pressing work on the multiplex telegraph, however, he waited more than three months before he took any patent action. In February of 1876 he finally put his ideas on paper. Instead of applying for a patent, however, Gray filed what was known at the time as a caveat, or warning.

A caveat was supposed to give the Patent Office formal notice of an inventor's basic concept. The idea was that after filing a caveat the inventor would develop the concept into a workable device and then apply for a patent. The procedure was intended to give a certain amount of protection to inventors' concepts. On this occasion Gray's patent attorney wrote to White, Gray's backer, that the inventor's "talking telegraph caveat" potentially interfered with a patent application from Bell. "As Gray's caveat was filed on the same day as Bell's application, but later in the day," the attorney reported, "the Commissioner holds that he is not entitled to an interference and Bell's application has been ordered to issue.... We could still have an interference by Gray's coming tomorrow and promptly filing an application for a patent. If you want this done, telegraph me in the morning, on receipt of this, and I will have the papers ready in time to stop the issue of Bell's



... 174,537	Table, ironing, J. V. Six	174,385
... 174,490	Tailor's drafting apparatus, P. Roudel	174,443
... 174,471	Tank filler, automatic, A. Haerle	174,420
... 174,5C2	Teapot handle, J. A. Graff	174,520
. . 174,474	Telegraphy, A. G. Bell	174,465
... 174,586	Tether, S. L. Boyles	174,407
... 174,559	Thill coupling, W. H. Cornell	174,413
... 174,582	Thrashing machines, spike for, C. Wilde	174,453

WAN
Po
Stationary
about as g

INDEX
CHINE
number o

PATENT NO. 174,465 was one of 270-odd granted by the U.S. Patent Office during the week ending March 7, 1876, three weeks after Bell had filed his telephone patent application. This "official notice" was published in the April 8, 1876, issue of SCIENTIFIC AMERICAN.

patent, but my judgment is against it."

When the letter reached White, the inventor happened to be visiting him. White firmly impressed on Gray that he should concentrate on multiplex telegraphy. Since the attorney too had advised against proceeding further, Gray dropped the matter. The telephone was left to Alexander Graham Bell.

The experts in telegraphy, including Gray, had concluded that the telephone was not worth serious attention. Bell had remained steadfast in his belief that a successful telephone would be an invention of the first importance. Most of his draft patent application dealt with the devices tested in July of 1875 (which came to be known as the "magneto-electric telephone"). As an afterthought, however, he inserted a description of a different kind of transmitter.

"Electrical undulations," he wrote, "may also be caused by alternately increasing and diminishing the resistance of the circuit.... For instance, let mercury or some other liquid form part of a voltaic circuit. Then the more deeply the conducting wire is immersed in the liquid, the less resistance does the liquid offer to the passage of the current. Hence the vibration of the conducting wire in a liquid included in the circuit occasions undulations in the current."

Bell had clearly misjudged the electrical properties of mercury; it is a low-resistance liquid, not a high-resistance one. Why he did so is something of a puzzle, because he had been familiar with the electrical properties of mercury since he had first read about them in Helmholtz. It is also not entirely clear just how Bell conceived of a variable-resistance transmitter, that is, what precedents he relied on. There is even some question as to how and when the descriptive paragraph found its way into his final patent specification. Finally, one may wonder why he filed an application rather than a caveat. His concepts had not been, in terms of patent law, "reduced to practice," a fact the Patent Office either overlooked or ignored. These questions have been sources of speculation for more than a century and need not detain us here.

In June of 1876 Gray saw Bell demonstrate his telephone at the Centennial Exhibition in Philadelphia. He later told an associate: "As to Bell's talking telegraph, it only creates interest in scientific circles. [Its] commercial value will be limited." So did the professional Gray continue to misjudge the importance of the telephone even after its successful realization. In contrast, the amateur Bell wrote his father two weeks after filing his patent application (and nearly two weeks before he was first to hear human speech through his instrument): "*The whole thing is mine*—and I am sure of fame, fortune and success."

THE AMATEUR SCIENTIST

More about edifying visual spectacles produced by laser

by Jearl Walker

Last August I described some of the dazzling and instructive visual effects that can be achieved with the pure, intense light of lasers. I said I would later set forth some ideas on how an amateur might prepare a laser-light display in the home or the classroom. I shall now do so, giving in addition some of the many suggestions that came from readers as a result of my earlier piece.

The musical group Genesis has been accompanied by laser for a long time, and it was at one of the group's concerts that I saw a particularly beautiful display. During one of the numbers stagehands blew onto the stage a thick fog of dry-ice vapor. A laser beam was directed upward through the fog at a small angle to the vertical. The beam was visible because small droplets of the fog scattered the light. A mirror mounted on a motor caused the beam to rotate about the vertical at a speed that matched the tempo of the music. The beam originated in a pulsed gas laser, so that the cone formed by it was chopped into discrete lines. I think the gas must have been a mixture of krypton and argon, which in a laser can generate more than one color, because what I saw was a glittering, rotating cone of light made up of four different colors. The effect was magical.

David Yoel, one of my students at Cleveland State University, devised a means of creating a more modest cone of light that my class would see end on rather than from the side. We had neither a multicolor laser nor a machine for making dry-ice fog, and so we worked with our 15-milliwatt helium-neon laser and smoke generated by a magician's stage explosive. The laser light was reflected by a mirror Yoel had mounted on the end of a rotating shaft by means of a hinge and a spring.

When the shaft was stationary, the mirror was perpendicular to it and reflected the laser beam direct to the middle of a screen. When the shaft was rotated, centrifugal force caused the mirror to tilt. The amount of tilt increased with the speed of the shaft, thereby increasing the deflection of the beam. The rotation swept the beam around in a cone centered on the middle of the screen. Yoel controlled the size of the cone by adjusting the speed of the shaft. The stiffness of the spring is important in this apparatus; keep trying springs until you find a suitable one.

If you have a multicolor laser, you can split the cone into different colors by putting into the beam a dispersing device such as a prism or a diffraction grating. Rig a motor that will make the device oscillate back and forth through the laser beam. The colors emitted by the laser emerge at different angles because of the dispersion. As the prism or grating oscillates, the emerging colors are swept across a slit the light must pass through to reach the rotating mirror. The color that goes into the air then varies as the prism or grating moves. Plans for making a laser that emits several colors appear in *Light and Its Uses,* which is cited in the bibliography for this issue [*page 170*].

The explosion of gunpowder serves the purpose of putting into the air a number of particles that will scatter the laser light. The particles augment the visibility of the cone. If the air were perfectly clean, neither the beam nor the cone would be visible except by reflection from the screen. We rig four flashbulbs that can be fired when a single button is pushed. They ignite lengths of magician's flash paper, which set off gunpowder sprinkled on the paper.

The particles are in the size range where they scatter light by both diffraction and reflection. As a result the light is strongly scattered forward and backward but weakly scattered to the sides. For demonstrations to students I put the laser on a stage and aimed it toward optical systems I had set up near the back of the room. The rotating mirror reflected the laser light back to a screen at the front of the room. (If you attempt a similar demonstration, make sure the beams are high enough to be well above the head of a tall person standing up in the room. Fix the laser and the optical systems rigidly in place to prevent an accidental misalignment from directing the beam into the audience.) When the laser was switched on just after the explosions, the air above the audience was laced with shimmering threads of red light.

Our laser's output of 15 milliwatts was more than enough for any one display, and so Yoel positioned small glass slides at an angle to the beam in order to achieve a variety of optical effects by splitting off part of the light. A slide transmits most of the light reaching it but reflects some. In order to control the direction of the reflected light he set up several solenoids, each with a small mirror mounted on its shaft. When a shaft was extended from a solenoid, the mirror directed the beam to a particular array of optical devices. When the shaft was retracted, the beam passed the mirror and went to another optical system on the other side of the solenoid. Yoel could shift the beam from one system to the other by means of a switch that controlled the current to the solenoid.

In my August piece I offered to send readers samples of a particularly useful Ronchi filter. Readers who obtained the filter can create sharper images in an interference pattern by peeling off the protective layer of Mylar on one side of the grating. I still have filters; if you would like one, send $1 and a mailing label to me at the Physics Department, Cleveland State University, Cleveland, Ohio 44115. The display generated by the filter is best when the laser develops at least a few milliwatts of power. It may be disappointing with weaker lasers.

James Watson of Ball State University called my attention to a Ronchi filter that is sold by the Holex Corporation (P.O. Box 27056, Philadelphia, Pa. 19118). The filter gives rise to a rectangular array of diffraction spots. Vic Rice of Campbell, Calif., sent me a sample of a halftone screen he employs to create the same optical effect. The screen can be either positive or negative.

In August I also described how interference fringes can be created by directing laser light through a sheet of clear plastic that has been coated with airplane glue or Duco cement. I tried adding small pieces of plastic and metal to the glue. The best results came from plastic strips, none more than a millimeter wide, that are sold under the name Diamond Dust by the Permafrost Division of Potter Industries, Inc. (377 Route 17, Hasbrouck Heights, N.J. 07604). The material is also sold in hobby shops. With this system I got not only bright and dark fringes but also bright spots resulting from the refraction of light through the pieces of plastic. When I rotated the sheet through the beam, which was expanding because I had put a convex lens in its path, the bright spots slid around on the screen.

Several of the people who sent for Ronchi filters also told me about their own experiments with laser light. Christopher Heilman has devised a way to project a laser image that resembles a photograph of the sun made in one color: the color of an emission line of hydrogen. He employs a lens to expand the beam from his five-milliwatt helium-neon laser and then directs the beam through a small glass sphere of uneven density. The result on his screen is an illusory sphere.

Roger Warden works with a textured plastic disk that he mounts on a slowly turning motor. When the laser beam intercepts one of the deep scratches he has made on the plastic, a starlike pattern appears on the screen. Warden also suggests trying small drops of water on the plastic. Frank B. Fadich uses cut lead-glass crystals and plastic figures to create patterns with his five-milliwatt helium-neon laser. He also creates interesting figures when he inserts in the beam a small transparent ball with many flat faces. Such balls are available in hobby shops. Gertrude Reagan generates interference patterns by directing a laser beam through the neck of a gin bottle and out through the bottom. The patterns vary as she moves the bottom of the bottle around in the beam.

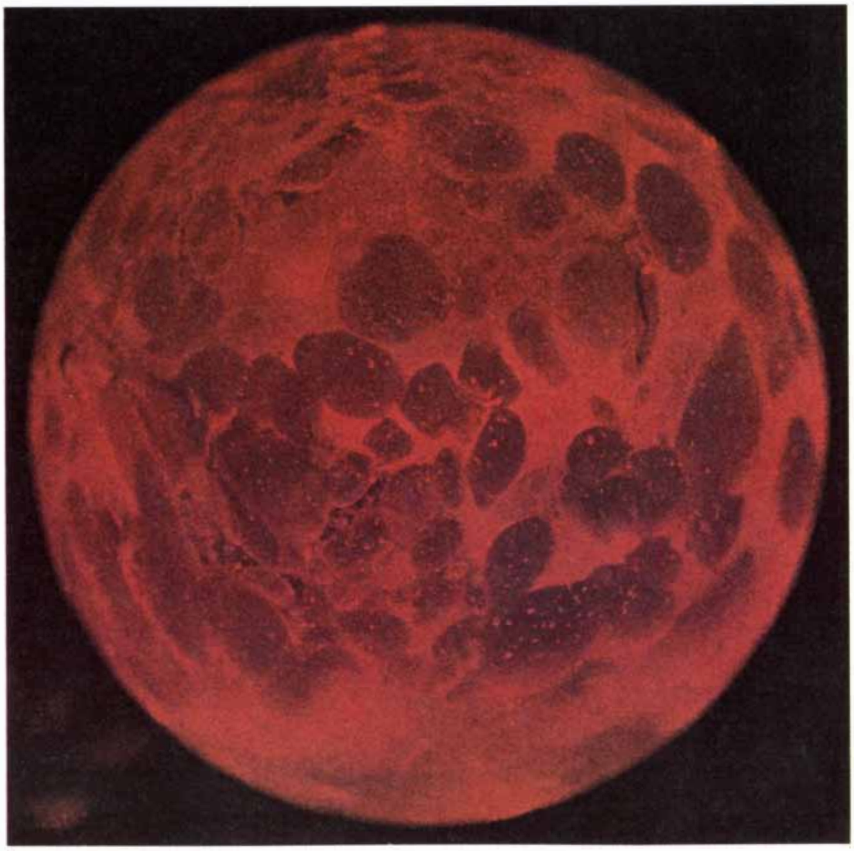
Gary F. Benedict of Chandler, Ariz., chops a continuous beam, such as one from a helium-neon laser, by inserting in it an opaque disk with teeth on its perimeter. As the disk is rotated the teeth periodically block the beam. This effect may be useful in obtaining a cone of many rays of light like the one I saw at the rock concert. Instead of totally blocking the beam you could work with a disk that has reflecting teeth. The light is reflected to other optical systems whenever the teeth are in the beam. Benedict also suggests using a rotating cylindrical glass rod with its long axis perpendicular to the laser beam. When the rod is inserted in the beam, it acts as a convex lens and redirects the beam. If the rod is rotated rapidly, the persistence of vision holds the pattern that is swept out.

Many people wrote to me about ways of reflecting a beam to form patterns on a screen. Tom Glanzman of Duke University reflects a laser beam from an encapsulated pool of mercury resting against a loudspeaker. The vibrations from the speaker generate ripples on the surface of the mercury. The reflection of the laser beam varies accordingly.

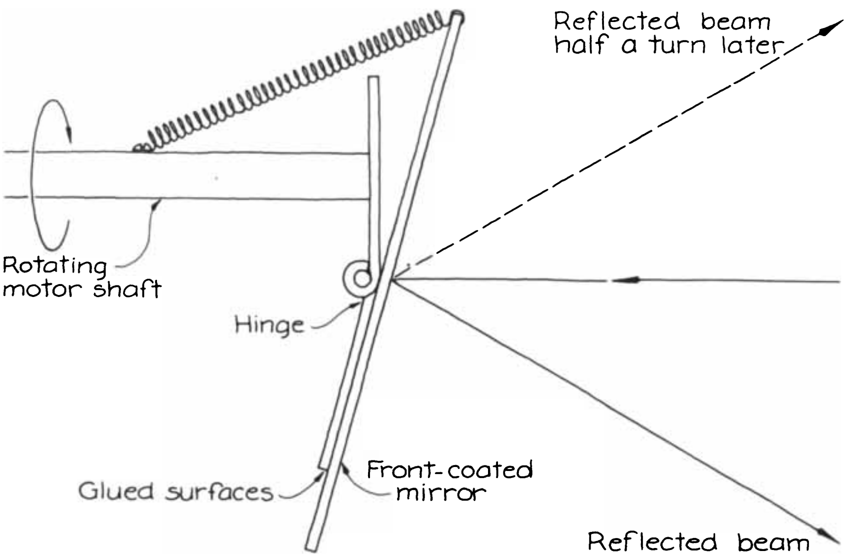
Robert J. Kearney of the University of Idaho mounts a small mirror on a rubber diaphragm that he attaches to an inexpensive speaker. The response of the mirror, he says, is relatively insensitive to the thickness of the rubber. The system can be adjusted by adding small weights to the rubber in order to control its patterns of vibration. (Any mirror used with a laser should be front-coated and of good quality, otherwise the beam will be excessively dispersed.)

One of the most interesting techniques came from Arthur Eisenkraft of

A pattern formed by laser light directed through a sheet coated with glue and plastic strips



The pattern formed through Christopher Heilman's glass sphere



David Yoel's rig for achieving a cone of light

Ossining, N.Y. He mounts a lightweight mirror on a speaker by means of a small loop of tape. The mirror is available from Metrologic Instruments, Inc. (P.O. Box 307, 143 Harding Avenue, Bellmawr, N.J. 08031). The mirror can move freely because of the looseness of the mounting. Eisenkraft tinkers with the position of the mirror until it reflects a laser beam both vertically and horizontally with equal strength. Without the adjustment the mirror would reflect the beam along an axis on the screen, thereby giving elongated patterns.

Eisenkraft was interested in the patterns that result when records of various kinds are played over the speaker. With "Sgt. Pepper's Lonely Hearts Club Band" by the Beatles he got a mixture of pure Lissajous figures and erratic designs. The pure patterns were probably generated by the electronic music the Beatles meshed into the song.

The best pieces for creating laser tracings on the screen were recordings of electronic music by Keith Emerson. In particular Eisenkraft suggests the *Karelia* Suite by Sibelius, which is part of the album "Five-Bridge Suite" by Emerson and the Nice.

I have had good luck with much of the electronic rock music recorded by European groups. The Tangerine Dream group works almost exclusively with synthesizers. When its music is played on the two-mirror speaker I described in August, hypnotic Lissajous figures appear on the screen. The single-speaker system I also described in August responds only to electronic music in the lower frequencies.

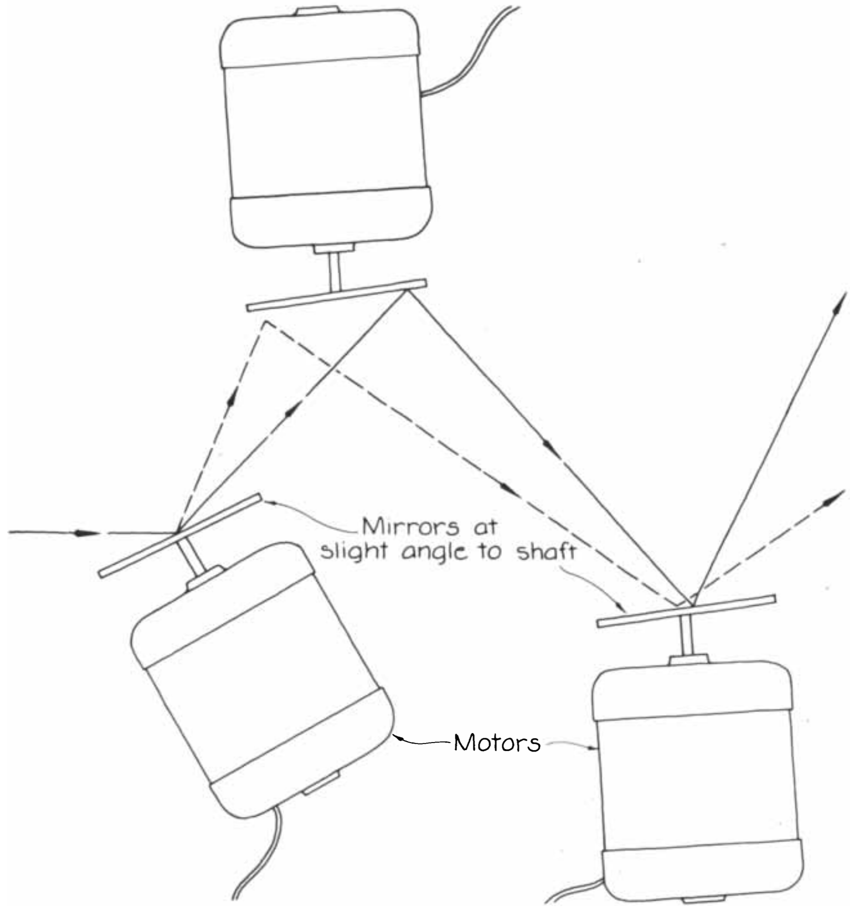
In addition to turning electronic music into patterns on a screen you can convert the light back into sound. Eisenkraft demonstrated this capability by inserting a photocell in the beam. The signal from the photocell went successively to a preamplifier, an amplifier and a speaker, thus recovering the electronic signals encoded in the original music.

Teachers whose classroom equipment includes a ripple tank can mount mirrors on the paddle that generates the ripples. Several people also suggested mounting a mirror on the shaft of a motor with the plane of the mirror at a slight angle to the shaft. When the motor is engaged, the mirror turns, forming the beam into a cone of light. If the beam is chopped before it reaches the mirror, the effect is similar to the cone I saw at the Genesis concert. On the screen you see not a continuous trace of a pattern but dots that appear to chase one another along a curved path.

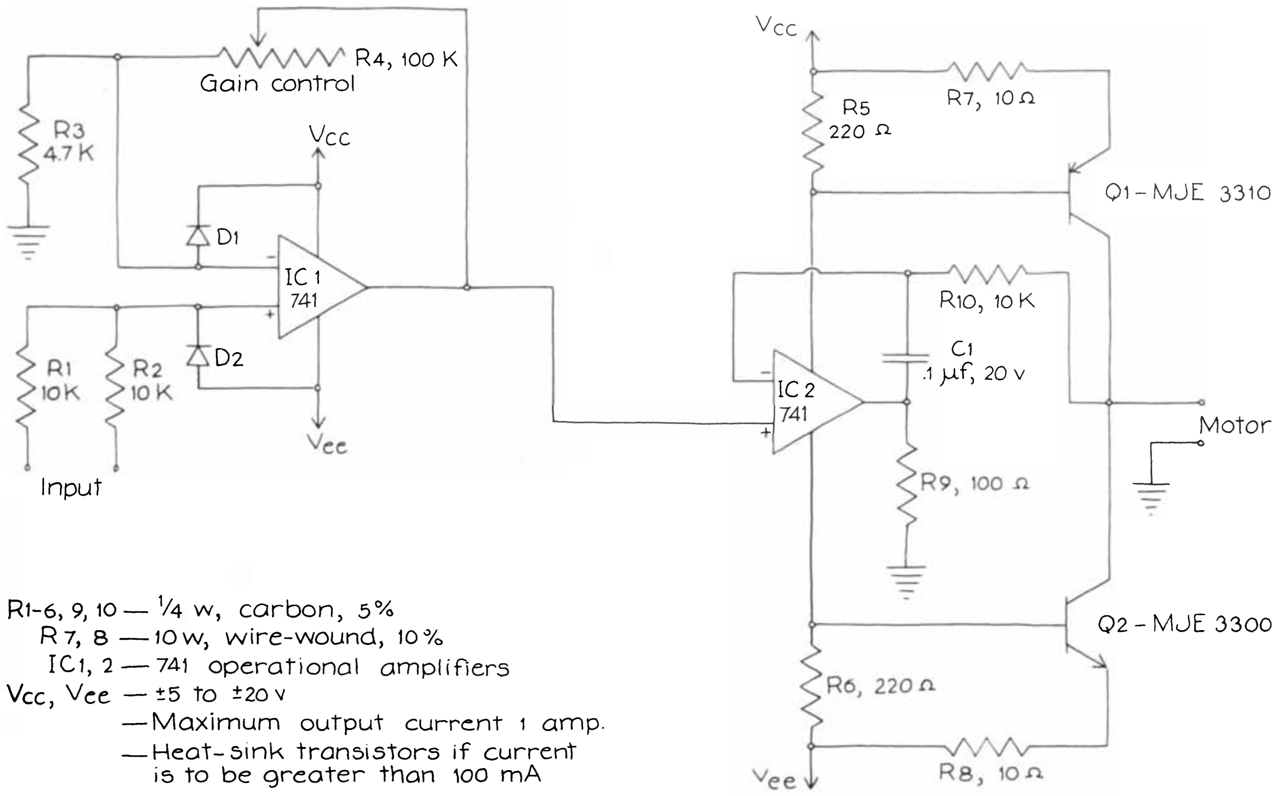
Gary E. Tomlinson of Grand Rapids, Mich., told me about a system of three such rotating mirrors. Each mirror is glued to the end of a shaft at an angle. By adjusting the angle of tilt of each mirror and the arrangement of the motors Tomlinson creates circles, ellipses, stars, squares and other figures on his screen. A patented system of this design is available from Lasertrace (26 Station Road, Westgate-upon-Sea, Kent, England CT8 8RT).

Gregory Yob of Palo Alto, Calif., has built a system of five motors with tilted mirrors mounted on their axes. He arranges for the mirror with the smallest tilt to be the first one reflecting the laser beam, which then goes to each of the other mirrors in succession. The power supply for the motors is shown in the illustration on the opposite page. Yob says the circuit is designed to make the motors operate synchronously.

Before the beam reaches the mirrors it



Gary E. Tomlinson's arrangement of three tilted, rotating mirrors



The circuit for the power supply of Gregory Yob's five-motor mirror system

is chopped by a wheel with 16 triangular teeth. The wheel is mounted on a motor that is attached to a lead screw. The motor is connected to a slide potentiometer so that Yob can control the speed of the wheel. The lead screw is driven by a gear-reduced motor. With this rig Yob can vary the extent to which the teeth intersect the laser beam. The more they cut into the beam, the less space there is for its transmission. Dots appear on the screen. When only the tips of the teeth intersect the beam, dashes appear.

Yob has devised a simple apparatus for modulating the reflection of a laser beam with his hand or his voice. His laser is mounted vertically in a stand and set on a tripod. The stand has a top about six inches square. The laser beam passes through a central hole in the top. This arrangement also protects the operator from the vertical beam.

To intersect the beam Yob fastened a lightweight mirror to a piece of rubber stretched over a fruit-juice can. The rubber, which was made for balloons, and the mirror, which is coated with aluminum on its front surface, came from the Edmund Scientific Company (101 East Gloucester Pike, Barrington, N.J. 08007). The other end of the can is left open. To direct the reflected beam Yob either moves the can in the laser beam or speaks into the open end of the can to make the mirror vibrate. He can also put textured plastic on the top of the laser housing. Moving the plastic through the beam creates the images I discussed in my August article. Yob says good displays can be obtained from a stained glass known as clear German crackle.

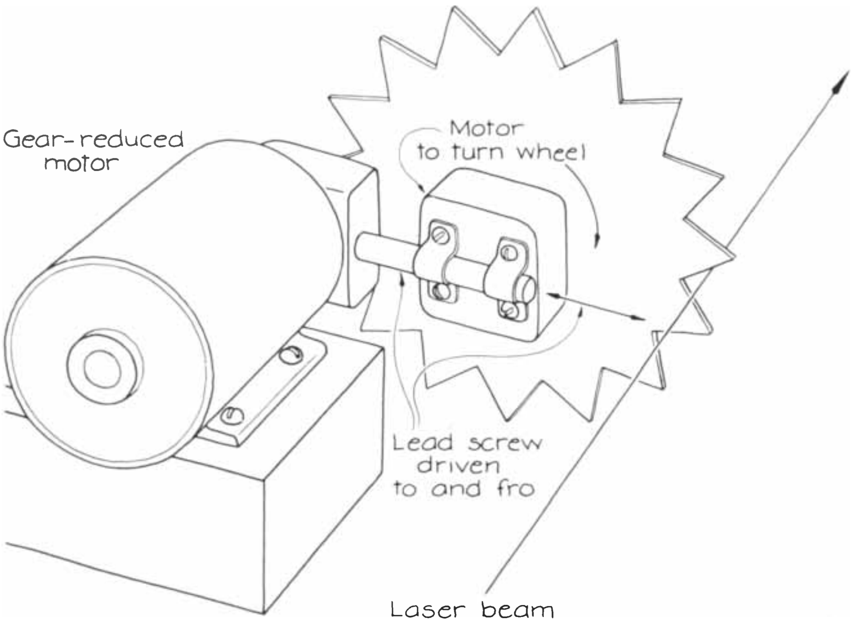
Abid Tanovic of Venice, Calif., wrote to me about his use of a dove prism to redirect the laser beam and form interesting patterns on a screen. A beam of light entering a slanted face of a dove prism parallel to the long side refracts toward that side, reflects from it and leaves the prism through the other slanted face, traveling in its original direction. At the long side the reflection is total, which means that because of the angle at which the beam hits that face and because of the index of refraction of the glass no light can refract out of the prism there.

If the prism is rotated about the axis of the incident light, the image is rotated twice as much. For example, if you look through the prism parallel to its long side while rotating it 180 degrees around your line of sight, the scene you see through the prism rotates 360 degrees. Tanovic incorporates this effect in his laser display by mounting a dove prism inside a tube he has fixed in a bearing block. The tube is rotated rapidly by means of a belt connected to a motor. He points out that the prism could be mounted directly in the center of the tube and that the motor should be variable in speed from 800 revolutions per minute down to zero in controlled steps.

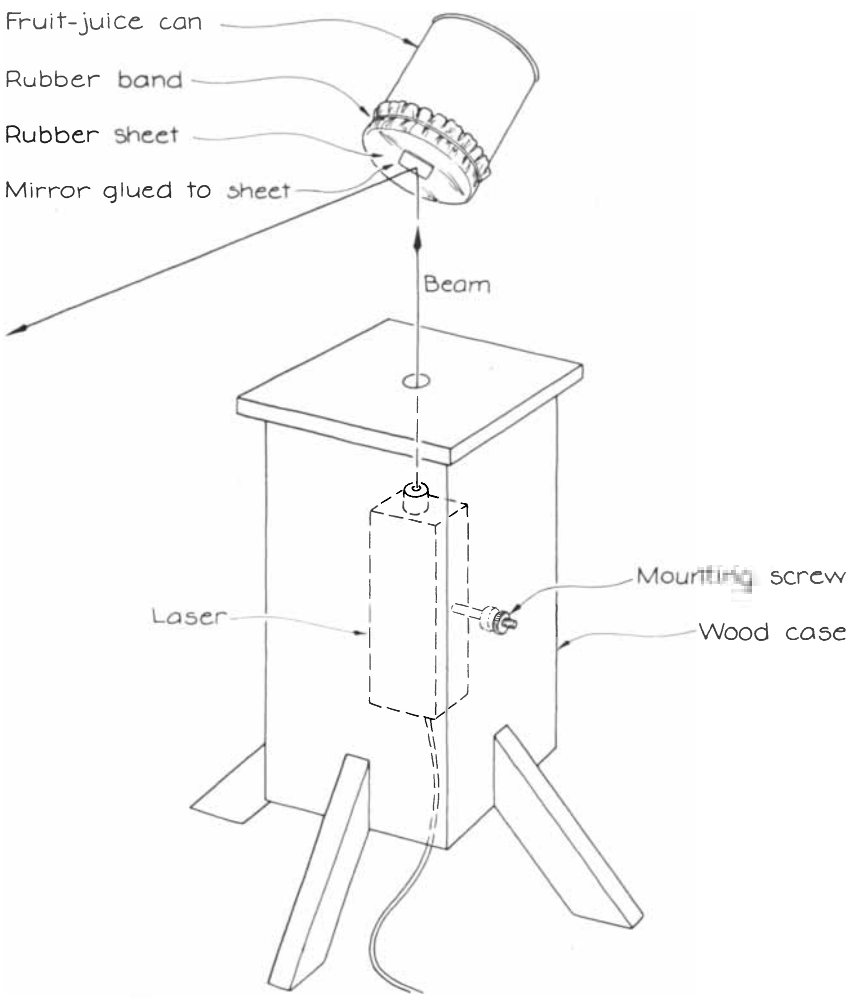
If a laser beam is sent through the rotating prism, it forms a circle on the screen. The size of the circle depends partly on how far from the center of rotation the laser beam entered the prism: the closer to the center the beam is on entering, the smaller the circle is. It also depends on the angle at which the beam enters the prism. To create more interesting formations on the screen Tanovic places near the prism diffraction gratings, Fresnel lenses, textured glass, shaped plastic or any of the other things I have described. The resulting images are fascinating as they form, rotate, spread and contract. With a pulsed laser or a chopped beam the images on the screen lie in a circle, and each successive image is oriented differently.

Tanovic lights his displays with a krypton laser emitting blue, green, yellow and red. He directs the beam through a regular prism in order to separate the colors. They fall at different places on a mirror he has mounted on a vibrating speaker. Then the light passes through his rotating dove prism and proceeds to the screen. The colors on the screen are in their spectral order from the center of the display to the perimeter. By rearranging the optical components he can make the central color either blue or red.

The techniques I have described are just a few of the possible ones for a laser-light show. To make laser beams visible oil dispersions blown into the air can be substituted for the dry-ice fog and the smoke from theatrical explosive. Beware of making a mess with the oil. I once ruined a screen with it. A laser



Yob's arrangement for chopping a beam



Yob's laser housing

beam will also give rise to interference patterns if it is directed through a viscous fluid that is being mixed by heating. As the light encounters varying refractive indexes in the fluid the beam interferes with itself and casts magnificent interference patterns on a screen.

Intriguing interference patterns develop when a laser beam is directed through a plastic material that is being slowly melted. Some of the commercial laser-light shows employ plastics that are melted by the laser beam itself, but this effect takes more power than an amateur's laser is likely to have. Finally, the entire laser show can be controlled by a home computer programmed to actuate the optical devices in time to the music from a tape recorder.

In December, 1979, I described a set of experiments involving sauce béarnaise, a warm emulsified mixture served over meats. The sauce consists primarily of dilute vinegar, wine, egg yolks and butter. In my experience it is a highly frustrating sauce to prepare because of its tendency to curdle, frequently just as I am serving it to a guest. In my discussion I raised three questions. What factors stabilize the sauce so that it remains smooth and appealing? What factors cause the droplets of butter in the sauce to flocculate, that is, to coalesce into unappetizing pools? If the sauce does fail in this way, what can the preparer do to salvage it? Colleen Kelly, Rachel Kleinman, Karen Mehlman and Craig Deutsche, who are students at the Westlake School for Girls in Los Angeles, wrote to me about experiments they did to answer these questions.

The answers I gave were not definitive because the present knowledge of the physics and chemistry of the sauce is sketchy. Two models have been formulated to describe the interaction of the butter droplets and the mechanism of a failure of the sauce. One model sees the droplets as being coated with negative electric charges. Surrounding the droplets is an "atmosphere" of positive charge. When two droplets of butter move close to each other, they repel each other because their atmospheres have charges of the same sign. With sufficient charge on the droplets and in their atmospheres they will not easily flocculate. The sauce is said to be stable.

In the other model the droplets are visualized as being coated with lecithin from the yolks of the eggs. Each molecule of lecithin is oriented with its lipophilic end pointing toward the butter droplet and its hydrophilic end pointing away from the droplet. The hydrophilic end binds water from the surrounding solution. As a result the butter droplets are covered with a protective layer of water molecules. Butter droplets that might flocculate are prevented from doing so by the water layer.

According to the first of the models,

the sauce fails if the charge on the butter droplets is too low. When droplets collide, they coalesce, eventually forming a pool of butter. In the other model failure results if the droplets have insufficient lecithin. Then they do not have enough protective water to prevent coalescence.

As an egg ages, the lecithin disintegrates, reducing the ratio of lecithin to cholesterol. Whereas lecithin is an emulsifier for butter droplets in water, cholesterol is an emulsifier for water droplets in butter. An aged egg may not provide the lecithin necessary to stabilize the sauce but may still provide the cholesterol that can ruin the sauce.

The students who told me about their work were concerned with the role of the egg yolks in the sauce. Mehlman made a successful sauce of vinegar, butter, lemon juice, cream and egg yolks. The egg yolks had been kept in the refrigerator, out of their shells, for four days. After preparing the sauce she added one and a half teaspoons of laboratory-grade cholesterol to a third of a cup of sauce. The mixture flocculated, but she could restore its smoothness by stirring it vigorously. The additional cholesterol apparently destroyed the emulsification of the sauce.

Kleinman made two batches of sauce from egg yolks, lemon juice and unsalted butter. In the first batch she used fresh eggs; the sauce was smooth and stable. For the second batch she aged the eggs for a week; the sauce flocculated. She then stirred in a tablespoon of liquid lecithin, which is available at health-food stores. The sauce was restored.

Half of this batch was left as an experimental control. To the other half Kleinman added two teaspoons of cholesterol. The sauce flocculated. These results appear to support the hypothesis that lecithin is a stabilizing agent and cholesterol is a destabilizing one. Still, the sauce is unpredictable. When Kleinman made a third batch and added several teaspoons of cholesterol, the sauce thickened but would not flocculate.

Kelly prepared a sauce with egg yolks, red-wine vinegar and margarine. When she added several teaspoons of cholesterol, the sauce flocculated. It was restored to smoothness when she stirred in liquid lecithin. In another batch of sauce she used eggs she had stored for six days. The sauce was not stable until she stirred in several teaspoons of lecithin.

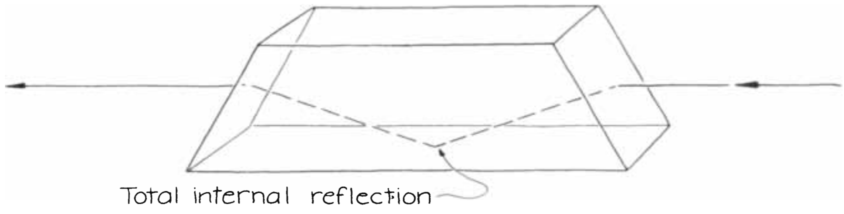
The students also tested their sauces for an effect of electric charge on the butter droplets. If the sauce is stabilized by charged atmospheres surrounding each droplet, the ions released by inorganic salts may deplete the charge of the atmospheres enough for the droplets to flocculate. In some cases salts releasing multivalent ions should promote flocculation more than salts releasing singly valent ions.

The students prepared samples of sauce to which they added variously sodium chloride, magnesium chloride, aluminum chloride, iron (III) chloride, sodium sulfate or sodium orthophosphate ($NaCl$, $MgCl_2$, $AlCl_3$, $FeCl_3$, Na_2SO_4 or Na_3PO_4). No consistent pattern emerged. The sauce showed no increased tendency to flocculate when multivalent ions were released by the salt. The students note that with different recipes different results may follow the addition of a salt. They conclude that the stabilization of a sauce béarnaise is more likely to be due to the lecithin than to charged atmospheres.

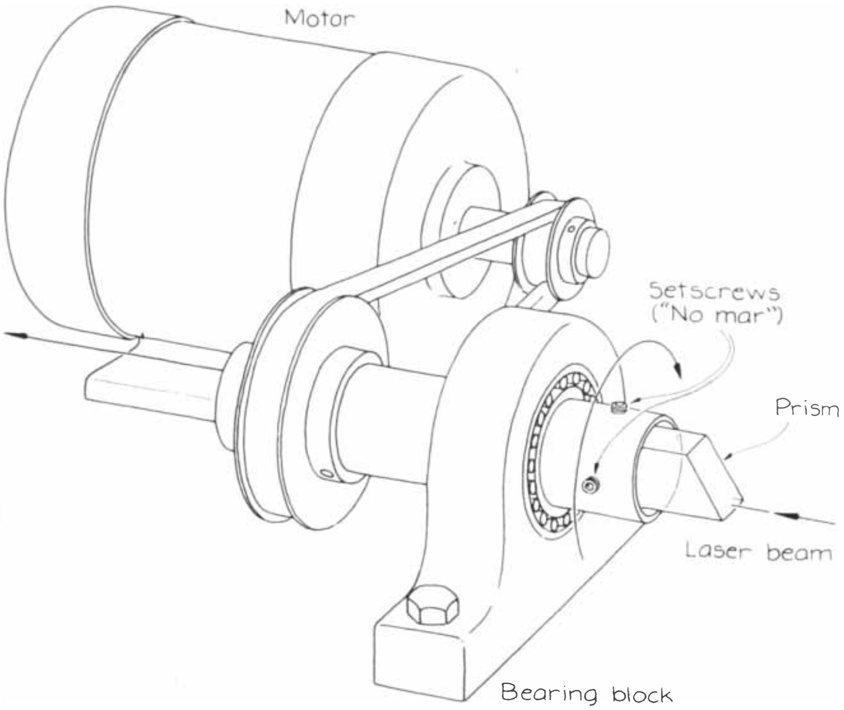
I also heard from Madeleine Kamman, a professional chef and cooking teacher who wrote *The Making of a Cook,* one of the few scientifically oriented cookbooks. She politely took me to task on my anatomy of a failed sauce. "The trouble is that you lovely scientists want to scientify everything. That sauce is fickle." I agree, but then again her own approach is hardly unscientific.

Kamman's secret for stabilizing a sauce béarnaise for as long as four hours, which is necessary in a restaurant, is to prepare an initial infusion that is more liquid than usual. The sauce is put into a pan that is put into another pan containing water at a temperature of 140 degrees Fahrenheit to keep the sauce warm. As liquid evaporates, the sauce thickens. The smart cook will know exactly when to remove it from the warming pan before the butter starts to separate into a layer. At that point a bit of salted water at a temperature of 120 degrees F. is whisked into the sauce. Its effect, Kamman says, is to reduce the acidity that builds up during evaporation. The addition of salted water is also partly for taste. Moreover, it and the whisking may be necessary to break up butter droplets that flocculate during the long period of sitting.

I think flocculation is less likely in the initial infusion because the extra water makes collisions of butter droplets less likely. As the water evaporates flocculation becomes more likely. The distribution of charge in the sauce may also change. Whisking and the addition of water are then needed to reverse flocculation and to make collisions of butter droplets less likely.



The effect of a dove prism on a laser beam



Abid Tanovic's design for mounting a dove prism

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

TOWARD A RATIONAL STRATEGY FOR OIL EXPLORATION

GIANT OIL AND GAS FIELDS IN THE UNITED STATES. Michel T. Halbouty in *The American Association of Petroleum Geologists Bulletin,* Vol. 52, No. 7, pages 1115–1151; July, 1968.

ENERGY RESOURCES. M. King Hubbert in *Resources and Man,* Committee on Resources and Man, National Academy of Sciences–National Research Council. W. H. Freeman and Company, 1969.

SCIENTIFIC USES OF RANDOM DRILLING MODELS. W. H. Menard and George Sharman in *Science,* Vol. 190, No. 4212, pages 337–343; October 24, 1975.

VIROIDS

BIOLOGY OF VIROIDS. T. O. Diener in *Slow Transmissible Diseases of the Nervous System,* edited by Stanley B. Prusiner and William J. Hadlow. Academic Press, 1979.

STRUCTURE AND FUNCTION OF VIROIDS. H. L. Sänger in *Slow Transmissible Diseases of the Nervous System,* edited by Stanley B. Prusiner and William J. Hadlow. Academic Press, 1979.

VIROIDS AND VIROID DISEASES. T. O. Diener. John Wiley & Sons, Inc., 1979.

VIROIDS: A CLASS OF SUBVIRAL PATHOGENS. H. J. Gross and D. Riesner in *Angewandte Chemie: International Edition in English,* Vol. 19, No. 4, pages 231–243; April, 1980.

THE TOTAL ARTIFICIAL HEART

AN ARTIFICIAL HEART INSIDE THE BODY. William J. Kolff in *Scientific American,* Vol. 213, No. 5, pages 38–46; November, 1965.

THE FIRST MECHANICAL HEART TRANSPLANT: INFORMED CONSENT AND EXPERIMENTATION. William J. Curran in *The New England Journal of Medicine,* Vol. 291, No. 19, pages 1015–1016; November 7, 1974.

CRITERIA FOR HUMAN TOTAL ARTIFICIAL HEART IMPLANTATION BASED ON STEADY STATE ANIMAL DATA. R. K. Jarvik, D. B. Olsen, T. R. Kessler, J. Lawson, J. English and W. J. Kolff in *Transactions of the American Society for Artificial Internal Organs,* Vol. 23, pages 535–542; 1977.

ASSISTED CIRCULATION. Edited by Felix Unger. Springer-Verlag, 1979.

TRANSIENT AND PERMANENT PROBLEMS ASSOCIATED WITH THE TOTAL ARTIFICIAL HEART IMPLANTATION. T. Murakami, K. Ozawa, H. Harasaki, G. Jacobs, R. Kiraly and Y. Nosé in *Transactions of the American Society for Artificial Internal Organs,* Vol. 25, pages 239–248; 1979.

THE ANDROMEDA GALAXY

SPIRAL STRUCTURE IN M31. Halton Arp in *The Astrophysical Journal,* Vol. 139, No. 4, pages 1045–1057; May 15, 1964.

GALAXIES. Harlow Shapley. Revised by Paul W. Hodge. Harvard University Press, 1972.

A RADIO CONTINUUM SURVEY OF M31 AT 2695 MHZ, I: OBSERVATIONS; COMPARISON OF RADIO CONTINUUM DATA. Elly M. Berkhuijsen and R. Wielebinski in *Astronomy and Astrophysics,* Vol. 34, No. 2, pages 173–179; August, 1974.

THE OPEN STAR CLUSTERS OF M31 AND ITS SPIRAL STRUCTURE. Paul W. Hodge in *The Astronomical Journal,* Vol. 84, No. 6, pages 744–751; June, 1979.

STUDIES OF LUMINOUS STARS IN NEARBY GALAXIES, IV: BAADE'S FIELD IV IN M31. Roberta M. Humphreys in *The Astrophysical Journal,* Vol. 234, No. 3, Part 1, pages 854–860; December 15, 1979.

THE WILD GENE RESOURCES OF WHEAT

HYBRIDIZATION BETWEEN AMPHIDIPLOIDS AND THE EVOLUTION OF POLYPLOIDS IN THE WHEAT (*AEGILOPS-TRITICUM*) GROUP. Daniel Zohary and Moshe Feldman in *Evolution,* Vol. 16, pages 44–61; 1962.

INTERSPECIFIC RELATIONSHIP IN *TRITICUM* AND *AEGILOPS*. Hitoshi Kihara in *Seiken Zihô* (*Report of the Kihara Institute for Biological Research*), No. 15, pages 1–12; 1963.

THE WHEATS AND THEIR RELATIVES. E. R. Sears in *Handbook of Genetics,* edited by Robert C. King. Plenum Press, 1975.

THE MIND-BODY PROBLEM

PSYCHOLOGICAL EXPLANATION: AN INTRODUCTION TO THE PHILOSOPHY OF PSYCHOLOGY. Jerry A. Fodor. Random House, Inc., 1968.

RES COGITANS: AN ESSAY IN RATIONAL PSYCHOLOGY. Zeno Vendler. Cornell University Press, 1972.

THE LANGUAGE OF THOUGHT. Jerry A. Fodor. Thomas Y. Crowell Co., 1975.

MIND, LANGUAGE AND REALITY: PHILOSOPHICAL PAPERS, VOL. 2. Hilary Putnam. Cambridge University Press, 1975.

READINGS IN PHILOSOPHY OF PSYCHOLOGY: VOL. 1. Edited by N. Block. Harvard University Press, 1980.

GELS

SWELLING OF NETWORK STRUCTURES: THEORY OF SWELLING. Paul J. Flory in *Principles of Polymer Chemistry.* Cornell University Press, 1953.

SPECTRUM OF LIGHT SCATTERED FROM A VISCOELASTIC GEL. Toyoichi Tanaka, Lon O. Hocker and George B. Benedek in *The Journal of Chemical Physics,* Vol. 59, No. 9, pages 5151–5159; November 1, 1973.

CRITICAL BEHAVIOR OF DENSITY FLUCTUATIONS IN GELS. Toyoichi Tanaka, Shin'ichi Ishiwata and Coe Ishimoto in *Physical Review Letters,* Vol. 38, No. 14, pages 771–774; April 4, 1977.

COLLAPSE OF GELS AND THE CRITICAL ENDPOINT. Toyoichi Tanaka in *Physical Review Letters,* Vol. 40, No. 12, pages 820–823; March 20, 1978.

PHASE TRANSITIONS IN IONIC GELS. Toyoichi Tanaka, David J. Fillmore, Izumi Nishio, Shao-Tang Sun, Gerald Swislow and Arati Shah in *Physical Review Letters,* Vol. 45, No. 20, pages 1636–1645; November 17, 1980.

TWO PATHS TO THE TELEPHONE

ALEXANDER GRAHAM BELL'S EXPERIMENTS WITH THE VARIABLE-RESISTANCE TRANSMITTER. Bernard S. Finn in *Smithsonian Journal of History,* Vol. 1, No. 4, pages 1–16; Winter, 1966.

BELL: ALEXANDER GRAHAM BELL AND THE CONQUEST OF SOLITUDE. Robert V. Bruce. Little, Brown and Company, 1973.

ELISHA GRAY AND THE TELEPHONE: ON THE DISADVANTAGES OF BEING AN EXPERT. David A. Hounshell in *Technology and Culture,* Vol. 16, No. 2, pages 133–161; April, 1975.

BELL AND GRAY: CONTRASTS IN STYLE, POLITICS, AND ETIQUETTE. David A. Hounshell in *Proceedings of the IEEE,* Vol. 64, No. 9, pages 1305–1314; September, 1976.

THE AMATEUR SCIENTIST

FUNDAMENTALS OF OPTICS. Francis A. Jenkins and Harvey E. White. McGraw-Hill Book Company, 1976.

A MATRIARCHAL SOCIETY: SAUCES. Madeleine Kamman in *The Making of a Cook.* Atheneum Press, 1978.

LASER ART & OPTICAL TRANSFORMS. T. Kallard. Optosonic Press, 1979.

LASERS. In *Light and Its Uses: Making and Using Lasers, Holograms, Interferometers, and Instruments of Dispersion.* With introductions by Jearl Walker. W. H. Freeman and Company, 1980.

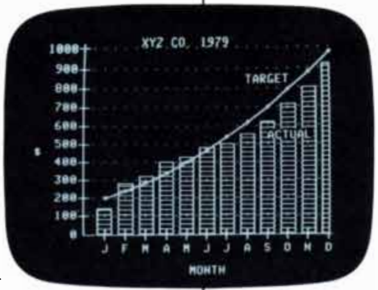
Three good reasons why professionals pick Apples.

1 In research

Apple personal computer systems help you collect, store and analyze data as fast as you can load a disk and execute a program. Because more than 100 companies offer software for Apple, you have the largest program library for manipulating your data in the personal computing world. Need special programs? Use any of Apple's development languages — BASIC, FORTRAN, Pascal.

2 In engineering

Apple personal computer systems let you define models, make trade-offs and refine prototypes. Want to study cause and effect of several variables? Apple computes new results instantly and displays them in colorful, easy-to-read graphs, charts or plots on a video monitor.



3 In production management

Apple personal computer systems make it easy to gather data, analyze productivity, measure yields and facilitate all phases of production control. Want to speed up repetitive tasks? Rely on Apple's word processing capabilities to write, edit and print your reports.

Apples grow with you.

Whichever system you pick, Apple never locks you into a single configuration. You can use up to four or eight I/O accessory expansion slots to add an IEEE bus, Apple's Silentype™ printer, a modem or a graphics tablet. Add memory up to 64K bytes or 128K bytes. Add up to four or six 5¼″ disk drives without adding any overhead.

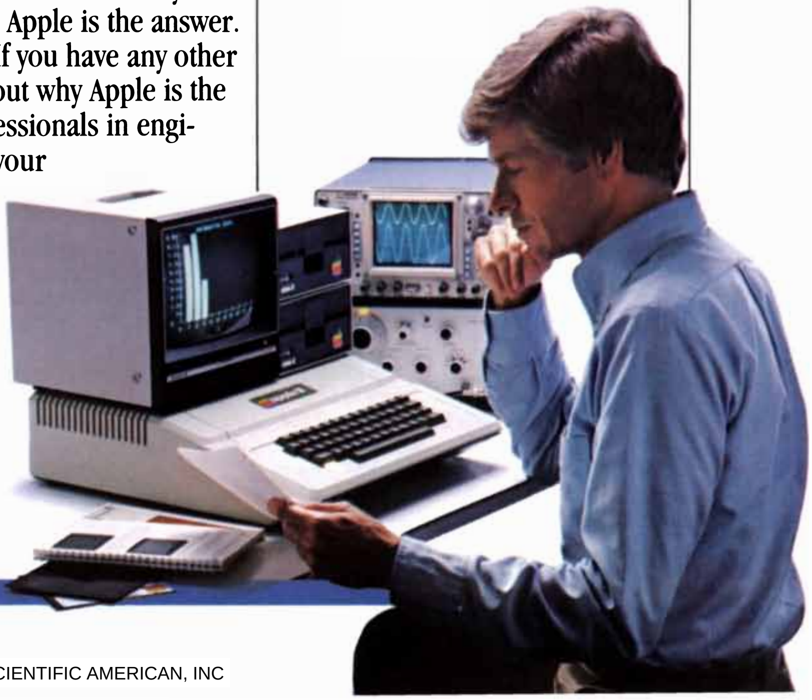
For support, service and the best extended warranty in the industry — Apple is the answer.

If you have any other questions about why Apple is the pick for professionals in engineering, see your nearest Apple computer dealer or call **800-538-9696**. In California, **800-662-9238**. Or write: Apple Computer, 10260 Bandley Drive, Cupertino, CA 95014.

	Apple II	Apple III
Maximum Memory Size	64K bytes	128K bytes
Screen Display	40 column (80 column with peripheral card) 24 Lines Upper Case	80 column 24 Lines Upper Case/Lower Case
Screen Resolution (B&W)	280 x 192	560 x 192
Screen Resolution (Color)	140 x 192 (6 colors)	280 x 192 (16 colors)
Keyboard	Fixed	Programmable
Numeric Key Pad	Accessory	Built-in
Input/Output	8 expansion slots	4 expansion slots plus built-in: disk interface RS-232 interface Silentype™ printer interface clock/calendar
Disk Drives	Add-on one to six drives	One drive built-in, plus interface to support three more drives
Languages	BASIC Fortran 77 Pascal Assembly Pilot	Enhanced BASIC Fortran 77 Pascal Assembly
Typical Configuration Pricing	CPU, 48K RAM, single disk drive, B&W Monitor (9″), Silentype™ printer, and BASIC. **$2875.00***	CPU, 96K RAM, integrated disk drive, B&W Monitor (12″), Silentype™ printer, SOS, Enhanced BASIC. **$4865.00***

*Suggested retail price.

apple computer inc.



How dare we call the new 1981 Saab 900 one of the most beautiful cars in the world, what with its wrap-around windshield, pointy front end, sloped rear end and just plain all-around strange looks?

Well, if you look at the Saab the way we do, you'll call it "beautiful" too.

For instance, the outside of the Saab is designed not to make your heart throb, but rather to slice through the wind.

Consequently, the Saab 900 has a low coefficient of drag, which affects performance, handling, and mileage all for the better.

At Saab, *that's* what we call "beautiful."

You may have noticed that Saabs ride a little high off the ground.

That's because we've found that adults like to sit upright in chairs, rather than sit on the floor.

So we provide the Saab with vertical seats that are up off the ground, to give you ample support for your thighs, back, shoulders, to make you as comfortable as is humanly possible. On some models, they're even heated!

And that unconventional windshield, combined with the height of the seat, gives you a phenomenal view of the road.

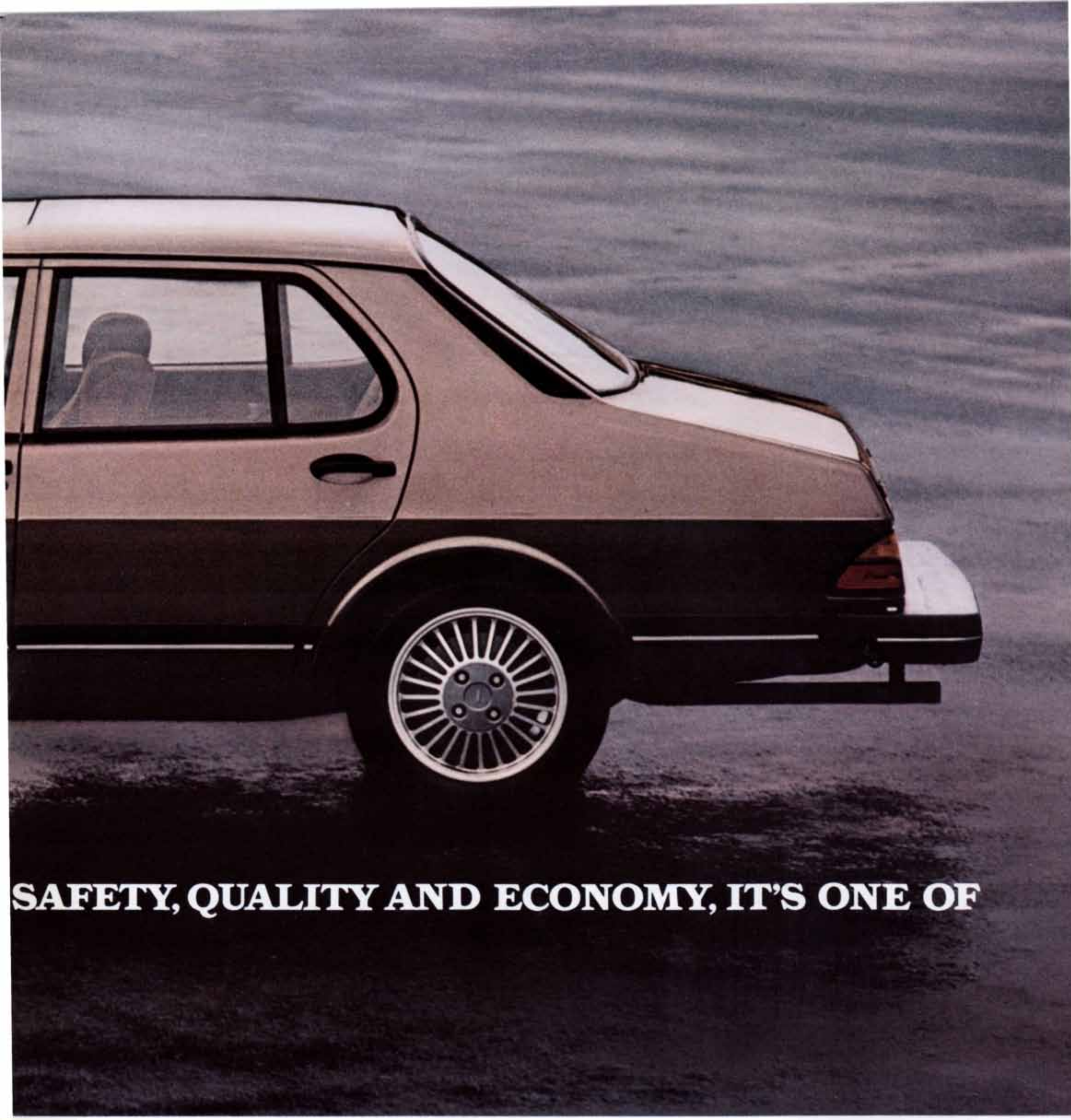
At Saab, *that's* what we call "beautiful."

Part of what the Saab looks like is a function of the safety features built into it.

For example, if you take away the sheet metal body, you'll see a cage of thick steel beams, pillars, cross members, floor plates, side panels, and so forth.

We call it the Saab "Safety Cage." We designed it to give you and your passengers extraordinary protection i the event of a collision.

Function, not form, was the primar objective in designing the basic shape of the Saab 900.


SAFETY, QUALITY AND ECONOMY, IT'S ONE OF

It's Technics SL-10 and it represents the most radical departure in turntable design since Technics first introduced the modern direct-drive turntable in 1969.

Not much bigger than a record jacket, the SL-10 combines a quartz-locked direct-drive motor, a servo-controlled linear-tracking tonearm and a moving-coil cartridge, complete with a built-in pre-preamp.

To play a record, simply place it on the platter, close the cover and push the start button. The SL-10's microcomputer automatically senses the record size and speed.

In addition to providing zero tracking error, the gimbal-suspended linear tonearm is dynamically balanced allowing you to play the SL-10 on its side or even upside down with no loss in accuracy or tracking ability.

Another reason for the SL-10's outstanding accuracy is its moving-coil cartridge. With its built-in pre-preamp, coreless twin-ring coils and pure boron pipe cantilever, the cartridge provides an extremely linear and flat frequency response as well as superb dynamic range.

Technics SL-10. The world's most unique turntable.

Technics
The science of sound