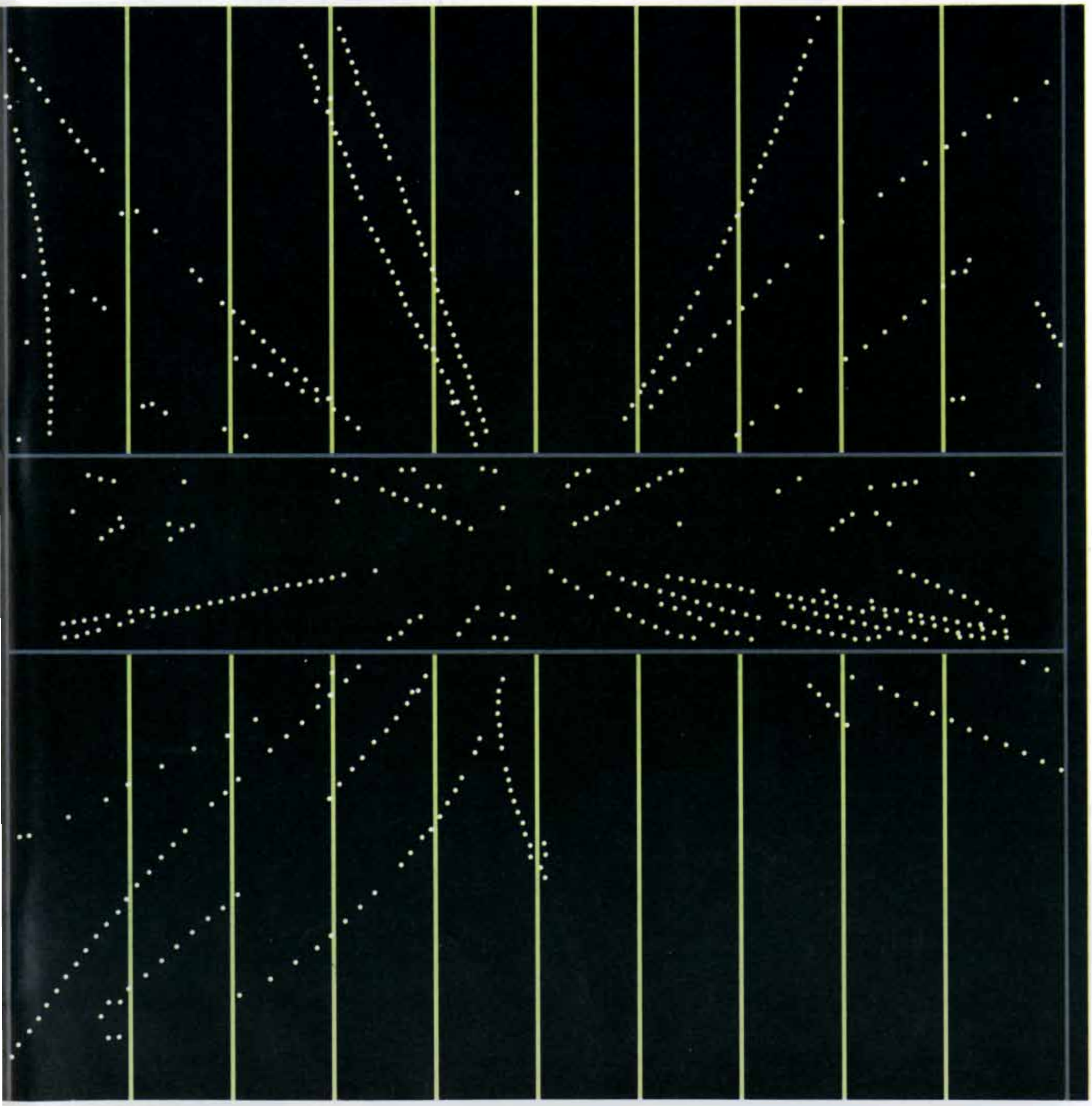
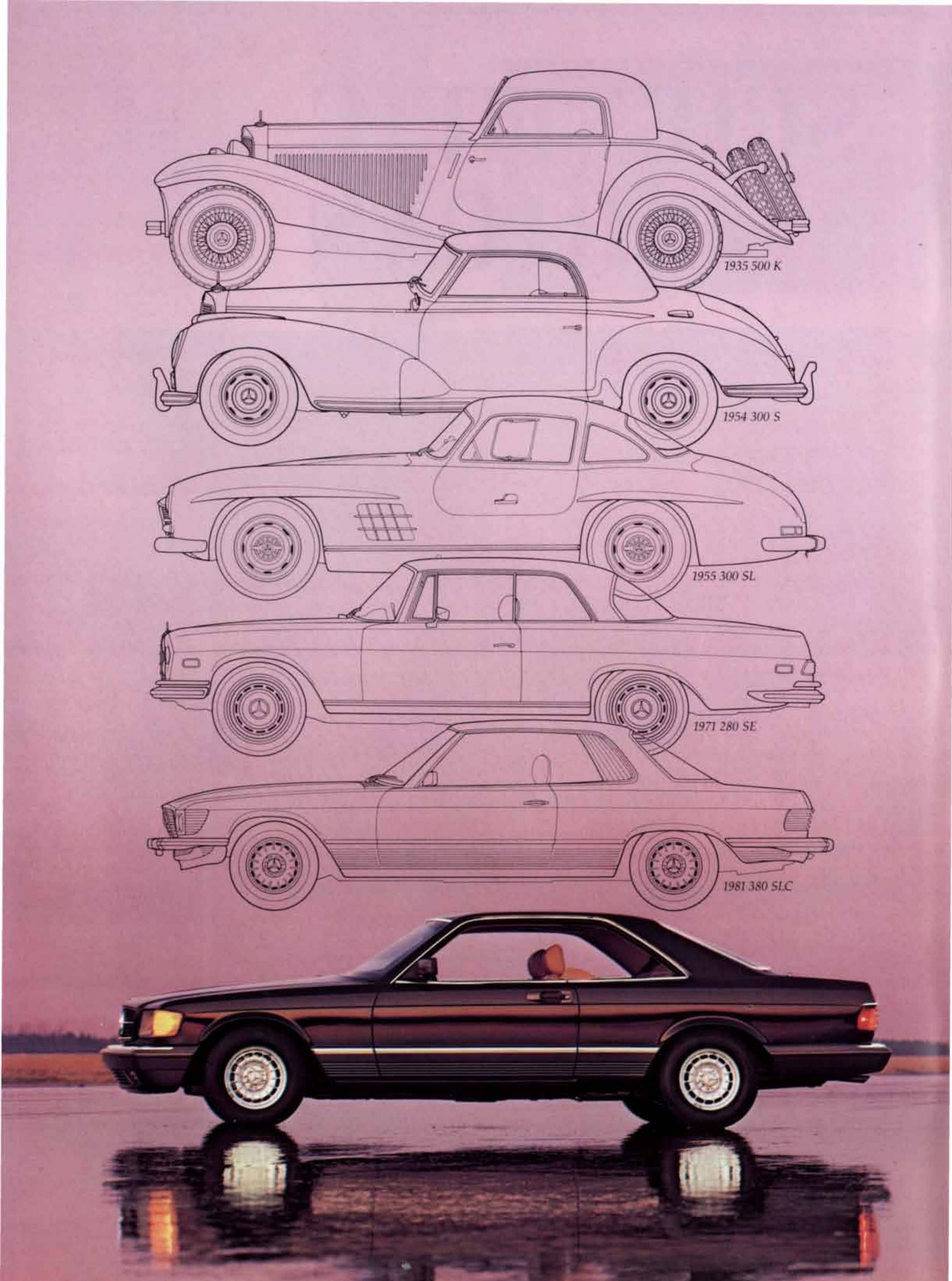
SCIENTIFIC AMERICAN



PROTON-ANTIPROTON COLLISION

$2.00

March 1982


1935 500 K
1954 300 S
1955 300 SL
1971 280 SE
1981 380 SLC

The evolution of the new Mercedes-Benz Coupe.

The new 380 SEC is the coupe that Mercedes-Benz has been readying itself to create for 83 years. There is no automobile like it, at $52,000* or any price.

A motorized carriage built by the co-founder of Mercedes-Benz in 1899 was the first automobile ever to be called a coupe.

This Benz "My Lord" founded a coupe lineage that must by now be the most aristocratic and most envied anywhere: evolving after 1927 from thinly disguised racing machines into roadsters and cabriolets as sumptuous as they were powerful; gaining fixed steel roofs; ever more civilized as driving itself became more civilized.

Until today when, whatever its character or type or age–if it is a Mercedes-Benz coupe, it is probably a collector's item.

There is now a new coupe, the 380 SEC. It is Mercedes-Benz at its technological zenith. List all those things a car of this kind should do well and point for point it is meant to score more highly, more often, than any other–even any other Mercedes-Benz coupe.

Its 0.35 drag coefficient makes this one of the most aerodynamically advanced designs in production today.

Inside, four persons are accommodated in four individual seats that feel sculpted in leather. They are served by virtually every electric assist and other amenity in the Mercedes-Benz repertoire.

Mechanical butler

Burled walnut veneer lends this cabin a warm and clubby atmosphere. So pampered is the driver that, as he turns on the ignition, he is served up his seat belt by a mechanical butler. "If you dare to ignore it," reports *Autocar*, "it continues to hold out the belt for half a minute, then–you *feel* the disapproval–it sadly retires again into its hole."

The 380 SEC, meanwhile handles and corners with the best sports cars. Its suspension is so supple, lending the tires such bite, that you can make good time even on atrocious roads.

Yet the car is so soothingly quiet that there is barely any wind noise at highway speeds.

The 380 SEC is engined with the most efficient V-8 in Mercedes-Benz history, a light alloy 3.8 liter whisperer. Your gearshifts are made for you hydraulically, by a highly automated four-speed automatic gearbox.

There is something notable about almost every aspect of the 380 SEC. This coupe provides a sedan-sized trunk of more than 14 cu. ft. capacity. The outside door handles wear aerodynamically shaped gauntlets. Every car has windshield washers, but the 380 SEC's washer nozzles are *heated*, as a deicing measure.

Quantities limited, significance universal

The 380 SEC will be minted in restricted numbers for a few fortunate customers per year. For the rest of the world it will stand as a landmark of automotive design–probably through this decade and beyond.

*Approximate suggested advertised delivered price at port of entry.

Engineered like no other car in the world


IMPORTED
Seagram's
V.O.
CANADIAN WHISKY
THIS WHISKY IS SIX YEARS OLD
868 PROOF
BOTTLED IN CANADA
Taste!
You be the judge. Discover why more V.O. is bought than any other imported distilled spirit in America. More than any Scotch, Canadian, Rum, Gin, Vodka, Tequila.
VO
Enjoy our quality in moderation.
CANADIAN WHISKY. A BLEND OF CANADA'S FINEST WHISKIES. 6 YEARS OLD. 86.8 PROOF. SEAGRAM DISTILLERS CO., N.Y.C.

Established 1845 **SCIENTIFIC AMERICAN** March 1982 Volume 246 Number 3

ARTICLES

DEPARTMENTS

SCIENTIFIC AMERICAN (ISSN 0036-8733), PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017 COPYRIGHT © 1982 BY SCIENTIFIC AMERICAN, INC. ALL RIGHTS RESERVED. PRINTED IN THE U.S.A. NO PART OF THIS ISSUE MAY BE REPRODUCED BY ANY MECHANICAL, PHOTOGRAPHIC OR ELECTRONIC PROCESS, OR IN THE FORM OF A PHONOGRAPHIC RECORDING, NOR MAY IT BE STORED IN A RETRIEVAL SYSTEM, TRANSMITTED OR OTHERWISE COPIED FOR PUBLIC OR PRIVATE USE WITHOUT WRITTEN PERMISSION OF THE PUBLISHER. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $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, NY 10017

COULD THE INDIVIDUAL HOLD THE KEY TO AN ENTIRE COMPANY'S PRODUCTIVITY?

THE COVER

The design on the cover shows an array of elementary-particle tracks emanating from the head-on collision of a proton and an antiproton, each of which has been accelerated to high energy in a colliding-beam machine recently completed at the European Organization for Nuclear Research (CERN) in Geneva. It is in the debris of such collisions that physicists expect to see the first direct evidence of a long-sought family of particles called intermediate vector bosons (see "The Search for Intermediate Vector Bosons," by David B. Cline, Carlo Rubbia and Simon van der Meer, page 48). The cover design is based on a new computer-assisted display associated with a large particle detector surrounding the collision point. The two opposed beams are confined to an evacuated tube (not shown) that runs along the axis of a cylindrical "drift chamber" at the center of the detector. Protons enter from the right, antiprotons from the left. The vertical yellow lines correspond to planes of closely spaced wires strung across the chamber at right angles to the beam tube (perpendicular to the picture plane in this side view). The electrically charged decay products from the collision ionize molecules of a low-pressure gas in the chamber; the ions drift to the wires, where they deposit their charge. Signals from the wires are processed by computer to reconstruct the trajectories of decay products.

THE ILLUSTRATIONS

Cover illustration by Gabor Kiss

Page	Source
18–28	Albert E. Miller
42	Wide World Photos
43–47	Albert E. Miller
49	European Organization for Nuclear Research
50–57	Gabor Kiss
58	Gabor Kiss (*top*), European Organization for Nuclear Research (*bottom*)
59	European Organization for Nuclear Research
60–61	Robert H. Evans, Otis B. Brown and James W. Brown
62–68	George V. Kelvin
69	Tom Prentiss
70	George V. Kelvin
80	Bunji Tagawa
82	G. Steven Martin (*top*), Bunji Tagawa (*bottom*)
83–84	Bunji Tagawa
87	Mark C. Willingham and Ira H. Pastan (*top*), Larry R. Rohrschneider (*bottom*)
88	Bunji Tagawa
89	J. Michael Bishop (*top*), Bunji Tagawa (*middle and bottom*)
90–92	Bunji Tagawa
95	Fiat Auto S.p.A.
96	Ian Worpole
99	Avco Everett Research Laboratory, Inc.
100–101	R.T.M. Institute
102	United Technologies Research Center
103	General Motors Corporation
107–108	Laird A. Thompson and Kitt Peak National Observatory
110–114	Allen Beechel
117	Eric A. Newman
118–119	Tom Prentiss
120	Tom Prentiss (*top and middle*), Edward R. Gruberg and Eric A. Newman (*bottom*)
121–127	Tom Prentiss
129	R. F. Bonifield, courtesy of the Museum of the American Indian, New York City
130–131	Ilil Arbel
132	Robert Byck, Yale University School of Medicine, and Alan Pooley, Peabody Museum, Yale University
133–140	Ilil Arbel
147–152	Michael Goodman

LETTERS

Sirs:

This letter is one of the most difficult I have ever written. It is a letter of protest, and I intend it to be a strong one. It is against the policies of a magazine that I greatly admire, and it is particularly difficult because it was necessitated by what I gather is the final column of one of my heroes, a man of reason whom I have been reading for 20 years, Martin Gardner. Nevertheless, it must be done.

For as long as I can remember the attitude of *Scientific American* toward economics has been one of disdain for any theory other than the welfare-statist/interventionist school. Usually, of course, opposing schools are not even mentioned, and so I suppose one should be grateful for their limited inclusion in the December "Mathematical Games." But the sneering attitude, the almost total contempt for anyone who dares to hold for some semblance of a free market, is too much. Perhaps your readers are to blame. To paraphrase Ayn Rand, if in the past any intellectual submitted an article advocating the controlling of world production, feeble voices might be raised against his (or her) statistics, but no one would question his right to enforce his way with a gun. I have never spoken up before, but I do now. For the record, I disagree totally with any individual who is against free markets and the free minds that make them possible....

JOHN L. QUEL

Kettering, Ohio

Sirs:

In the December issue Martin Gardner dedicated "Mathematical Games" to his views of the economic theories put forward by the present Administration. In making his comments Mr. Gardner resorted to the use of innuendo, put-downs and the display of "purported fact." With the last of these I refer to his statement that his so-called neo-Laffer curve is "based ... on a sophisticated statistical analysis ... of the best available data for the U.S. economy over the past 50 years" and that "since the data are represented ... by a swarm of densely packed points, the actual shape of the curve is somewhat arbitrary." He then drew a curve begging data not presented, contending that it is representative of these data. That this curve is worth less than nothing is immediately obvious, and the conclusions drawn from it are meaningless as well.

Within the pages of *Scientific American* I would have expected possibly to find an article or a series of articles presenting various economic theories, their pros and cons and the data either supporting or refuting them. I never expected to see such a sham as Mr. Gardner's article. It is particularly distressing to me that you, the editors, would see fit to print this article in a section supposedly devoted to mathematical games and diversions, and not labeled as an editorial. Such a presentation is a blatant disservice to the readers of *Scientific American*.

No matter what economic views you may espouse, if they are not presented in such a manner as to be open to the scrutiny of scientific examination and inquiry, they have no place in *Scientific American*. I can only hope that such a disgrace as this will not be repeated and that Mr. Gardner and the editors will publicly apologize to their readers.

DIRK DUNNING

Gresham, Ore.

Sirs:

I too believe both in free markets and in free minds. How Mr. Quel could take my spoof of Lafferism to be an attack on either beats me. Mr. Dunning's letter is funnier than my column. Did he really think my neo-Laffer curve was based on a "sophisticated statistical analysis"? Yes, Mr. Dunning, the curve is a total sham. That was the whole point of it.

MARTIN GARDNER

Hendersonville, S.C.

Sirs:

It was a pleasure to read Martin Gardner's column about the Laffer curve in the December *Scientific American*. Lest it be said that there are no interior laughs in the dismal science, I should like to report briefly on my own research in the area of economic-policy implications of the Laugher curve.

What Laffer asserts and almost everyone believes is that the tax-revenue function, with the tax rate as an argument, has a maximum value. What only fools believe, or cynical opportunists pretend to believe, is that this maximum has been attained by the U.S. economy in the 1970's. At any rate no proof has been requested and none has been offered. This is of course an application of the free-market mechanism to the fields of science and ethics. Briefly, if an idea is still being believed, then it must be a good one since the competition of ideas in the marketplace drives bad products out of business. Any form of standard is intolerable because the standard calls for the intervention of regulators.

When I first realized that the main problem of the U.S. economy is inflation caused by low productivity and high spending, I came up with an institutional solution: reintroduce slavery. In this way one can guarantee low wages and high productivity and thus increase output and cut spending. I don't talk that way any more; I was being taken seriously in some quarters.

I therefore felt a theoretical solution was called for. If one derives the supply of labor by maximizing the worker's utility subject to a survival constraint (namely that people have to eat), then the supply of labor varies directly with real wages up to the point where the survival constraint becomes binding. Reduce wages below that and people must work harder to make ends meet. This can be done until the worker expires near the zero wage level. This is the Laugher curve, and it rather than the supply-side scenarios should be the basis of rational economic policy. Indeed, since the survival constraint is a rectangular hyperbola, the wage bill is fixed and output increases. Hence profits also increase and, as everyone knows, profits are gainfully employed. Hence output increases and everyone lives happily ever after, except for those who died of exhaustion or violated the survival constraint.

MOHAMED A. EL-HODIRI

Professor of Economics
University of Kansas
Lawrence

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 75 cents.

Subscription correspondence should be addressed to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 5969, 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 ______________________

____

____

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

MARCH, 1932: "The rocket turbine for airplanes, recently patented by Robert H. Goddard, Director of the Physical Laboratory at Clark University, is designed to solve a major problem in the use of rocket propulsion for craft capable of carrying human beings. The object of this rocket plane is to provide highly efficient transportation at ordinary airplane speeds and at the same time to make possible travel at great altitudes through the stratosphere, where propellers and ordinary airplane engines are useless. Dr. Goddard's rocket plane has two propellers at the rear, each being surrounded by a streamlined housing containing turbine blades. When the plane is in the dense part of the atmosphere, the rocket blast from the rear of the plane impinges on the turbine blades and the energy of the blast is transmitted to the propellers in such a way as to make them revolve. When the plane is flying at a very high altitude, the propellers and housings are moved to one side out of the way of the gas stream. Then the rocket blast alone furnishes propulsion."

"Two methods of preventing diphtheria deaths now exist: one is early treatment of the disease with antitoxin, the other is prevention by immunization with toxoid. 'The prevention of diphtheria by active immunization is the greatest thing in public health since Jenner's development of vaccination against smallpox,' says Dr. Walter T. Harrison of the National Institute of Health. He predicts that diphtheria, once the dreaded scourge of childhood, will disappear in a few more years. Credit for the development of immunization against diphtheria goes to Dr. William H. Park, director of laboratories of the New York City Health Department, although toxoid was first introduced by Prof. Gaston Ramon of the Pasteur Institute in Paris. Toxoid is toxin to which a small amount of the common disinfectant formalin has been added, and which is then kept at a temperature of 100 to 102 degrees Fahrenheit for from three to six weeks."

"The Curtiss Condor transport plane has recently been put into service on Eastern Air Transport lines. This excellent airplane illustrates the tendencies of the best modern design. Practice in transport planes seems to be swinging to compartments. Perhaps the reason is that air-sick passengers ought to have a compartment to themselves! At any rate, the Condor's comfortable, deeply upholstered chairs are arranged in groups of six in three separate compartments, each partition having a sliding door and two large non-shatterable glass windows. Each chair is equipped with fixtures for individual luncheon trays and in addition large tables are provided for card playing or writing. The cabin ventilation and heating have been carefully studied, and the ventilation is perfectly adequate with all windows shut. It is considered by the best authorities that decrease of noise and vibration and the provision of perfect heating and ventilating are the best possible safeguards against air-sickness."

SCIENTIFIC AMERICAN

MARCH, 1882: "There are in chemistry two leading methods of research that have contributed equally to the advancement of our knowledge of the relations of the various elements to one another. One of them is analysis, or decomposition; the other is synthesis, or combination. By various methods—some extremely simple, others highly complex—all substances, animal, vegetable and mineral, within our reach have been analyzed, or decomposed, until we have arrived at a number of bodies that we call elements—about 65 in number—and that we have never yet been able to tear asunder into any simpler form of matter. The forces by which bodies may be dissociated are various, but by far the most important is heat. By the action of heat all compound bodies are resolvable into their elementary constituents, and if we had a world the temperature of which was sufficiently high, we should have simply a collection of elementary bodies."

"The Common Council of Cincinnati, Ohio, has taken hold of the smoke problem in a vigorous way by passing an ordinance making it an offense punishable by fine to maintain a furnace that needlessly pollutes the air with smoke. It is unreasonable to suppose that the problem is incapable of solution. The products of the perfect combustion of the smokiest coal are solid ashes, which remain in the furnace, and colorless gases, which make no visible addition to the atmosphere. The presence of smoke is always proof of imperfect and wasteful burning. It is the business of our inventors to accomplish as perfect a combustion of the fuel used as is possible in the laboratory, either primarily in the furnace or by the subsequent reburning and washing of the sooty and volatile products that so largely pollute the air of our cities and manufacturing towns. The demand for such inventions is wide and urgent. The action of the Common Council of Cincinnati is likely to be generally imitated."

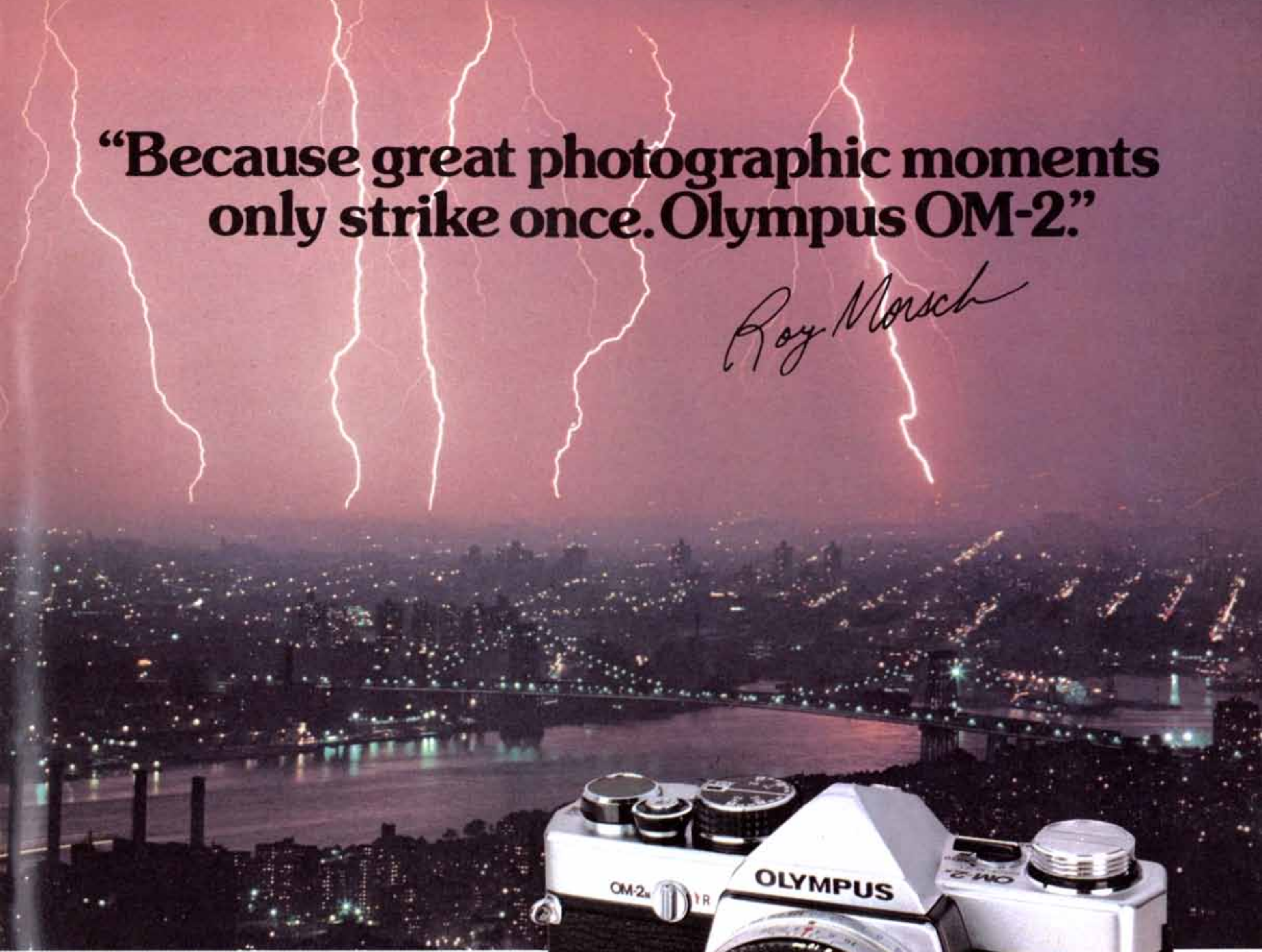
"The distance between England and the United States is gradually being annihilated, and the Atlantic passage begins to lose some of its terrors when little more than seven days need be spent in making it. The recently built Cunard steamer *Servia* reached Queenstown last week after having made what is claimed to be the shortest trip from New York on record. The actual time occupied was seven days, seven hours and 41 minutes, a remarkable performance taking into account that the route followed by the Cunard liners is 90 miles longer than that adopted by other steamers."

"M. Blennard has recently completed his investigations upon the composition of albuminoid substances at the Collège de France. Albuminoid substances exposed to the influence of barium hydrate are decomposed and yield products that vary according to the nature of the albuminoid substances treated. A deer's horn, for example, treated with barium hydrate becomes hydrated and produces ammonia, carbonic, oxalic and acetic acids, and finally a mixture of amido acids. Among these last it is important to recognize the bodies designated as gluco-proteines. M. Blennard has studied by the same method the action of barium hydrate on isinglass, osseine and gelatine, which decompose in a similar manner. Wool, feathers and human hair give results that are comparable. They differ in the quantity of alanine, amidobutyric acid, etc. The fur of the rabbit acts very much like the fibroine of silk in its decomposition."

"The anti-Mongolian mania threatens to be general. Already it has raged in Congress to a degree calculated to humiliate all sane Americans now and for years to come. The picture that Congressmen draw of the certain submergence of Christian civilization in this country by swarming hordes of heathen Chinese is so appalling that Congress threatens to pass a law to prevent it by stopping immigration from that side of the world, thereby adopting toward the Chinese the same policy of exclusion that the Chinese so long exercised against the 'outside barbarians.' Having compelled China to open her ports and allow Americans to go thither to trade and to upset by missionary operations the social and religious order of the empire, the superior race now finds itself in the position of the fisherman in *The Arabian Nights.* After he had forced the cork of the magically sealed bottle he would fain have the genie shut itself up again. The case would be pitiful if it did not originate in craze and lead to national dishonor."

"Because great photographic moments only strike once. Olympus OM-2."

Roy Morsch



"My OM-2 worked perfectly where most automatic SLR's couldn't work at all," says award-winning commercial and *NY Daily News* professional photographer Roy Morsch. "It automatically made a perfect picture from 5 successive lightning bolts, without guesswork. My OM-2 never misses a shot!"

What makes OM-2 perfect for professionals like Roy Morsch makes it unbeatable for you. Its revolutionary OTF™ system sets exposures *as* you shoot, not before. Off-the-film. Automatically fine-tuning changing light in micro seconds. As no ordinary SLR can do.

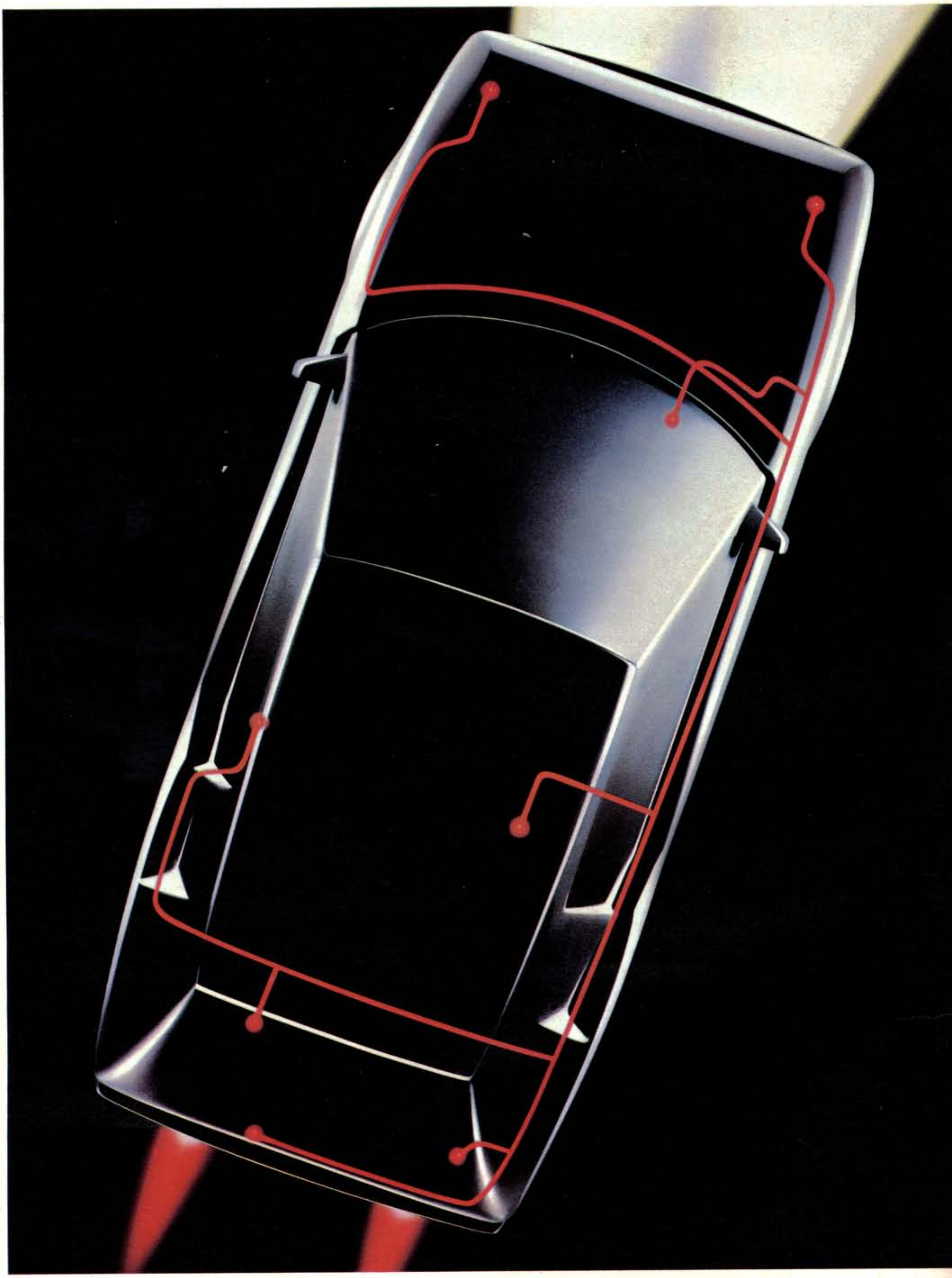
OM-2 even makes pro-quality flash shots automatically easy. With its OTF™-controlled Olympus thyristor flash units. So you command more creative options than ever before. The OM-2's LED viewfinder ready light even confirms exposure accuracy. Automatically.

And OM-2 is more than an incomparable camera. It's the heart of the incomparable OM system. With 300 superb lenses and accessories to let you do almost anything, anytime. See OM-2 now at your Olympus dealer. You'll see why it's taking the pros by storm...

For information, write Olympus, Woodbury, NY 11797. In Canada: W. Carsen Co. Ltd., Toronto.

OLYMPUS® OM-2

The camera with the press card



Running Light through a Running Car

Computerization came first; now wiring

Great strides have been made in microcomputerized automotive engine systems to improve fuel economy and reduce exhaust emissions. The microcomputer controls fuel injection, ignition timing, EGR rate and engine speed under idling conditions. Similar systems have been developed to adjust the headlight beams as a car takes curves, prevent skidding when braking on slippery roads, and detect malfunctions. But there is a significant problem: Today's passenger car already contains 1–2 kilometers of wire weighing up to 10 kilograms, most of which must be installed by hand. And these innovations require even more bulky and complex wiring systems.

Hitachi takes advantage of opto-electronics

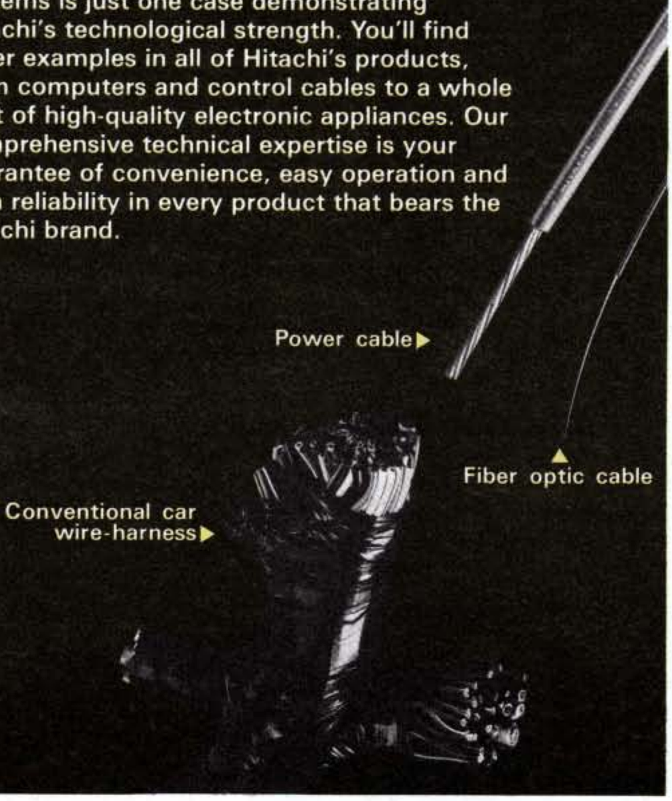
Hitachi's approach to the problem was to modify the optical electronics process whereby laser light emitters are used to transmit millions of digitalized signals per second through optical glass fiber to diode light receivers. Our researchers came up with light-emitting diodes (LEDs) capable of sending infrared signals through plastic fiber with a core diameter of about 1 mm. An entire automobile wiring system incorporating this fiber will take up less space than the conventional network and weigh 1–2 kilograms. And unlike copper wire, plastic optical fiber is not susceptible to electrical noise which might disrupt the weak signals used for microcomputer control.

The wiring system of the future

The optical fiber multiplexing system will interlink a central control unit (CCU), comprising an LED module and 8-bit microcomputer, with local control units (LCUs) that convert CCU optical transmissions into electrical signals in order to carry out the driver's instructions. What's more, two-way, time-shared multiplex transmission of pulse signals means that information can be sent back to the CCU through the same connecting fiber: if a taillight burns out, for example, the related LCU can trigger an alarm on the display panel in front of the driver. Eventually, multiplexing will be expanded to allow electronic navigation, danger avoidance, and even fully automatic driving.

Technical excellence is embodied in all Hitachi products

The development of optical fiber wiring systems is just one case demonstrating Hitachi's technological strength. You'll find other examples in all of Hitachi's products, from computers and control cables to a whole host of high-quality electronic appliances. Our comprehensive technical expertise is your guarantee of convenience, easy operation and high reliability in every product that bears the Hitachi brand.



HITACHI

A World Leader in Technology

THE AUTHORS

CHARLES B. KEELY ("Illegal Migration") is research associate at the Center for Policy Studies of the Population Council in New York. He received both his bachelor's degree (1965) and his Ph.D. in sociology (1970) from Fordham University. In 1969 and 1970 he was assistant professor of sociology at Loyola College in Baltimore. In 1971 he joined the faculty of Western Michigan University, leaving in 1975 to return to Fordham as associate professor. He moved to the Population Council in 1977. His main research interest is immigration. He has studied the effects of immigration on population growth and employment in the U.S. and also the flow of temporary workers from the Middle East to Europe.

DAVID B. CLINE, CARLO RUBBIA and SIMON VAN DER MEER ("The Search for Intermediate Vector Bosons") are all working in the effort to detect the carrier particles of the weak nuclear force. Cline divides his time between the University of Wisconsin at Madison, where he is professor of physics, and the Fermi National Accelerator Laboratory (Fermilab). His undergraduate degree is from Kansas State University. His Ph.D. in physics (1965) is from the University of Wisconsin. Rubbia, born in Italy, was trained in physics at the Normal School of Pisa and at the University of Pisa, where he was instructor of physics in 1958 and 1959. He came to the U.S. in 1959 as a research fellow at Columbia University and returned to Italy in 1960 to join the faculty of the University of Rome. From 1962 to 1970 he was senior physicist at the European Organization for Nuclear Research (CERN) in Geneva. Rubbia joined the faculty of Harvard University as professor of physics in 1970 and now divides his time between Harvard and CERN. Van der Meer received his diploma in engineering from the University of Guelph in 1956 and since then has been on the staff of CERN. His main work has been on the design of accelerators, and lately he has had a central part in the development of proton-antiproton colliding-beam machines capable of creating intermediate vector bosons.

PETER H. WIEBE ("Rings of the Gulf Stream") is associate scientist at the Woods Hole Oceanographic Institution. He received his B.S. in 1962 from Arizona State College; his Ph.D. in biological oceanography (1968) is from the University of California at San Diego. In 1968 and 1969 he was a research fellow at Stanford University's Hopkins Marine Station. At the end of that time he moved to the Woods Hole Oceanographic Institution as assistant scientist. His main research interest is the factors that affect the size and distribution of populations of ocean animals.

J. MICHAEL BISHOP ("Oncogenes") is professor of microbiology at the University of California at San Francisco. He was graduated from Gettysburg College in 1957 with an A.B. in chemistry. He got his M.D. from the Harvard Medical School in 1962. Between 1964 and 1968 he worked at the National Institute of Allergy and Infectious Diseases in the field of virology, first as a research associate and then as a senior investigator. He joined the faculty of the University of California in 1968, becoming a full professor in 1972. Bishop's main research interest is the theme of his article: how normal mammalian genes are incorporated into viruses and how such genes may initiate tumors.

ALDO V. LA ROCCA ("Laser Applications in Manufacturing") is on the staff of Fiat Auto S.p.A. in Turin. His responsibility is the application of advanced technology to the company's manufacturing processes. He received a degree in mechanical engineering from the University of Naples in 1950. In 1951 he came to the U.S. as a Fulbright fellow to study at the Polytechnic Institute of Brooklyn. From 1951 to 1955 he studied gas dynamics in the institute's Supersonic Aerodynamics Laboratory, obtaining a Ph.D. in applied mechanics. In 1955 he joined the General Electric Company, where his research concerned the propulsion of aerospace vehicles. He returned to Italy in 1972 to join the staff of Fiat; there much of his work has been on the industrial applications of lasers.

STEPHEN A. GREGORY and LAIRD A. THOMPSON ("Superclusters and Voids in the Distribution of Galaxies") are astronomers with a particular interest in the structure and evolution of galaxy superclusters. Gregory's bachelor's degree is from the University of Illinois at Urbana-Champaign; his doctorate in astronomy is from the University of Arizona. From 1973 to 1977 he was on the faculty of the State University of New York at Oswego. In 1977 he went to Bowling Green State University as assistant professor of physics. Thompson received his undergraduate training in astronomy and physics at the University of California at Los Angeles. His Ph.D. in astronomy was awarded by the University of Arizona in 1974, the same year as Gregory's. After two years at the University of Nebraska he moved to the University of Hawaii at Manoa, where he is current-

YOU RARELY SEE USED SAABS FOR SALE BECAUSE THERE'S RARELY A REASON TO SELL THEM.

The longer Saab owners hold on to their cars, the more they like them.

But that's no surprise when you consider what there is about a Saab to like.

For superb performance, Saabs are equipped with a fuel-injected engine that combines excellent drivability with the economy of 4 cylinders.

When it comes to cargo space, Saabs have as much room as a station wagon—up to 56.5 cubic feet.

And as far as styling goes, Saabs have no sharp angular corners, so there's no cause for air turbulence. Even its rear end is shaped to give it efficient air breakaway.

So if you find yourself interested in a Saab, come test drive a new one.

You might not be able to get your hands on a used one.

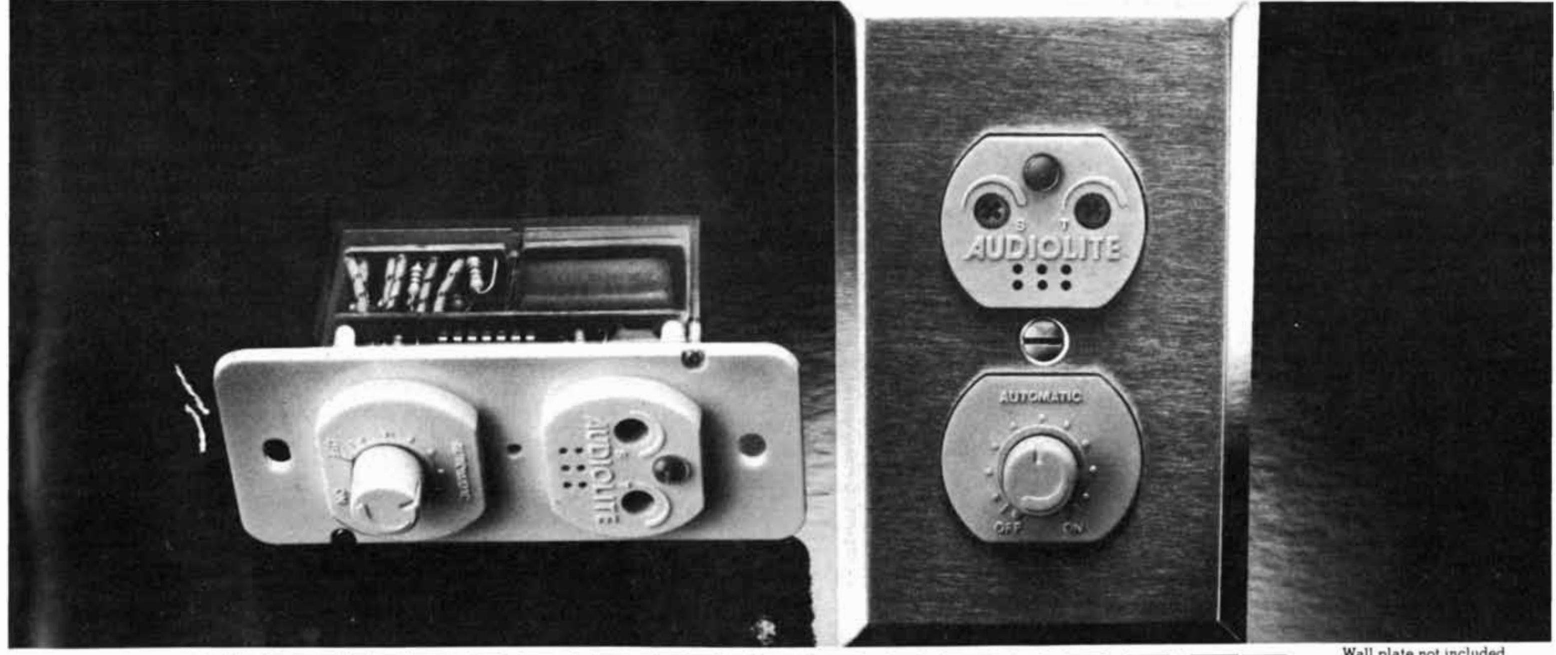
SAAB

The most intelligent car ever built.

ly a member of the staff of the Institute for Astronomy. His research interests include (in addition to superclusters) supernovas, high-resolution imaging and the color photography of celestial phenomena.

ERIC A. NEWMAN and PETER H. HARTLINE ("The Infrared 'Vision' of Snakes") are respectively assistant scientist and associate scientist at the Eye Research Institute of the Retina Foundation in Boston. They have collaborated in work on the infrared sense of snakes since 1978. Newman obtained his bachelor's, master's and doctoral degrees from the Massachusetts Institute of Technology; his Ph.D., granted in 1977, is in biology. Hartline received his undergraduate education at Swarthmore College. He earned a master's degree at Harvard University and, in 1969, a Ph.D. in neuroscience from the University of California at San Diego. From 1969 to 1974 he was assistant research scientist at the University of California at San Diego. In 1974 he went to the University of Illinois at Urbana-Champaign, and in 1977 he moved to the Eye Research Institute.

CRAIG VAN DYKE and ROBERT BYCK ("Cocaine") are psychiatrists who share an interest in drugs and drug abuse. Van Dyke is associate professor of psychiatry at the University of California School of Medicine in San Francisco and Chief of Consultation–Liaison Psychiatry at the San Francisco Veterans Administration Medical Center. He was graduated from the University of Washington in 1963 with a bachelor's degree in oceanography and received his M.D. from the same institution in 1969. After completing an internship in medicine he spent two and a half years in the Navy. In 1971 he moved to the Yale University School of Medicine, where he was a psychiatric resident and then a member of the faculty. Van Dyke joined the faculty of the University of California in 1980. Byck is professor of psychiatry and pharmacology at the Yale School of Medicine. He received his undergraduate degree in psychology from the University of Pennsylvania and his M.D. from the University of Pennsylvania School of Medicine. From 1960 to 1962 he was a research associate at the National Institute of Mental Health. From 1962 to 1969 he was a member of the faculty of the Yeshiva University School of Medicine; in 1969 he moved to Yale. He is the recipient of a Research Career Development Award from the National Institute of Mental Health for work on the somatosensory cortex in the brain of the cat. Byck's other research interests include the drugs that are employed in the treatment of psychiatric disorders and the effects of freezing on the tissues of the central nervous system of mammals.



Wall plate not included.

NEWEST THIEF STOPPER. A LIGHT SWITCH WITH EARS.

Imagine entering a darkened room that lights up at the sound of your footsteps. A room that's also 'smart' enough to flick off lights when it's empty.

Imagine too, an intruder attempting to pry open your window at midnight. Suddenly, room lights are ablaze, frightening him off—and your home and family remain untouched.

And who's your fearless protector? A light switch named AudioLite.

An invention that offers security and remote convenience straight from the pages of science fiction.

A slave to your every move.

AudioLite is also a wall switch you never have to touch—because it's sound activated. It turns on lights the second someone enters a room and turns them off when they leave (works outside too—in driveways, patios, porches).

How much sound does it take? You can adjust AudioLite's sound sensor to precisely any level. At high sensitivity even the clicking of a key in a lock or the opening of a door will turn on the lights. You can also lower its sensitivity so an extremely loud verbal command or noise will be required. Or choose an infinite number of adjustments in between.

Once activated, lights remain on for the time duration you set, anywhere from 7 seconds to 7 minutes.

Don't worry about lights going off when you need them. Each time AudioLite hears a new sound, it resets and begins counting over—keeping lights on without interruption. For quiet reading, simply use the manual On/Off override. Easily adjustable timer and sensitivity controls are recessed and child proof.

Save money in more ways than you'd think.

Hall lights usually left on for hours are now on only when you need them. Likewise bathrooms, closets, garages and patios. You'll find about half the light usage in such areas is saved—lowering your utility bill accordingly.

Because AudioLite is also a full range dimmer, you use exactly the amount of light you need—as well as extend bulb life by 15 times or more.

And you needn't turn AudioLite off during the day—a daylight-sensitive photocell assures no unnecessary lights get turned on during daytime hours.

Turn on lights with the snap of your fingers—even the creaking of your footsteps.

Magic that won't fade.

From the first snap of your fingers, AudioLite will be a constant source of amazement to you and your friends. And because it's solid state electronics—with few moving parts to wear out—it will outlast conventional light switches by many years. Its microcomputer circuitry is so reliable there's a one year total replacement warranty: should anything go wrong, the manufacturer will send you a new one, free of charge. After a year, prompt service is available by mail—but it's unlikely you'll ever need it.

AudioLite replaces your present light switch in minutes. All you need is a screwdriver. Works with any incandescent bulb up to 300 watts.

As soon as you try one, you'll want to replace all your conventional wall switches with AudioLite. That's why there's a special price if you order four at a time—$99. (For one only, the price is $34.)

However, you may order four on trial. If not perfectly satisfied after trying one, return them all within 30 days for a prompt, courteous refund. You're protected by The Sharper Image guarantee of satisfaction.

Call now toll-free, and protect your home with this newest brainchild of remote control technology. Be among the first to own a light switch with ears, eyes and brains.

ORDER TOLL FREE.

Credit card holders use our toll-free numbers. Or send a check. Order product #SM653 for the package of four: $99 plus $4.50 delivery. Order #SM652 for one: $34 plus $2.50 delivery. Add 6% sales tax in CA. And please mention this magazine.

800 227-3436
In California 800 622-0733
In Alaska and Hawaii 800 227-3826

THE SHARPER IMAGE™
755 Davis Street Dept. 8304
San Francisco, CA 94111
© 1981 The Sharper Image

For a year's subscription to our catalog of innovative products, send a dollar with your address.

METAMAGICAL THEMAS

Is the genetic code an arbitrary one, or would another code work as well?

by Douglas R. Hofstadter

It all began with a pesky student of mine named Vahe Sarkissian. I was telling my computer-science class about one of my favorite notions: the analogy between the complex machinery in a living cell that enables a DNA molecule to replicate itself and the clever machinery in a mathematical system that enables a formula to say things about itself. To my mind the resemblance is deep and fruitful; it has afforded me sharper perception in both domains. Although Vahe appreciated the analogy, he doubted the validity of a significant aspect of it, and so he brought the matter up in class. His challenge forced me to think the issues through carefully, and en route I encountered some fascinating details of cell biology I might otherwise never have known. What I find gratifying is how quickly people can come to appreciate these intricacies without having studied molecular biology.

Both of the profound 20th-century discoveries involved in the analogy depend crucially on codes: curiously arbitrary-seeming mappings from one set of entities to another set of entities. In metamathematics the code is Gödel numbering, in biology it is the genetic code. In Gödel numbering code numbers are assigned to various mathematical symbols (parentheses, for example), just as license numbers are assigned to cars or area codes to cities. This mapping connects entities from two intrinsically unrelated domains, one typographical and the other abstract.

The genetic code is likewise a mapping between two mutually unrelated domains. In this case, however, both domains consist of chemical units. To someone who is not familiar with chemical terminology the two domains might not seem sufficiently different for the startling quality of their connection to be apparent. The fact remains that in the course of evolution a scheme was worked out in which a chemical unit of one species is assigned as a code for a unit of another species. Actually a *triplet* of units of what I shall call Species I is assigned to a unit of Species II. In fact, sometimes several different triplets are assigned. For the moment, however, that is beside the point. The main thing is simply that members of two entirely unrelated species of chemical units are "mapped" onto each other. Vahe's question has to do with how arbitrary this mapping really is.

Species I is the nucleotides. Species II is the amino acids. If these words are not in your vocabulary, don't panic! You need not know word one about chemistry to be able to imagine this correspondence, this match-up between members of two different chemical species. All you need to know is that to each triplet of nucleotides (whatever they are) there is matched an amino acid (whatever that is). That is what the genetic code is.

The purpose of the code would be hard to describe without adding a little about the constituents of the cell. The "personality," or character, of a cell is stored in its genes. The genes, however, are essentially static, like words in a book. For them to come alive they must be translated into dynamic agents. Those agents are proteins; their actions realize the potential of the genes. They "express" the genes and thereby create the character of the cell. Genes are strings of nucleotides, and proteins are strings of amino acids. The cell's personality is therefore written in the passive chemical units of Species I. Through the genetic code this description can be converted into a vast population of dynamic agents made out of chemical units of Species II. Therefore thanks to the genetic code the cell's personality, implicitly defined by its genes, can emerge and bloom.

There are 20 different amino acids, and so you might think there would be 20 different nucleotide triplets. It is not quite that simple. There happen to be four different nucleotides involved in the genetic code, denoted A, C, G and U (which stand for adenine, cytosine, guanine and uracil). Every possible triplet (beginning with AAA, AAC, AAG and going all the way to UUU) stands for some amino acid. (Well, not quite. Three triplets do not, but for the moment that is just a detail.) How many such triplets are there? Sixty-four, of course: $4 \times 4 \times 4$. Hence 61 $(64 - 3)$ different triplets are matched up with 20 amino acids. Consequently some amino acids are coded for by more than one "codon" (a triplet of nucleotides). Indeed, there are some amino acids that get six different codons, some that get four, some that get three and some that get two; only a couple get one. The complete genetic code is shown, for your convenience, in the top illustration on page 23.

Back to Vahe. He noted that Gödel's numbering scheme was quite arbitrary. Kurt Gödel could have made just about any number correspond to each of the mathematical symbols involved; it would not have made the slightest difference to the success of his work. This is quite true. On the other hand, Vahe had the feeling the genetic code was deeper. It seemed to him intuitively that each amino acid is related to its particular codon (or codons) for some compelling reason, that there must be some fundamental chemical necessity for the relation. To caricature Vahe's position

Alanine	Ala
Arginine	Arg
Asparagine	Asn
Aspartic acid	Asp
Cysteine	Cys
Glutamine	Gln
Glutamic acid	Glu
Glycine	Gly
Histidine	His
Isoleucine	Ile
Leucine	Leu
Lysine	Lys
Methionine	Met
Phenylalanine	Phe
Proline	Pro
Serine	Ser
Threonine	Thr
Tryptophan	Trp
Tyrosine	Tyr
Valine	Val

The 20 amino acids

Adenine	A
Guanine	G
Uracil	U
Cytosine	C

The four nucleotides


VSOP
Fine Champagne
COURVOISIER
The Brandy of Napoleon
COGNAC
COURVOISIER. THE COGNAC OF NAPOLEON

What if you chose as a technical

"At Foxboro, our HP computer handles quality assurance data, and cuts paperwork by 70%."

Foxboro ranks among the world's leaders in process management and control systems. At their Highland Plant, in East Bridgewater, Mass., General Manager Joe Golemme tells us, "We market over 1,000 different instruments and systems and use an HP 1000 computer to control our entire QA data program. Because our HP 1000 delivers accurate information instantaneously, we can spot and prevent potential manufacturing problems on line. What's more, we've eliminated 70 percent of the paperwork.

"The same HP system maintains spare parts inventory and calibration programs for all our production and test equipment. Automatic calibration saves us over $80,000 a year in operating expenses. And because we know what spare parts to have on hand and when, we've reduced costly equipment downtime by 20 percent."

HP can be your personal computer partner, too!

You no longer have to compromise on versatility, performance, quality, and availability. The HP family of personal computers offers both a range of capabilities to meet your changing needs and a professional level of service. This growing family of personal computers is available through HP's network of qualified dealers, or through HP's computer sales organization.

For example: the HP-41/HP-IL, a battery-powered, portable computer system with mass storage, printer, and software application packages.

Or the portable HP 85, a fully integrated computer with video display, CPU, magnetic tape data storage, and thermal printer. Versatile, it can handle technical and business problems, run peripherals, and control instruments.

Hewlett-Packard computer partner?

"At Corning, HP personal computers reduce product development time and pay for themselves in one month."

Corning Glass Works, the world's leading manufacturer of specialty glass and ceramics, is also a major producer of diagnostic test kits. At the Medical and Scientific Division in Walpole, Mass., the Product Development Department uses five HP-85 computers to develop and validate all the tests they market.

Dr. Loretta Lee, Special Projects Manager, told us, "With HP-85s we've gained the on-line interfacing, graphics, and printout capabilities of a larger system—at a fraction of the cost.

"HP's enhanced BASIC lets us easily program our systems to handle other department needs, without additional expense. And now that we can directly interface test equipment, we've eliminated manual procedures and reduced throughput time by 20 percent."

Or the powerful HP 125. Specifically designed for the office, the HP 125 guides the user

through procedures step by step. It helps people do their jobs better with its impressive word processing, versatile management software, and graphics capabilities. Powerful as a stand-alone computer, it can also communicate with big computers. The HP 125 is supported by more than 100 software programs.

More good reasons to consider a working partnership with HP.

For more information on any of these personal computer systems, write to A.P. Oliverio, Vice-President, Marketing, Hewlett-Packard Co., 3000 Hanover, Palo Alto, CA 94304.

When performance must be measured by results

HEWLETT PACKARD

00241


COUNTRY CUB
The Lynx four-door wagon is designed for more important things than golf clubs, bicycles, or a week's groceries.
It is designed for people.
That's why it is the only wagon made in America* with a transverse engine, front-wheel drive, and fully independent suspension—a system that not only gives you a smooth ride, but also more room to ride in.
And while we've created a surprising amount of space, it hasn't come at the expense of fuel efficiency. In fact, the Lynx wagon has the best gas-mileage rating of any American-built wagon—either manual transaxle or optional automatic transaxle.
MANUAL 28** EPA EST. MPG 44 EST. HWY.
AUTOMATIC 29** EPA EST. MPG 41 EST. HWY.
The Lynx wagon. Good things do come in small packages.
1982 LYNX
LINCOLN-MERCURY DIVISION
Ford
LYNX FROM LINCOLN-MERCURY
*Excludes other Ford Motor Company products. **Compare these estimates 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 highway estimate. (California estimates lower.) Buy or lease from your Lincoln-Mercury Dealer.

one might say: "Gödel's code is the work of a mere mortal, but the genetic code is the work of God. Therefore the genetic code must be perfect, inevitable and unalterable."

I was quick to retort that, as far as I could tell, such was not the case. I said the genetic code seems every bit as arbitrary as the way the telephone company assigns area codes, every bit as arbitrary as Gödel's numbering scheme. I drew on the blackboard some pictures of the molecules involved to show my reasons for thinking this way. But as I stood there at the board a few things began to nag at me, places where I was not entirely sure of what I was saying, "facts" I knew I really ought to check up on. This desire to prove to my class the arbitrary nature of the genetic code has led me down some fascinating paths in molecular biology, and my findings are what I wish to report here.

Let us return to the cell. A cell is a little hotbed of activity, rather like a miniature town. There are basically two kinds of entity in this town. There are passive objects, "lumps" that just sit around and wait for something to do something to them, and there are active agents, "doers" that always want to get in there and do something. These active agents are for the most part enzymes. Each enzyme has a specific job it can carry out, and it does this job on lumps of a specific type. Actually it need not always be acting on lumps; enzymes can act on other enzymes. How an enzyme works is not our business here. We can assume that enzymes do their thing, whether it involves splitting some lump in two, welding two lumps together or performing some other chemical act.

FIRST POSITION	SECOND POSITION				THIRD POSITION
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

The genetic code. Sample reading: C **(left)**, *A* **(top)** *and G* **(right)** *code for Gln (glutamine)*

A marvelous thing about cells is that they are so elegantly designed that for many purposes one can ignore their chemistry and think just about their logic. In fact, that is the only way I know to think about the goings-on inside cells, since I am not a biochemist. Although I use the chemical names for things, my true image, deep down inside, is hardly one of chemicals. It is really one of little objects that somehow magically behave in certain ways specified in biochemistry books. My view of the chemicals involved in the processes of life is like most people's view of cars: they know what cars will do in all kinds of situations, but they do not really understand how cars work. I get a kick out of batting around technical terms of biochemistry when in fact I understand only their logic. ("The molecular logic of the living state," Albert L. Lehninger calls it.) The fact that one can get away with this is one of the beauties of molecular biology, and it is that beauty we are celebrating here.

An enzyme is just a kind of protein. And all proteins are strangely curled molecules made up of amino acids. The curliness is critical. Here is how I think about it. First, I imagine a large number of amino acids strung together like cars of a train. (Like train cars, amino acids have couplings that allow them to hook up at the front and the back to other amino acids in an arbitrary sequence.) Then I imagine holding this long string of amino acids between my two hands, tautly stretched to form a straight chain. Now I let the string go. Sproing! The crazy critter rapidly twists itself up into a tight little ball about the size of a fist. Now *you* try it. Here. Grab the two ends inside the ball and slowly pull them apart. The protein chain is resisting, of course, but if you are careful not to jerk it, you can uncurl it without breaking it anywhere. Got it all straightened out? Good. Now let it go. Sproing! What do you know? It returns to exactly the same curly shape. Now hand it back to me. Thank you.

It seems that a protein *likes* to be curled up in its little ball. That three-dimensional shape is called its tertiary structure. The tertiary structure of each type of protein is unique. The sequence of amino acids totally determines the tertiary structure; it is what makes the protein fold up every time into that same shape. The one-dimensional sequence of amino acids is the primary structure. Thus we can say that a protein's primary structure determines its tertiary structure. (Some proteins have a secondary structure as well, an intermediate level of coiling like a telephone cord, but tertiary structure is the essence of a protein.)

But so what if a protein always has a tertiary structure? So what if it folds up? The answer is that this folded shape is what determines the kind of doer this protein is. (If indeed it is a doer. Some proteins are not enzymes but mere lumps; henceforth we shall be concerned only with enzymes.) An enzyme's tertiary structure is characterized by certain bumps and clefts, like the nose and ears of a person's face, except that enzymes differ more radically than faces and are more convoluted. Certain parts of an enzyme are called its active sites. They are where the enzyme fastens itself, leechlike, to the lumps it is going to act on.

It is by the match-up between the shape of its site and the shape of its target lump that an enzyme "figures out" what to attach itself to. The enzyme and its substrate (the lump) are often likened to a lock and a key. No wrong substrate will fit. (Actually wrong ones sometimes

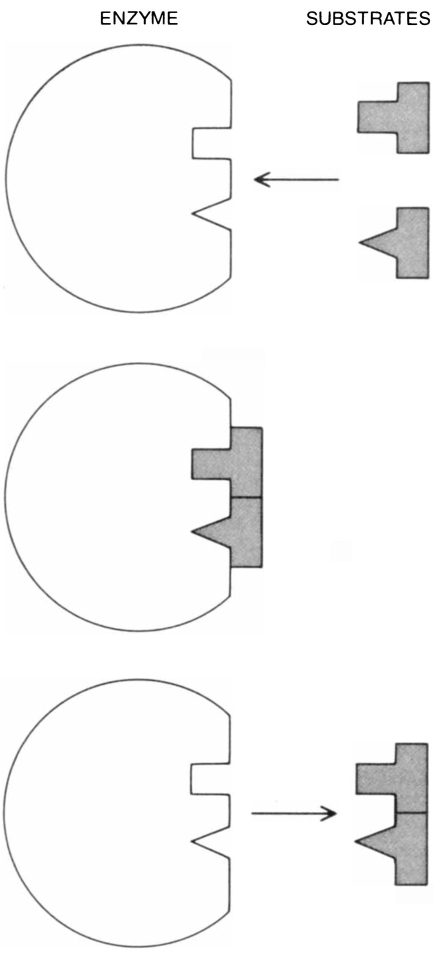
PROTEIN		Phe	Val	Asn	Glu	Asp	Leu	Cys	Gly	Ser	
mRNA	C	U U U	G U U	A A U	G A A	G A U	U U A	U G U	G G U	A G U	A

Amino acids in a protein chain **(top)** *are coded for by triplets of nucleotides in messenger RNA* **(bottom)** do under special circumstances, but we need not go into that here.)

An enzyme is very specific; it is tailor-made for a certain task and no other. Once an enzyme is hooked up to its substrates it starts churning, like a laundromat washing machine into which the proper coins have been inserted. The enzyme may rip parts off one substrate and attach them to another, it may bind two substrates together—whatever its thing is, it does it. Then it lets go of the product (or products), which are now free to go off and drift about inside the cell.

The upshot of all this frenzied activity by thousands of busy enzymes is the creation and sustenance of a unique living organism. It is these enzymes, these proteins, these coiled-up chains of amino acids that carry out the master plan stored passively in the cell's genes, which are chains of—but wait! We are getting ahead of the story.

There is nothing more important to know about cells than how enzymes work. They are what make cells run. There is one other thing, however, that is equally important. That is *which* enzymes are present and how they got there. Not all cells have the same set of enzymes, not by a long shot; that is why not all cells have the same character. What is more, a given cell's complement of enzymes can change, depending on both internal and external circumstances. Where do the enzymes come from? Ultimately, of course, they come from the genes, which are like blueprints, but that answer does not help at this point. What we need to understand is how enzymes are actually built, not where their blueprints are stored.



How an enzyme joins two substrate molecules

Remember that an enzyme is a protein, and a protein is a long chain of amino acids linked end to end and curled up into a ball. You might think that since enzymes are so good at taking things apart and putting things together, it is they that build proteins. The job is so delicate and specialized and critical, however, that a different kind of machine exists to do it. That little machine is the ribosome. A ribosome is partly composed of protein, but it is also partly composed of nucleotides. Its exact composition does not matter to us, though. After all, we are concerned only with the *logic* of the cell.

Is there one ribosome for each kind of protein? Hardly. That would lead to a frightful infinite regress. What would construct each type of ribosome? A specific metaribosome? And then what would construct each specific kind of metaribosome? In reality a ribosome is not specific to any protein; it knows nothing about the various proteins it builds. A ribosome is simply a general-purpose amino acid hooker-upper. It follows that something must tell it which amino acids to hook up, and in what sequence to do so. But what? For example, suppose it is to be lysine-leucine-glycine-proline-cysteine-histidine-tryptophan. (I chose the sequence purely for its musical sound. It is too short to be a real protein. Proteins are usually many dozens of amino acids long, like the seemingly endless freight trains that rumble across the Middle Western plains and hoot outside your motel room in Wibaux, Mont., late at night.) What tells the ribosome to start with lysine and finish with tryptophan?

At the risk of seeming to invoke an infinite regress, I shall now reveal that there is *another* train, this one composed of the chemical units of Species I: nucleotides. This train runs right through the middle of the ribosome, like a train through a station. Its cars, taken in triplets, tell the ribosome which amino acid goes first, second, third and so on. It is this train that, through the genetic code, tells the ribosome which amino acid comes next.

This train is mRNA: messenger RNA (ribonucleic acid). A molecule of mRNA is a long chain of A's, C's, G's and U's. An RNA chain is much, much longer than a protein chain. It may consist of thousands of nucleotides strung together like beads. An mRNA chain has special markers along it telling where the code for the protein begins and ends. That is why there are three special codons that do not stand for any amino acid. They stand for a quality such as "caboosehood." They convey to the ribosome: "Cut this protein off right now! Don't add a single amino acid more!"

We are coming to the crux of the matter. Where is the genetic code stored? I have made it sound as if ribosomes "know" the code, but they do not. Although ribosomes do the translating, they know neither language involved. How can this be?

Imagine yourself at the United Nations. An important speech is about to be given by Mr. Na, the flamboyant ambassador from Nucleotidia. A simultaneous interpreter of great skill, Meri Boso, is summoned. Ms. Boso has no idea, however, what language the speech will be in or what language it must be translated into. It looks bad. At the last moment, just before the speech begins, the members of a rescue team rush into the translating booth, where they suspend from the ceiling a huge number of tiny flash cards. Each card has on its front a word of Nucleotidian (curiously, all the words consist of three letters) and on its back the word's translation into the target language, which happens to be Aminoacidian. Meri Boso is saved! All she must do is listen carefully to Mr. Na, and then for each word she hears to find with lightning speed its flash card. Having found the card, she deftly flips it around so that she can speak its Aminoacidian translation in the nick of time into the microphone before her. Next word, please!

It's no sweat, being a ribosome. All you need to do is find the right flash card in a jiffy. But where are the flash cards in the cell? Even more to the point, *what* are they? At this juncture it seems that the genetic code has receded from view a little; it has become more decentralized, harder to localize. Whereas at first one might have guessed that the genetic code was somehow stored inside ribosomes, now it seems to lie in those flash cards. Hence if we want to determine how arbitrary the genetic code is, we must determine whether the flash cards could be changed and, if so, how.

The cell's flash cards are tRNA's: transfer RNA's. The term suggests that they are made out of the same stuff as mRNA is: A's, C's, G's and U's. This is true, except that some nucleotides are occasionally modified by enzymes, but for our purpose we can ignore that fine detail. At birth a tRNA is just an ordinary snippet of RNA. Then, quite unlike mRNA, which stays long and snaky, a tRNA folds up just like a protein, assuming a specific tertiary structure. This is in contrast to mRNA, which merely forms rather aimless curls over short stretches. The curling of mRNA is nonfunctional, whereas the curling of tRNA is functional. All tRNA's fold up into roughly the same shape: a chubby *L*, rather like the bent arm of Mr. America. At a more detailed level, however, the tertiary structures of tRNA's differ.

Once a tRNA is folded up it acts like a flash card in that it has an amino acid at one end of the *L* and a codon at the other. Actually it is not a codon but an anticodon. For our purpose, however, they are just two sides of the same coin. For each codon there is a corresponding anticodon, and vice versa. To make one from the other you merely interchange A with U and C with G. (A and U are said to be complementary, and so are C and G.) Therefore codon CUC's anticodon would be GAG and codon GAG's anticodon would be CUC. To be more explicit about tRNA, one end of it simply *is* an anticodon. The other end is a site where an amino acid can be attached.

In a nutshell, a ribosome is a translating mechanism between the two intracellular languages of Nucleotidian and Aminoacidian. The words of Nucleotidian are codons; the words of Aminoacidian are amino acids. The mRNA is a long speech whose sentences are written in Nucleotidian. The ribosome is a quick but ignorant simultaneous interpreter that uses tRNA molecules to assemble proteins, which are the word-by-word translations of the mRNA sentences into Aminoacidian. (By "quick" I mean the following. In a bacterial cell under normal conditions a ribosome can translate about 20 codons per second. In a rabbit cell things are slower: a little better than one codon per second.)

The mRNA "speech" is constantly clicking through the ribosome one codon at a time. On encountering a new codon the ribosome must seek out a matching tRNA, one whose anticodon corresponds to the codon. Of course, a ribosome has no eyes and cannot scan about as Meri Boso does. It must try one tRNA after another, the way the prince found Cinderella by her slipper. A mystery is how a ribosome can find a matching tRNA so quickly. In any case, having found one and clicked its anticodon into position against the mRNA codon, the ribosome snips off the tRNA's amino acid and snaps it onto the growing protein chain; then it releases the "nude" tRNA, which is free to pick up a new amino acid.

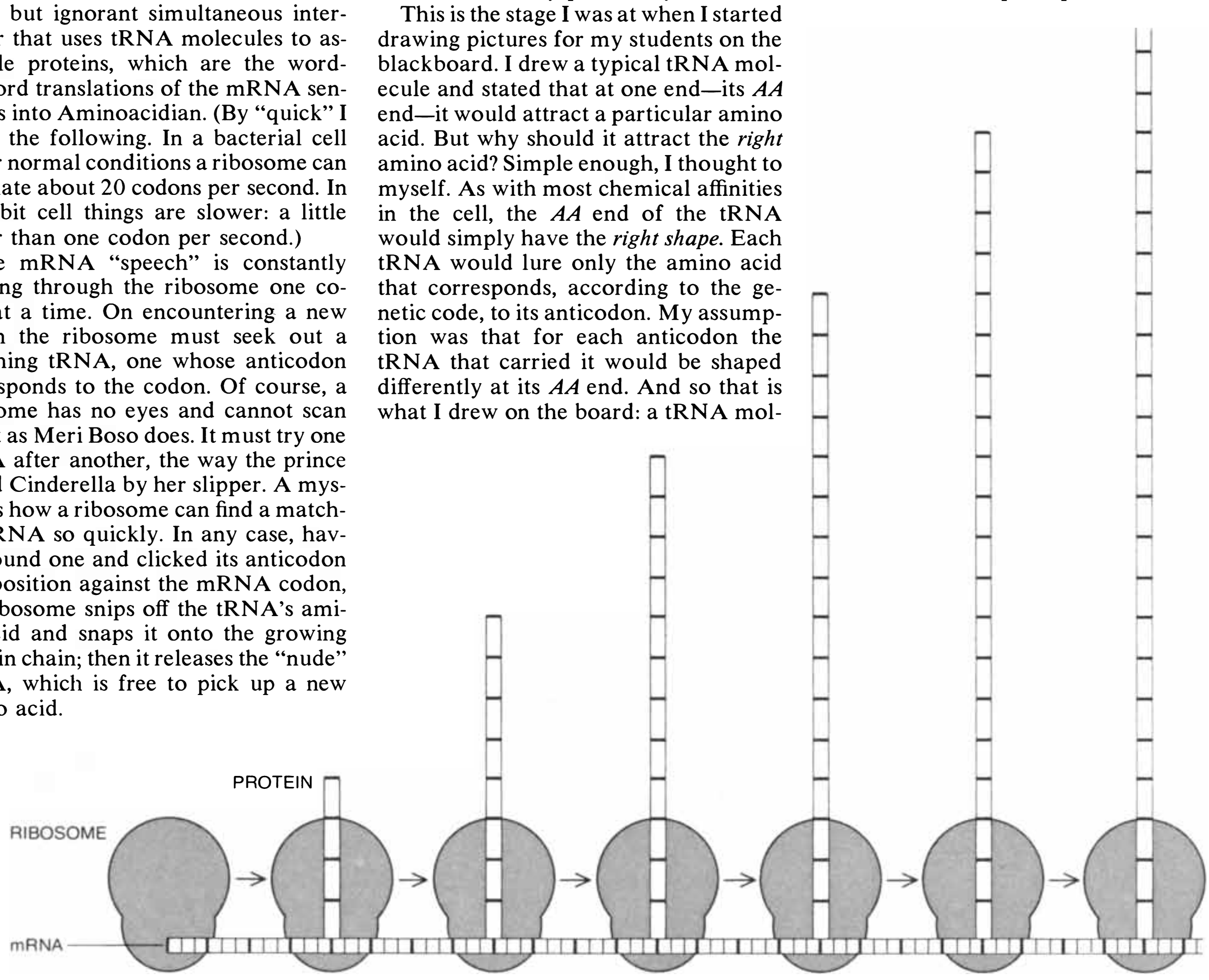
This is a salient difference between the metaphorical flash cards and tRNA molecules. Whereas flash cards can be used over and over again, each time a tRNA molecule gets used it has to be "recharged" with the right amino acid. Just where and how does this take place? Which amino acid should it get charged with? How is this determined? What determines it? All of a sudden these questions loom large. We shall return to them.

It is now apparent that if the genetic code is stored anywhere, it is in a spread-about fashion, distributed among the thousands of tRNA's floating in suspension in the cell near the ribosomes. Could these tRNA's somehow be subverted? Could they falsely guide the translation process? Certainly we can imagine the UN rescue team rushing in with the wrong set of flash cards, hanging them all up in Ms. Boso's booth and having her translate Mr. Na's speech into a completely inappropriate language. Could there conceivably be produced an entire set of "bad tRNA's": tRNA's with wrong amino acids attached to them, tRNA's that would fool the ribosomes into manufacturing nonsensical proteins? What could perpetrate such a nasty practical joke?

This is the stage I was at when I started drawing pictures for my students on the blackboard. I drew a typical tRNA molecule and stated that at one end—its *AA* end—it would attract a particular amino acid. But why should it attract the *right* amino acid? Simple enough, I thought to myself. As with most chemical affinities in the cell, the *AA* end of the tRNA would simply have the *right shape.* Each tRNA would lure only the amino acid that corresponds, according to the genetic code, to its anticodon. My assumption was that for each anticodon the tRNA that carried it would be shaped differently at its *AA* end. And so that is what I drew on the board: a tRNA molecule with a specific anticodon at one end and a specific "attractive shape" at the other end, a shape that would presumably combine with just one kind of amino acid.

Here a good question arises. Why does each tRNA attract the *right* amino acid for its anticodon, "right" being the amino acid defined by the genetic code? Why could some tRNA not fold up in such a way as to attract some other amino acid? Or is there some intrinsic connection between the two ends of the tRNA? Does the anticodon, for instance, somehow tell the other end of the tRNA how to fold up? This was one thought Vahe had.

I explained to the class that neither end of the tRNA knows anything about the other. You could surgically replace the anticodon with some other anticodon and the *AA* end would not know the difference. Conversely, you could surgically lop off the specially shaped *AA* tip of the tRNA and graft on an alien *AA* tip, which would then lure the wrong amino acid and make the tRNA embody a false piece of genetic code. I concluded by saying: "Since the two ends of any tRNA are independent, the genetic code can in principle be subvert-



A ribosome "reads" the sequence of nucleotide triplets in mRNA and spins a strand of protein ed and is therefore arbitrary." Then I blew the chalk dust off my hands and turned to another topic.

Well, it turns out that the picture I had drawn was right in spirit but wrong in detail. All tRNA molecules have at their *AA* tip precisely the *same* structure. For instance, the last three nucleotides at the *AA* tip are always CCA. Thus the site where the amino acid gets attached is completely nonspecific. There is no special chemical affinity between the *AA* tip of a tRNA and the amino acid that goes there! When I first found this out (after class was over), I was somewhat at a loss. How, I wondered, does the tRNA always end up with the right amino acid attached to it? What lures it there? Could it be the anticodon, even though it is at the other end of the tRNA? And if it is the anticodon, does that mean there is, after all, some special and intrinsic link between the anticodon and its amino acid partner? Is the genetic code, after all, inevitable?

By talking with friends and looking in books I found the answer. To my relief it supported my side, but matters turned out to be subtler than I had suspected. Although the *AA* end of the tRNA is indifferent to the amino acid that docks there, so that it can in principle accept *any* amino acid, under normal circumstances only one amino acid will get attached. This is owing not to the anticodon but to the tertiary structure of another region of the tRNA: its DHU loop. This is a loop that every tRNA molecule has, and it bends around in a characteristic shape in each different kind of tRNA. It is therefore a kind of three-dimensional signature by which the tRNA's type can be recognized from the outside.

But what could accomplish such recognition? Why, an enzyme, of course, one of the type known as an aminoacyl-tRNA synthetase. Such an enzyme has two active sites. One of them is the site that recognizes the tRNA's three-dimensional signature. The other binds an amino acid. That site, unlike the *AA* end of the tRNA, is *not* indifferent to the amino acid. It will bind one and only one amino acid, namely the one that is coded for by the tRNA's anticodon. To be sure, the synthetase itself never looks at the anticodon. All it does is "sniff" the DHU loops of various tRNA's, and when it finds one it "likes," it fastens its amino acid tightly to the tRNA and bids it farewell. For each amino acid there is one (or more) synthetase.

So here we have a funny thing. There are molecules floating about in the cell whose purpose it is to instruct the tRNA's in the genetic code. They pack each tRNA with a little burden and let it trudge off to encounter a ribosome somewhere. Can it be said, then, that the tRNA's themselves know the genetic code? No, they do not. They have to be instructed. What instructs them? The synthetases. And so can it be said that the synthetases know the genetic code? No. They merely match up DHU loops of various shapes with amino acids. Hence in the end we find out that *nothing* in the cell knows the genetic code.

To be sure, that is a slight exaggeration. The truth, again, is simply that "knowledge" of the genetic code is spread out. It is shared by the entire set of tRNA's and synthetases and cannot be claimed by either one alone. And yet there is a place where one might contend that the genetic code is stored all in one piece, namely the DNA. You might have been wondering when we would come to DNA, which is usually the star in tales of molecular biology. Well, this is the moment.

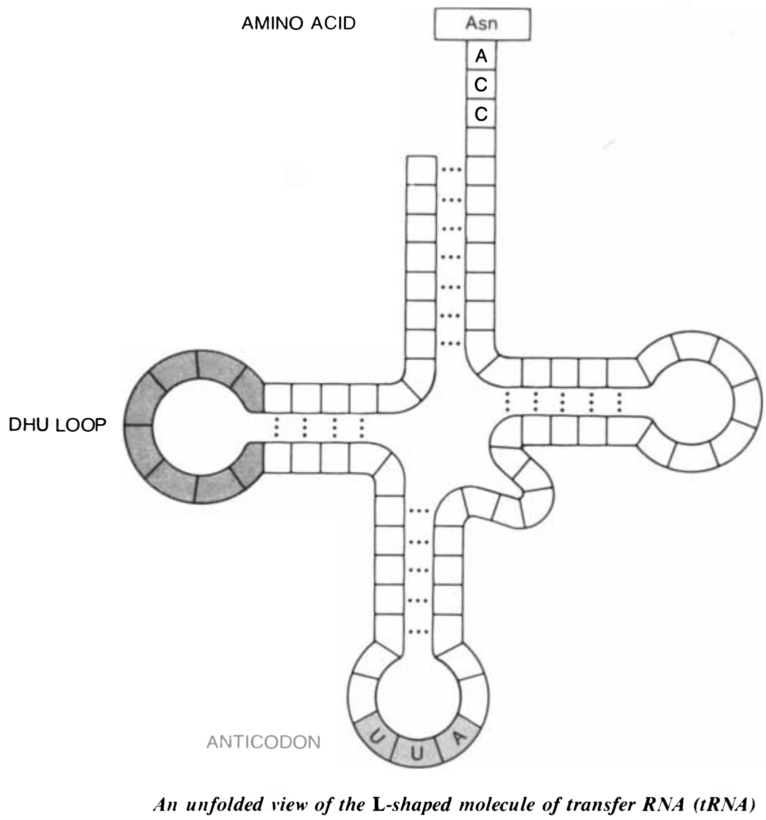
One can regard DNA as a big fat, lazy, cigar-smoking slob of a molecule. It never does anything. It is the ultimate "lump" of the cell. It merely issues orders, never condescending to do anything itself, quite like a queen bee. How did it get such a cushy position? By ensuring the production of certain enzymes, which do its dirty work. How can it make certain that a given enzyme will get produced? Ah, that is the trick.

DNA is a set of blueprints for all kinds of cellular constituents, lumps and doers alike. If you want to know where something in a cell comes from, the chances are the answer is: It is coded for in the DNA. The piece of DNA that codes for some specific entity is the entity's gene. The entity may be a protein, it may be a tRNA molecule or it may be some RNA that will eventually become part of a ribosome. Whatever the constituent is, there is a gene for it in the long, twisty DNA molecule. Indeed, that is why DNA is so long. The length of the DNA for a mere bacterium can be a million nucleotides, and for a human being it can be thousands of times longer than that!

This brings us back to nucleotides. DNA, like RNA, is made up of nucleotides, but instead of U it uses T (which stands for thymine). In DNA, A and T, like C and G, are complementary. For every strand of DNA there is a complementary strand that twists around it, making the entire supermolecule look like a double vine. The reason DNA does this and RNA does not is that A and U do not fit together as tightly as A and T, and so the twists of the would-be RNA double helix are not as stable as those in DNA. RNA can form a double helix for short stretches but not for long ones. That is also why tRNA's have short double-helix hairpin turns but are not double helixes all the way.

At the beginning I mentioned that an entity's gene is a *coded* version of the entity. Now where there is code there must be decoding. But there are two possible layers of decoding DNA. First of all you can decode it into RNA. This is done merely by complementation: A codes for U, T for A, and C and G for each other. The second layer of decoding is the decoding of the message contained in the RNA. That, of course, is the job of the ribosome.

If the cell wants to make, say, a tRNA

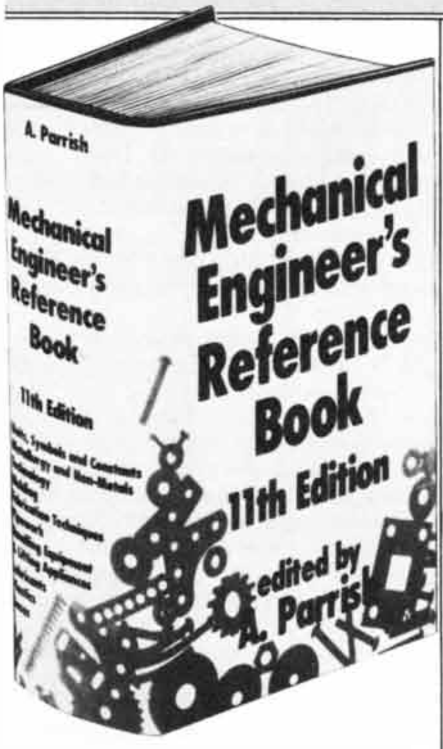


An unfolded view of the L-shaped molecule of transfer RNA (tRNA)

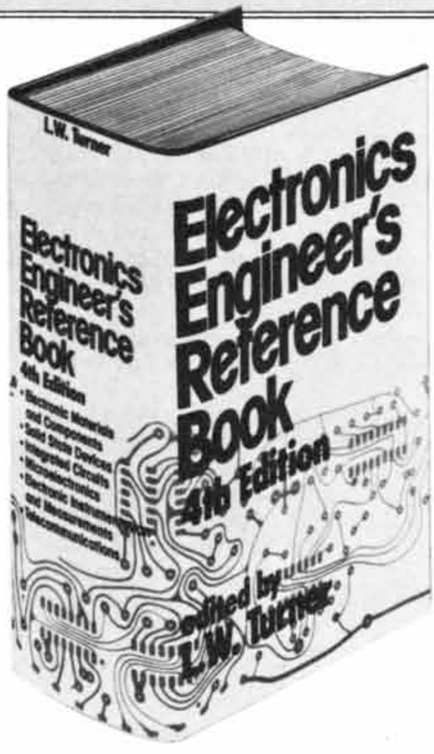
Take 1 of three Engineer's Reference Books for only $2.95

when you join any one of our three Professional Engineering Book Clubs—*the Electrical and Electronics Engineering Book Club, the Professional Civil Engineering Book Club* or the *Professional Mechanical Engineering Book Club.* Whichever club you join, you simply agree to buy 3 more books—at handsome di counts—within the next 12 months.

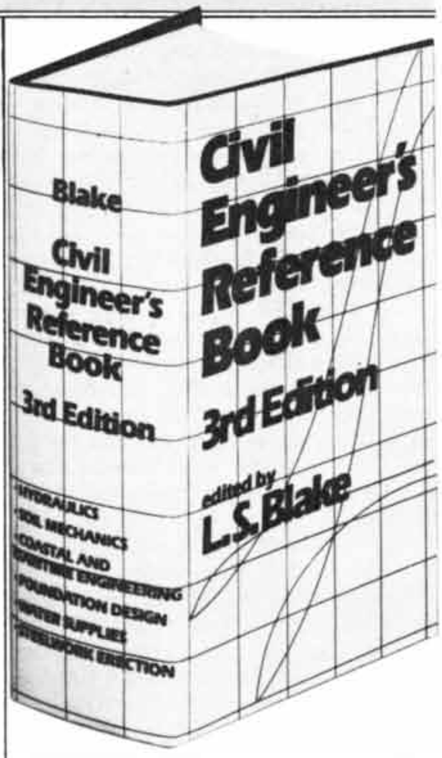
Each of the Professional Engineering Book Clubs provides a convenient and inexpensive way to improve your professional skills. The books we offer keep you up-to-date about new concepts, techniques and systems which are constantly being developed. Whichever field in engineering is your specialty—electrical and electronics, civil or mechanical—you can be sure that we have the ideas and answers to help you stay on top of your profession. So take the Engineer's Reference Book of your choice and join one of our three clubs.



61655-3. Mechanical Engineer's Reference Book. *Edited by A. Parrish.* Revised and updated. Comprehensive description of theory and design data and fabrication of materials. Chapters on thermodynamics, fluid mechanics, metallurgy, lubricants, computers, productivity measurement, instrumentation and much more. Extensive symbols and abbreviations. Thousands of charts, diagrams, formulas and illustrations. 37 engineering consultants contribute to 20 sections. 1700 pages. Publisher's Price: $79.95

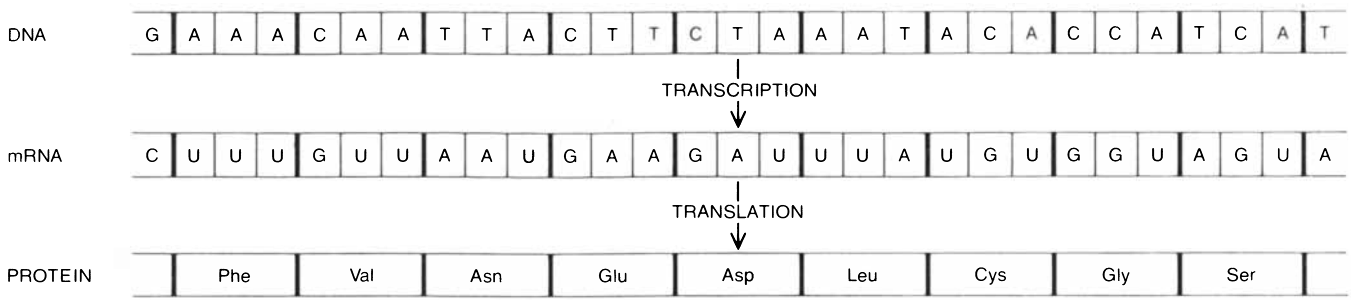


44145-3. Electronics Engineer's Reference Book. *Edited by L. W. Turner.* Completely rewritten and updated. Includes chapters on electronic materials and components, solid state devices, integrated circuits, microelectronics, telecommunications, electronics in industry, electromagnetic and nuclear radiation, sound and video recordings, and much more. 62 contributors write on 27 different sections. 1600 pages. Publisher's Price: $78.00



38250-3. Civil Engineer's Reference Book. *Edited by L.S. Blake.* Latest edition of this engineering classic. Concise chapters on mathematics and statistics, site investigation, hydraulics and surveying, soil and rock mechanics, sewage disposal, contract management and insurance. Current practices in design, structures and use of materials. Thousands of diagrams, charts, formulas, graphs and illustrations. 1800 pages. Publisher's Price: $84.00

If reply card has been removed, please write to Professional Engineering Book Clubs, Dept. -AE7, Riverside, N.J., 08075 to obtain membership information and an application.



How the genetic code of mRNA is transcribed from DNA and translated into protein

molecule, it uses only the first stage of decoding ("transcription"). It finds the tRNA's gene and in effect asks the DNA-decoding enzyme (an RNA polymerase) simply to manufacture the corresponding (complementary) segment of RNA. If the cell wants to make a protein, it uses *both* stages. First, as before, the cell gets an RNA polymerase to transcribe the gene for the protein. That gives rise to a long piece of mRNA. Then the mRNA is worked on by a ribosome, and out comes the desired protein. This second stage is called translation. And so you see that in our UN scenario Mr. Na did not write his own speech. Since he was merely the ambassador from Nucleotidia, he got his speech from the big boss back at home: the DNA. Mr. Na is merely a mouthpiece, a tool. He is just reading a transcript of the DNA's speech, and that transcript is slavishly translated by Meri Boso into Aminoacidian, according to the genetic-code flash cards. And even those flash cards were dictated by the big boss! You see what I mean about the lazy DNA being in control of everything?

The DNA incorporates coded versions of all the tRNA molecules, of all the synthetases and polymerases, not to mention the constituents of the ribosomes. Thus the DNA contains coded versions of its own decoders! By decoding their own genes the decoders manufacture more copies of themselves. You can see that this is a grand loop indeed. The genetic code is *locked in,* because decoders cannot help but produce more copies of themselves. Not only that, they produce enzymes that replicate the DNA, ensuring that new cells will have the same DNA using the same code.

Now, a code that is locked in is not the same thing as a code that is inevitable. For instance, the French language is locked into France, because not only do the adults in France speak French among themselves but also they teach it to their children. Moreover, they publish dictionaries and grammars that stabilize the language. This does not mean, however, that French is the only possible language in the world. The French word for something is not intrinsically tied to that thing by some God-given rule. French is an arbitrary code, as any other human language is.

Could it be that the genetic code is also changeable, in spite of its being locked in? How could one conceivably subvert it? What would be the cellular equivalent of my tampering with the table on page 23, making up my own genetic code? What kind of magic wand would I have to wave over a strand of DNA if I wanted it to use my own personal genetic code?

Let us set up the following hypothetical scenario. We take an ordinary functioning cell, reach into it and magically remove all its mRNA and tRNA. We throw those molecules into the garbage can. Then we reach back in and remove all the DNA (but we keep it), leaving behind a lot of random flotsam and jetsam, including some enzymes and ribosomes. Now the enzymes and ribosomes have nothing to do, since there is nothing left for them to transcribe or translate. If they will just be patient, however, we shall be right back. We go off and tamper with the DNA we extracted, and we then inject it back into the unsuspecting cell. Is it possible that not only this cell but also its progeny will now and forever use our new genetic code? What kinds of changes would we have to have wrought on that piece of DNA for the cell still to function exactly as before, except with the new code?

What does "function exactly as before" mean in this strange context? It means that the cell should look, to an outside observer, as it did before. What determines its overall functioning from that global point of view? The answer: Its complement of proteins. Proteins are what endow a cell with its character, its personality. Given this fact, how can we ensure that our cell's external personality is unchanged even though its internal language has been subverted?

Well, the moment we insert the altered DNA into the cell many RNA polymerases will start working on it. They will transcribe it into strands of RNA, namely short tRNA snippets and long mRNA trains. The tRNA snippets will fold up into their characteristic *L* shape. At this point various synthetases encountering the fresh tRNA's will slap amino acids onto them. Then the ribosomes will obediently use the charged-up tRNA's to translate the mRNA's. Hence if we are to produce the *same proteins* as before, we have to make sure of two things: (1) the new tRNA's must embody the new genetic code and (2) the new protein genes must be written in the new code.

Goal 1 is tantamount to making sure each tRNA has the right anticodon, according to the new code. To achieve this goal it will suffice for us to change three nucleotides in each tRNA gene in the DNA. Therefore the first set of changes to be made is to find all tRNA genes. In each gene switch the DNA codon, which, when it is transcribed, will become the tRNA anticodon.

To accomplish goal 2 we simply rewrite all the "literature," that is, all the genes for proteins, in the new language.

Now will things really work as we hoped? For instance, will the synthetases really do the right thing? Well, each tRNA will fold up exactly as before. (Remember that the anticodon has no effect on the way the rest of the tRNA folds up, in particular no effect on the DHU loop.) Now the synthetase comes along and encounters a familiar-seeming DHU loop. It sticks on the very same amino acid it would have stuck on before. The enzyme has been fooled in the way we wanted. It has become an accomplice to our treachery, because according to the *old* genetic code this tRNA is carrying the *wrong* anticodon for that amino acid, but according to the *new* code it is carrying the *right* one.

If you think this scheme through, you will see that it really works. A piece of "weird" DNA can be inserted into a cell whose layer-one decoding machinery (a set of RNA polymerases) is present, and the cell will proceed to manufacture new layer-two decoding machinery (ribosomes and tRNA's) and therewith to produce all the proteins coded for in the "weird" way in the DNA. Collectively these proteins will then imbue the cell with the same *external* character as it had before, when it was using the ordinary genetic code.

Thus we have succeeded in our aim of showing that the genetic code is just as arbitrary as Gödel numbering. And as a matter of fact it turns out that inside the mitochondria of many organisms just such a code switch has taken place! (Mitochondria are semiautonomous organelles inside cells whose purpose is to carry out respiration and to produce energy-carrying ATP for consumption by the host cell. They have their own ribosomes, DNA, tRNA's and so forth.) The genetic code of mitochondria is closely related to the standard genetic code: it differs only in four codons. Hence it is a *dialect* of Nucleotidian, rather than a completely new language, somewhat as Joual, spoken in parts of Quebec, is a dialect of French. Joual is just as locked into those areas as Parisian is to Paris. Mitochondria have their own tRNA's and their own genes, which would not be properly understood in the main body of the cell, but they get along perfectly well. Thus they confirm our original contention.

This excursion through the workings of the cell has given only the barest glimpse of the complexities and subtleties of the interlocking mechanisms that add up to life. Why do there have to be so many stages and so many intermediaries? Why are things accomplished so indirectly?

I am reminded of a visit I made to the plant in Chicago where *Scientific American* is printed. I was astonished by the degree of indirectness—that is, layers and layers of intermediary processes—of the complex machines. I kept asking about wheels and gears and pulley systems, "What is *this* for?" It always turned out that it gave the plant an extra degree of flexibility in some way that might not have been anticipated at first. In the development of almost any machine the earliest model is crude. Only the most straightforward applications and circumstances of use are taken into consideration. Then refinements introduced over the years result in levels of complexity that make the evolved system hard for someone not familiar with it to understand all at once. This certainly holds for automobiles, airplanes, radios, televisions, computers, even pianos.

In this light it is not surprising that the cell has so many delicately balanced mechanisms, some of which are there just to compensate for errors made by others. Sometimes biologists and biochemists write about these things in a way that makes it seem they have a wonderful view of the trees but have forgotten about the forest. The way I see the machinery of the living cell, the type of counterfactual thought experiment that comes to my mind—in short the view presented here—is assuredly not the way a specialist sees it. What I get out of a lucid and thorough treatise such as Lehninger's *Biochemistry* is a silhouette: the shadow projected by cellular processes into the space of information-processing concepts. I hope this is not wrong, because the shadow has a beautiful shape.

BOOKS

The banality of I.Q., the domestication of animals and a definition of physics

by Philip Morrison

THE MISMEASURE OF MAN, by Stephen Jay Gould. W. W. Norton & Company, Inc. ($14.95). A persuasive chronicle of prejudice in science, founded on scrupulous examination of the record, enlivened by the talent of a gifted writer, this volume takes on some of the sinister appeal of a tale of heinous crime. The epigraph bears the weight of the story. The young Darwin wrote: "If the misery of our poor be caused not by the laws of nature, but by our institutions, great is our sin."

Who did it? Mostly professors, both great and small. The famous naturalist Louis Agassiz, born in Switzerland, split the human species after he encountered black servants in a Philadelphia hotel. In the Harvard manuscript of a letter (which was published long ago without its key lines) Agassiz displays a pathological repugnance not to be repeated here, ending with "God preserve us from such a contact!" His articles and letters, couched in the terms of natural history, defended before and after Emancipation the social policy of strict segregation. The equally renowned Paul Broca was a Paris brain-weigher, and on his weighings he rested a view of all human worth based on the ineluctable verdict of the balance. He argued from the conclusions: white males were the best because their brains were the heaviest. He introduced very plausible corrections to crude weight once a study showed that German brains outweighed French ones! He forgot corrections for body weight, however, when he took data on women.

The physician Cesare Lombroso pursued a more complex path: his specialty was not brains but whole bodies, a search for the apish in the lowly. He found born criminals to be evolutionary throwbacks. The pictures and citations from this work are hard to take seriously now, although it cannot be doubted that in their day the ape stigmata were influential. (Count Dracula himself was fitted to Lombroso's account of the born criminal; be wary of an aquiline nose, bushy eyebrows and pointed ears.) If you want still more facts, you can examine the atavistic tattoos and the drawing of the prehensile foot of prostitutes.

So far we have glimpsed half of the rueful comedy. The story gets more modern and more serious. Now we enter the domain of the I.Q., a single numerical ranking said to be invariant, heritable, full of social power. It was manufactured in America, after a French invention. Alfred Binet of the Sorbonne decided to institute psychological tests for identifying schoolchildren who needed some form of special education. His early work on craniometry had led him almost to surrender: "The idea of measuring intelligence by measuring heads seemed ridiculous." Multiple tests with quick everyday tasks seemed practical. The particular tasks were numerous, almost random, themselves unimportant. The theorist Binet was not self-deceived. He saw that his scheme was altogether empirical. Intelligence could not be captured by any simple single number. He feared ranking and the self-fulfilling prophecy, and he refused to rank all pupils; he wanted merely to select those at the lower end who needed some help. He sought not to set a limit but to open up an opportunity.

The American developers of I.Q. ignored all Binet's caveats. The men whose work is discussed here are not widely remembered today, although their bias lives. It even outlived its effect in the minds of the pioneers. First H. H. Goddard, who devised the word moron and the once famous name Kallikak for his exemplary account of a century of familial feeblemindedness, recanted. That came a generation after his book, in which one can view the feckless expressions of several Kallikaks out in the Jersey pine barrens where Goddard's institution was. "There can be no doubt that the photographs... have been retouched." The manipulation was crude by modern standards; the expert verdict is sure. Those oafish appearances have clearly been heightened for effect.

The famous Army Alpha tests of World War I redesigned Binet's craftsman's improvisation for the mass market. Again the expert himself finally saw through the elaborate misuse of the rich but faulted data. A young Princeton professor, C. C. Brigham, was the postwar author of a key monograph on the Army Alpha tests praised by Harvard's R. M. Yerkes, once the colonel in command of Army psychological testing. By 1930 Brigham recognized in print that his effort to find racial differences by combining I.Q. scores was formally inconsistent, and moreover that his tests were culturally biased. "One of the most pretentious of these comparative racial studies—the writer's own—was without foundation." This remarkably disarming admission came too late. The Immigration Restriction Act of 1924 had become law. Cool Cal said, "America must be kept American," as he signed the bill. The quotas limited immigrants from southern, central and eastern Europe to a trickle. With nowhere to go, many were caught in the Final Solution, to whose crimes the eugenicists and the testers had made the unintended contribution such ideas can bring.

The most elaborate chapter is the one devoted to factor analysis. It begins with the story of Sir Cyril Burt, whose extraordinary career of fraud and deception seems unique even in this company. Burt was the leading British factorist of his generation. The technique is a procedure for drawing inferences of statistical correlation from multidimensional measures. Gould presents a clear account of this technique without resorting to algebra, by simple and persuasive geometrical arguments. His exposition is a fine piece of popular mathematics. There was a long wrangle between the British school that saw a single factor for intelligence (roughly the average among all the measures) and its powerful American critic L. L. Thurstone. Thurstone sought instead a set of vectors in the abstract space behind the tests that could show positive correlation for a number of the measures, where the single-factor approach implied negative correlations for many less important measures. In this way one could find not a hidden single factor but the several "vectors of mind."

Thurstone was plainly right; the single factor is an artifact of assumption, but so in truth were his few vectors. Correlation is not cause; Professor Gould observes that the cost of gasoline correlates nearly perfectly with his own age. The "automatic reification" of merely formal factors is invalid. That the current best-known American hereditarian still defines intelligence as the single Spearman *g* factor behind all the tests fixes the place of this unusual account in the book. People might do well or poorly in many tests either because they were born smart or stupid or because they lived as children in an enriched or a deprived way. The correlations are the same, but their causes lie beyond the reach of such statistics. Neither naive theory is apt to hold much truth.

Our species might indeed have developed into a set of isolated subspe-


It's a Girl!
IMPORTED
Canadian Club
Hiram Walker & Sons
THIS WHISKY IS 6 YEARS OLD

We believe that fuel injection, like front wheel drive, is a necessary component for any automobile that is truly of the '80s. We hope to explain why on the adjacent page. And, we also hope after reading it, you will agree.

How Volkswagens work:

Keeping an eye on the engine.

The fuel injection.

cies with important genetic differences. There has not been enough time. Local variations in gene frequency still far outweigh the variations among races in the well-measured blood-cell antigens. It is a fact that a few traits give rise to perceived external differences. Such visible signals, intrinsically insignificant, have been put to powerful social uses, epiphenomena of the complex course of history. The mind is not likely to be adapted for much other than learning well; we are in fact the "learning animals." We need to learn more.

DOMESTICATED ANIMALS FROM EARLY TIMES, by Juliet Clutton-Brock. University of Texas Press, British Museum (Natural History) ($24.95). Two distinct intelligent predators have hunted in packs across the tundra of Eurasia by day or by night over the past 50 millenniums. The old established hunters were the packs of the true wolf, *Canis lupus,* whose culling of the grazing herds had gone on for a million years. The second form was an interloper, a tropical primate newly emigrated to the subarctic, bands more deadly than the wolf pack, denless, able to carry the helpless young along or to bring them back food in bags made of skins. Once dependent mainly on the gathering of plants, the human animals were probably learning to count more on meat in the strongly seasonal growing climate of their cold new homelands. Pets came easily to women and children, who would tame young animals of any kind that would stay near the camp. Most animals would leave on approaching adulthood; inherently solitary, they would move out to seek food and a mate.

The wolf, however, shares many of the social responses of our species: we are both organized as hunting teams in a hierarchy of ritual dominance and submission, serious fighting turned into a playful symbol among kindred. The wolves stayed. By the slow selection of animals for curled-up tails, or for black and white pelts instead of wolf gray, or for some other endearing trait, a new manmade species of mammal arose: the dog. The first dog bone we have is 12,000 years old; it is a distinctly different jaw. All over the world man tamed wolf pups; the ancestry of the dog may include several subspecies of wolf. The breeds of modern dogs are not, however, to be traced back to wolf forms. The genetic variety now found among the "scavenging curs" of any peasant village seems enough to explain the slow formation of many breeds by selection down through the years.

A breed is a distinct form reproductively isolated by human intervention; it may or may not be localized, whereas a subspecies is almost always geographically restricted. The greyhound and the spotted bulls of Apis, bred for religious reasons all over ancient Egypt, to be buried at death in one underground tomb at Saqqara, are among the few breeds of domestic animal for which we have ancient pictorial evidence. The characteristics human beings select are often superficial, in the sense that they are not clearly reflected in the archaeological record built on skeletal remains.

The "enfolding" of another species within the society of our own is a subtle matter that clearly depends on fine details of the behavioral fit between the two species. For example, the African hunting dog of the game-rich plains, *Lycaon pictus,* is fully social: it lives and hunts in packs. *Lycaon,* however, has not developed dominance hierarchies, and it relies for much of its binding communications not on the visual and tactile signals of expression and posture seen in the wolf, which we easily emulate, but on the exchange of regurgitated food. It was these animals that Isak Dinesen once saw moving at midday across the Masai grass country, not in the usual pack of a few or a dozen but in a procession of 500 moving in a slow canter, "in the strangest way, looking neither right nor left.... They hardly seemed to see us.... I have told this tale to many people and not one of them has believed it. All the same it is true, and my boys can bear me witness."

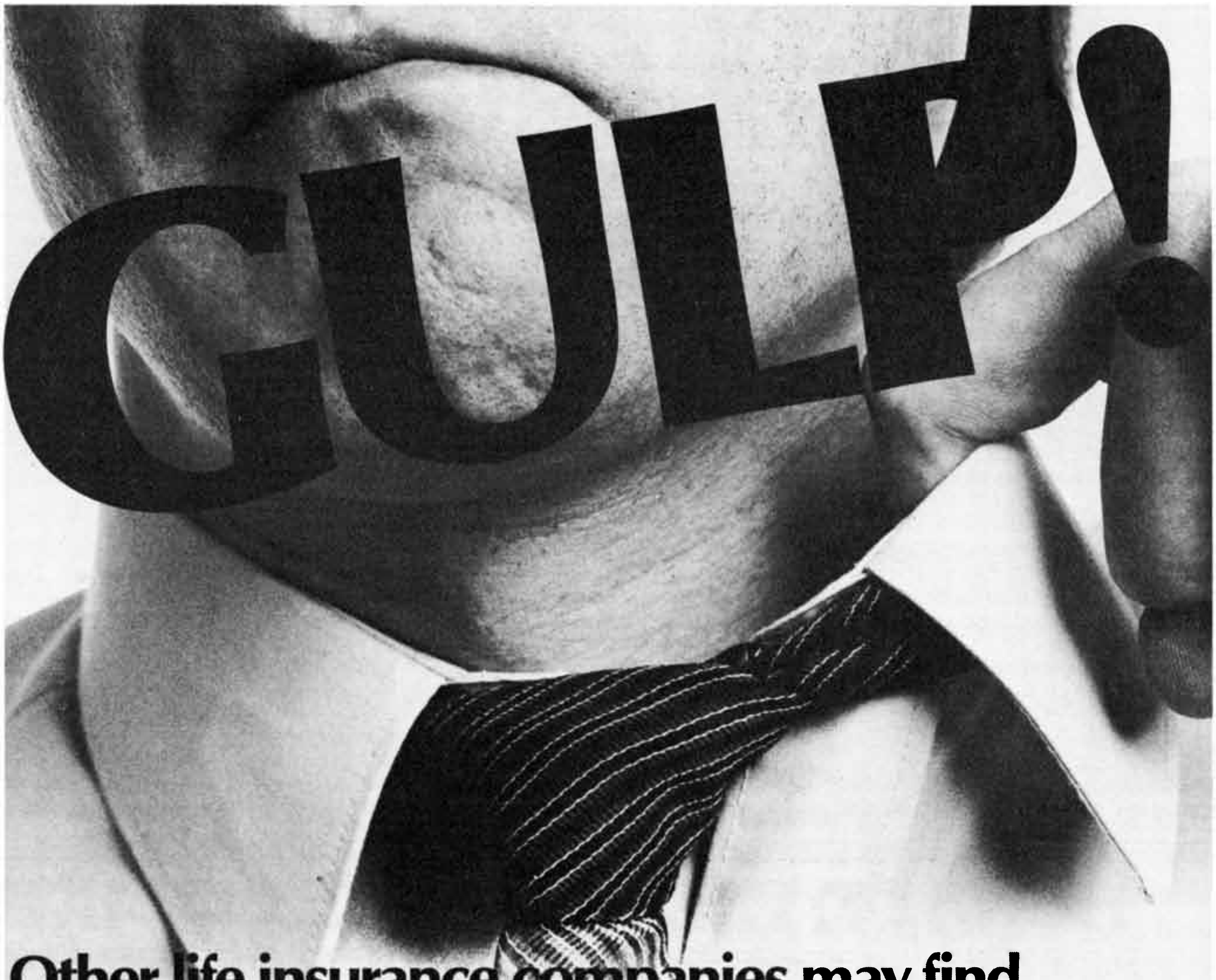
Those indifferent aliens have never been domesticated. Nor have we domesticated most species of deer, antelope or gazelle, even though the red deer was the prime game animal of the European Paleolithic; its remains are found in 95 percent of a large sample of sites. Even their antlers were collected on shedding each spring; they were used for tools, mining picks and hafts for stone axes. Male red deer, however, are both territorial and adapted for running off to escape fleet predators. Fencing them in can cause them to panic; the rutting males are unpredictable, even dangerous. The deer park is the antique answer, but these are game for royal field sport, not herded cattle. The first domesticated ungulates were the sheep of the foothills of the mountains of southwestern Asia. They are predisposed to follow a leader, flourishing in a local home range even in crowds; for them survival is finding food in the harsh land, not fleeing the swift predator. Goats were manmade about as early; they were still hardier creatures from the higher uplands.

The author of this readable, straightforward, brief but comprehensive review is a mammalogist with much archaeological research experience. The title of the book could mislead a reader; it is in fact limited to mammals and says nothing about birds or other domesticated species. Twenty-one species of mammals (out of some 20,000 in all) are forms whose reproduction is so closely controlled by man that they now represent populations quite different from those of their wild counterparts, most of which were known to Linnaeus. One photograph shows a white reindeer, a strain giving evidence of human control. Camels, although at least the one-humped dromedary has no living wild counterpart, have not changed much under human control, so well-fitted are they in their stubborn inflexibility to the intolerant desert. There are dromedaries bred for racing, but many herds of camels still live half-free, the herds guarded only during the season of rutting.

Elephants are hardly bred. They are "exploited captives"; their life cycle is long and their mating behavior is difficult to control. The tamed wild elephant has a history going back at least 3,000 years. The elephant of war is well documented; like many another new weapon it was important mainly as a symbol. Elephants very sensibly retreat under heavy attack, generally to damage their own army more than a resolute enemy. Pliny describes elephants trained to pick their way among seated guests to find their alloted place at a banquet. Ivory is still taken in large amounts; ice-age men in the Ukraine already made the framework of their huts from elephant bone and tusk, in a design not unlike that of the Ituri hunters of today. Long before that elephant was eaten. For half a million years human beings have exploited the elephant without ever breeding it in captivity. Zoos have now begun to do it, but it calls for long foresight.

Pigs, reared both as rangy animals in the forest and fat swine penned in the sty, are domesticates old in western Eurasia and even older in the Far East. Why pigs are favored in China and in New Guinea but taboo to Islam and to the Jews is hard to say. It is of interest that up to 3000 B.C. pigs were commonly eaten in Sumer and in Egypt. Yet Herodotus reports that later Egyptians were so anxious to cleanse themselves from accidental contact with a pig that they would plunge fully clothed into a river. At such a point we lose contact with adaptive simplicities, although there may have been some simple but transient reason.

This book does not deal with modern breeds of domestic animals; it treats the issue as of about the time of Rome, with allowances for other lands. It has much to say about cattle for meat, milk, draft and ritual sacrifice. Several chapters cover horses and donkeys, human lactase deficiency, onagers, yaks, mithan and water buffaloes. We are a species that has strong relations with other species, sharing even the diseases of our domesticates (a topic omitted from the book). There is a chapter on new domestication attempts, with the eland, the elk, the musk ox and more. The elands are doing well as a foundation stock for meat and milk, not so much in Africa as



Other life insurance companies may find Northwestern Mutual's new Extra Ordinary Term policy a little hard to swallow.

Until now you had two choices. Whole life or term insurance.

Until now.

Recently, Northwestern Mutual Life has unveiled an entirely new life insurance plan.

We call it Extra Ordinary Term.

You might just call it the answer.

Unlike most term policies, our Extra Ordinary Term plan provides level premiums plus the cash-building advantages of whole life. Never before has Northwestern Mutual Life offered such an innovative insurance plan. But innovation such as this is nothing new for a company that for the past 125 years has been a leader in low-cost long-term life insurance. A record that is the envy of the industry.

And a record that has been maintained by constantly improving the products we offer you. Our new Extra Ordinary Term plan is just another reason why other life insurance companies have come to realize that we're a tough act to follow.

Northwestern Mutual Life
The Quiet Company

A tough act to follow

in the southern Ukraine. The llama, the dormouse, the guinea pig, the ferret and the mink are all included at least briefly.

We have one symbiont that walks by itself, the only species with which the relationship is as a rule equally beneficial. The territorial carnivore *Felis catus* has been persuaded to share the core area of its home range with man by offerings of food, warmth and affection. Even manmade cats can still, however, flourish indefinitely on their own; the stray cat copes worldwide, half wild or even as feral carnivore, "treated with respect by humans who cannot resist the compulsion to feed and protect these shadowy but endearing creatures." We were and we remain nurturant forms, with an instinct to share; someday we may cease to be wolves to members of our own species.

THE INVESTIGATION OF THE PHYSICAL WORLD, by G. Toraldo di Francia. Cambridge University Press ($19.95). A Florentine physicist, learned, skeptical, above all lucid and direct, offers a definition of physics as that which can be studied with profit by the method attributed to Galileo. It is experimental at heart, interrogatory, but its questions are put to nature with due regard to the simplicity of the answer expected and to its quantitative form. His book is an adult introduction to physics in this mode, touched a little by the ironic. It is not mathematical, making no technical demands beyond those of high school algebra (if one or two integral signs do creep into a few pages); its aim is to describe principles, not by case studies drawn from history or by special explanatory examples but by presenting an explicit broad introduction to those principles as they appear today, "keeping in mind at all times their epistemological aspects and the critique of their foundations."

That suited exactly the original audiences, philosophy students at Florence in a seminar on the fundamentals of physics. It should suit very well a wider readership of teachers and students, of serious general readers in science and indeed all who welcome a reflective, consciously appraising exposition of the main topics of physics. It does not escape Professor Toraldo's notice that the Galilean scheme has made headway in a number of disciplines once far beyond its reach: chemistry, geology, mineralogy, biology, astronomy. With this shift (which would have been most congenial to the Grand Duke's Mathematician) has come a gradual shift away from the sharp distinction between general laws and their factual realization. The contingent begins to find its own causes in the past; the elementary, like the once structureless atoms, begins to show a certain contingency. The most refractory of units are modular only within a given context; even the proton has its quarks, if they are constituents with an unexpectedly stubborn cohesion.

In such a style the reader is led from a key initial chapter on method through four substantive and substantial essays. The first chapter treats the physics of the reversible, Professor Toraldo's crisp title for the classical (and relativistic) study of the well-determined; the second chapter extends our view to the irreversible, to heat and thermodynamics and on to probability and information. The third examines microphysics, taking a view of quantum mechanics that emphasizes its power and rationality without dodging its rich and perplexing novelty. The final chapter extends physics into its historical domains: the origins and fate of stars, galaxies, the earth and other planets, even life and thought. The last page is a reflection on the reflective, on the fact that finite human animals can somehow model the ample universe. "However, this book ends here, and another should begin; I shall refrain from starting it."

Examples of the approach to a few much discussed epistemological issues will suggest the flavor of this book to the discriminating. Possibly the most helpful is Professor Toraldo's careful account, based on a qualified adherence to the idea of the operational definition of physical quantities, of the limits of validity of the laws of physics. The truth of physical equations is always modulo some interval of precision. That does not represent early error, blunder or mistake but rather an inescapable limitation of view, which will likely enter in the same way, although at a different point, in any replacing "revolutionary" new discoveries.

Was Lavoisier wrong in saying that the mass of reagents was invariant through chemical reactions? To this day that fact governs our powerful chemistry; "it would be ridiculous to teach physics and chemistry" without it, even though of course it is not relativistically true. There is a hard-hitting section on how we see; its conclusion is compelling. It is naive to regard visual data as immediate and clinching sources of knowledge. They call for their own brand of verification; seeing is not the same as believing. Time's arrow is well explained; Professor Toraldo leans in the end to the importance of branching systems and to the psychological locus of some of the more conventional questions that arise under this heading.

On quantum mechanics the text concludes: not all the properties of objects belong solely to themselves, some are related to interactions with the means of observation. Reality is not to be prescribed by the preconceptions of the mind; we need to extend our understanding of reality by our study. Induction is subtle, and even falsification is no sure road to truth. Not even the invariance of the laws in space and time is a necessary postulational platform for the approximating physicist with his wits about him.

An informed reader will not be persuaded by all Professor Toraldo's ingenuities, nor is that to be expected. An air of freshness nonetheless drifts through the entire account, particularly in the frequent assessments of the current claims of the various schools of the philosophy of science. His is a world view that is unpretentious yet secure from unreasonable doubts. We know for the most part what we know, which is by no means to say everything. The footnotes and the side glances of Professor Toraldo are a running pleasure. It appears that this book is reworked from the Italian version of some five years ago, brought up to date and well translated by the author himself.

TINNITUS, edited by David Evered and Geralyn Lawrenson. Ciba Foundation Symposium 85. Pitman Books Ltd., London, distributed by Ciba Pharmaceutical Company, P.O. Box 12832, Newark, N.J. 07101 ($35). Tinnitus is the term for noises in the head. Sensations of sound that are not responses to any external signal are not a rare condition. Most people with some hearing impairment complain of false signals; perhaps 1 percent of the entire population report severe annoyance from ringing or buzzing, and more than one person in 10 knows the effect, without counting mere occasional sounds or those that follow exposure to "discos, shooting, or noises at work."

This volume is a symposium report, very up-to-date and expert, bringing together clinical and research people in London from Europe, the U.S. and beyond. The topic is not a new one, but interest is new. Something is stirring; the exhortation "Learn to live with it" no longer seems enough. Reading the papers and the carefully reported free discussion is itself a glimpse of a modern medical specialty in its bewildering variety. The first paper seeks definitions and classifications. (A working party was mustered to find them, and it reports plausibly at the end of the volume.) A dozen more papers treat of epidemiology, of matching and masking tests for the nature of the tinnitus (its pitch, loudness, bandwidth), of neurophysiological studies in cats and in human beings, of personality correlates, causative drugs, surgical therapy, therapy by masking sources of noise and finally the condition as it stands in courts of law.

Drugs? Aspirin and quinine are famous inducers of the symptom. For therapy the mild anesthetic lidocaine can bring brief but dramatic relief, at some price in side effects. Surgical treatment, even cutting the nerve, may have little or no effect. Like pain, the condition can be relieved by biofeedback training toward muscular relaxation. Local passage of electric current, under careful control, has helped. The most effective weapon of the new arsenal is the provision of a masking sound, often with a device that at the same time amplifies hearing. The masking is often best when the masking signal is narrow-band sound at just the frequency of a well-defined, high-pitch tinnitus, but the patient is then not helped; a less obtrusive broader-band masking sound is more acceptable. A lucky patient may "depart from the Tinnitus Clinic not only delighted with the relief afforded by wearing a masker, but also amazed at the ease with which it seems to have been achieved." The relief may continue for months after the masking signal is removed.

The most heroic of the therapies for profound hearing impairment is the cochlear implant. A microphone and amplifier are packed by the patient; their audio signal modulates a 16-kilohertz carrier that energizes a coil embedded in bone behind the ear. The coil feeds the amplitude-modulated current to electrodes set into the cochlea. The neurons there pick it up and furnish a sensation of sound. All 50 implant patients so far have gained sound from this device. The implants seem to relieve inner noises as well.

But it is not the energetic and diverse search for therapy that is the greatest interest of this volume for the general reader. The excitement arises—a fourth of the book is devoted to it—from recent unmistakable findings that the ear is not only a subtle microphone but also a loudspeaker. For about a decade there has been evidence that the cochlea can emit acoustic signals, not particularly loud ones but tones plainly heard and recorded by the ear-canal microphone of the investigators. They are narrow-band tones, some 10 hertz wide, somewhere in the audio range, mostly from one kilohertz up. Often the tone is not disturbing to the person emitting it; this public tinnitus "may not be the type for which alleviation is most frequently sought." It has been found in an elderly guinea pig, in aspirin-drugged cats, in an entire family of human subjects and now more widely.

It is striking that this very phenomenon was predicted in 1948 by Thomas Gold, well known by now for his original insights into many realms of the external world. His prediction was largely overlooked; it depended on his recognition that the sharp-tuned performance of the ear's pitch analysis was not consistent with the fluid-drenched mechanics of the soft cochlear lyre. These tuned resonators had to be regenerative to work so well, with neuroelectrical energy driving them to enhance their response. Such a system, like the regenerative radios of a generation ago, would surely sometimes oscillate on its own. It is thus given to some people who ask why they hear a tone in the inward ear to be told "because it is there." (A tape is available for the skeptic.) The clinical importance of the effect and the bearing it has on hearing theory are still issues before the experimenters. Sonar-like echo techniques have confirmed the effect in another way as well.

The study of defects of human perception is not only of therapeutic importance, not only part of a wider understanding of how the senses function, but also of general interest to all who pursue science. The errors of evolved perceptions and their recognition serve as a parable for the growth and improvement of science, itself a kind of cultural perception. The acoustics of the cochlea was meticulously studied in a lengthy and celebrated investigation. Those results seem now to be of little use; the study was made on dead ears, all feedback circuits off, about as relevant as careful resonance measurements within an amplifier never plugged in.

TO THE ENDS OF THE EARTH, by John Perkins, with the American Museum of Natural History. Pantheon Books ($27.50). "Much of this story is told in the explorers' own words, and their photographs show us what they first saw. When they went exploring, we were caterpillars, measuring the earth step by step. Now we are moths, flying through the air toward distant lights."

Across the large pages of this pictorial account of scientific travels the author has constructed a small work of art largely out of the photographs, letters, field notes and reports of four expeditions sent out by the great New York museum over the 50 years from the 1890's, before that metamorphosis to flight. He has gone farther afield to add context to the narratives, but all except a handful of the photographs are from the holdings of the museum. An anthropologist and filmmaker, the young author has by his expressive writing, his brilliant choice of pictures and his sensitivity to larger historical issues turned a simple nostalgic record into a book rewarding both to the knowing reader and to the innocent.

The four sorties are those of Robert E. Peary into the Arctic, north from Baffin Bay to the pole, the five-year venture of the young pair Herbert Lang and James Chapin, mostly to collect natural-history specimens along the Congo (it took four volumes to describe only their birds), the half-dozen ventures of Roy Chapman Andrews by car and camel into the Gobi after fossils, and the remarkable Jesup North Pacific Expedition headed by the young Franz Boas, in fact a cluster of a dozen ethnological expeditions to the two sides of the Bering Strait.

The first igloo Peary ever built was square. Overnight the wind ate away the angle where the roof and the end wall met, and Peary woke to snow in his face, abruptly tutored in the proper form of the Eskimo igloo. Peary climbed the rock wall over Umanak Harbor in 1896 to photograph the still, cold, berg-dotted waters of the bay, six kayaks small and intrepid beyond the foreground rocks, an evocative, lonely Arctic picture hard to rival. The Congo chapter closes with a photograph of the entire caravan of the expedition, after a forest march of 22 days, the bearers lined up 100 strong for the long shot, their head burdens lying before them. The expedition had engaged a total of 38,000 bearers during the 15,000-mile venture, most of the men serving on a march of only a few days. Knocked-down crates had been shipped out by the thousand, but the nails had rusted and the boards had been turned to dust by termites. The local smiths hammered new nails out of old musket barrels, and planks were sawn in plenty for transporting the 54 tons of the collections.

The Gobi, which stands as background for tales of Chingis Khan and his heirs, yielded to Manhattan the famous dinosaur eggs, as Greenland sent back the great iron meteorite and Siberia the shaman's kit. Here is the paleontologist Walter Granger at work on the Flaming Cliffs of Outer Mongolia, removing the eggs. Two of the 13 eggs yielded delicate embryo skeletons, identification beyond doubt. Others had found such eggs earlier, although none were known to science: in Stone Age human deposits not far away the investigators found bits of dinosaur eggshell worked into squares. The original discoverers of the remarkable fossils had considerable priority.

The shores of the Sea of Okhotsk disclosed a striking sight to Jochelson's lens in 1901. A village of the maritime Koryak had sacrificed four fine dogs, impaled high above the snow. What the people sought from destiny was a respite from measles, brought in 1900 to the country by the Russians. The plague was borne, so the people said, by spirits that "came running from the direction of the sunset in the guise of colts." Before the Russians came the Koryak did not know the horse any more than they knew the virus.

VIC-20

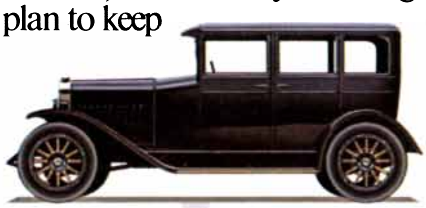
The Friendly Computer

$299.95

- Full Computer Keyboard
- 16 Colors/4 Sound Generators
- 66 Graphics Characters
- Memory Expansion to 32K RAM
- Includes Microsoft™/PET® BASIC
- Connects to TV or Monitor
- Recreation, Education & Personal Computing Programs
- Peripheral Equipment: Joystick, Paddles, Disk Drive, Printer, Cassette unit
- Telephone Modem for Communicating with outside Information Sources
- 4 Programmable Function Keys

For more information on the VIC-20's exciting
Space Games, Personal Software, and Programming capabilities,
visit your Commodore Dealer, or write:
VIC-20, Commodore Business Machines, 681 Moore Road, King of Prussia, PA 19406

commodore
COMPUTER



Established 1845 SCIENTIFIC AMERICAN March 1982 Volume 246 Number 3

Illegal Migration

The influx of foreign workers probably does not damage the economy, but migrants denied the protection of the law and rendered vulnerable by their illegal status may pose a threat to the integrity of U.S. society

by Charles B. Keely

It is generally agreed that illegal migration to the U.S. is a problem that should be solved by the implementation of a coherent Federal policy. There agreement ends. Each of the past four presidents has appointed a cabinet committee or task force to study the problem. Presidents Carter and Reagan both proposed measures intended to end illegal migration. No legislation has been passed by Congress, however, and no new Federal program has been created. As in other cases where demographic trends have become a matter of public concern, the making of coherent policy depends in part on the quality of the available data. The information on illegal migration is extraordinarily sketchy; in the absence of reliable information, conflicts among interest groups have paralyzed policymaking.

In the 1970's estimates of the number of illegal migrants ranged from two to 12 million. The disparity is so great that it renders the group of estimates useless. The most recent and soberest assessments indicate that there are between 3.5 and five million people living in the U.S. illegally, or very roughly 2 percent of the total population. The largest group consists of illegal migrants from Mexico, of whom there are probably between 1.5 and 2.5 million. Although the data are far from complete, the number of illegal entries per year appears to have risen sharply in the 1970's. One estimate that has been widely accepted (although it has no empirical basis) is that about 500,000 illegal residents per year are now being added to the U.S. population. The problem of illegal migration has recently been compounded by the arrival of large numbers of refugees from Cuba and Haiti. U.S. policy toward them has been as uncertain as it has been toward other illegal migrants.

Policy measures of three kinds have been proposed to deal with the problem of illegal migration. One approach is to make it harder for illegal residents to get jobs, which presumably would reduce the incentive for coming to the U.S. The second is to make it harder to enter the country illegally, primarily by increasing the staff and budget of the Immigration and Naturalization Service (INS). The third proposal is to offer amnesty to those who now live illegally in the U.S. and at the same time to adopt a sterner policy toward new illegal entrants. In order to choose rationally among these alternatives and others it is necessary to understand what effects illegal migration has on U.S. society. Where illegal migrants have settled, what jobs they hold and how they affect the employment of U.S. citizens are known with even less precision than the total number of illegal residents. Discussions of the issue have focused mainly on the competition for jobs, but the effects of illegal migrants are not uniformly harmful; indeed, it is probable that the economy as a whole does not suffer from their presence. On the other hand, certain noneconomic aspects of illegal migration are truly disturbing, and precisely because of its illegality. The existence of a large group denied the rights of citizenship may threaten the integrity of U.S. society.

The question of how many illegal residents there are in the U.S. first received substantial official attention in hearings held from 1971 to 1973 by the House Subcommittee on Illegal Immigration. In the hearings Raymond F. Farrell, then commissioner of the INS, estimated that there were slightly more than a million illegal residents. The total included both those who had entered the U.S. by illegal means and those who had been admitted properly but who had overstayed or violated the conditions of their visas. Illegal entrants from Mexico were estimated to make up 80 percent of the total.

Farrell's estimate was based on an extrapolation from the number of people apprehended by the INS for violating immigration laws. Such an extrapolation relies on an estimate of the ratio of those who are apprehended to those who are not; the ratio utilized was derived from little more than the intuition of INS officials. The report of the subcommittee, published in 1973, nonetheless maintained that there were between one and two million illegal migrants. The general conclusion of the report was that the number of illegal residents was large and, judging from apprehension statistics, increasing.

One of the groups with a strong interest in the number of illegal migrants is the INS itself. Increases in funding and manpower for the agency depend on public perception of illegal migration as an important problem. In the mid-1970's Leonard F. Chapman, the newly appointed commissioner of the INS, said many times that the flow of illegal migrants was increasing and that the agency was not equipped to stop it. At different times Chapman estimated that there were between four and 12 million people living in the U.S. illegally. When he was asked by Congress to make a more precise estimate to serve as the basis for allocating funds to the agency, he commissioned two studies.

The first study was done in 1975 by Lesko Associates, a consulting firm in Washington, D.C. The firm asked six people with expertise in the field of migration to estimate the size of the illegal population. The estimates ranged from 4.2 to 11 million; their average was 8.2 million. The method of the Lesko study was roundly criticized by demographers in the Bureau of the Census and by the Congressional Research Service. In response a second study was begun; it relied on estimates made by local INS officials of the number of illegal residents in their districts.

The second survey concluded that the total was about six million. The regional directors of the INS, who had been responsible for polling their subordinates to provide a total, stressed that the estimates were based on nothing more than guesses by those working in the field. In spite of their caution, the head of the Planning and Evaluation Division of the INS wrote in 1977: "We do have strong evidence that there are about six million illegal aliens in the country."

Although the number of illegal residents cannot be determined directly, demographers have applied statistical methods to the data that are available in an effort to make the estimates more precise. Clarise Lancaster and Frederick Scheuren of the Social Security Administration employed a technique that relies on comparing two counts of the U.S. population. One count is presumed to include illegal migrants and the other is presumed to exclude them. The difference between the counts yields an estimate of the illegal population.

In Lancaster and Scheuren's study the size of the population including illegal migrants came from Social Security files, from a list of individuals who have filed Federal income tax returns and from the Current Population Survey conducted by the Bureau of the Census in March, 1973. Illegal migrants are presumed to appear in such lists if they are employed. A mathematical technique was utilized to determine how many people were not represented in any of three lists; this number was added to the total. The size of the population excluding illegal migrants was extrapolated from the results of the 1970 census. Analyses of samples of census returns have shown that few illegal migrants were included in the 1970 count. Lancaster and Scheuren estimated that as of March, 1973, there were 3.9 million illegal migrants from age 18 to 44 living in the U.S. They added that their estimate was not precise; there might be as few as 2.9 million or as many as 5.7 million illegal residents.

J. Gregory Robinson of the Bureau of the Census compared death rates for various groups with their numbers as recorded in the 1970 census. Robinson presumed that few illegal residents had been counted in the census. Mortality statistics, on the other hand, which are based on death certificates, were presumed to include the illegal residents. An increase in the death rate for a particular group could therefore indicate the presence of illegal migrants. Robinson found such an increase among white males and estimated that in 1975 there were between three and four million white males from age 20 to 44 living in the U.S. illegally. He argued that there had been two waves of illegal migration: one to the Eastern states in the late 1960's and a second to the Southwest in the 1970's.

Robinson found no evidence for a substantial illegal population of blacks or of white females. Lancaster and Scheuren, however, had concluded that there was a sizable illegal black population. Moreover, their estimate included both men and women whereas Robinson had found only men; the rough agreement between the estimates is therefore largely fortuitous.

In 1979 Congress established the Select Commission on Immigration and Refugee Policy. The commission, made up of four senators, four congressmen, four cabinet members and four citizens-at-large, was chaired by the Reverend Theodore M. Hesburgh of the University of Notre Dame. At the request of the commission Robinson, Jacob Siegel and Jeffrey Passel, also of the Census Bureau, reviewed about a dozen estimates of the illegal migrant population. On the basis of this review they contended that there were almost certainly fewer than six million illegal residents in 1978. They "cautiously speculated" that the total was between 3.5 and five million. The number of illegal residents of Mexi-

SOME 700 CUBAN REFUGEES wait in an aircraft hangar at Eglin Air Force Base in Florida to be interviewed by representatives of the Immigration and Naturalization Service (INS). The photograph was made in May, 1980. After preliminary processing the refugees were moved to tents at the base until the INS decided which of them would be allowed to become permanent residents of the U.S. Although U.S. policy toward political refugees has traditionally been generous, President Carter did not allow the Cubans immediate refuge. About 132,000 Cuban and Haitian refugees have been given temporary status as "entrants." Some entrants have been resettled but many have spent long periods in camps such as the one at Eglin Air Force Base. Political refugees are a small minority of the 3.5 to five million illegal migrants living in the U.S. Most come to the U.S. because of higher wages. The largest single group consists of Mexicans.

can origin was then probably between 1.5 and 2.5 million.

A serious problem with most estimates of the illegal population is that they do not distinguish between illegal permanent residents and those who enter the U.S. illegally but leave shortly thereafter. The latter group is made up mostly of Mexicans; it includes seasonal workers in agriculture and also workers who take permanent jobs (in the garment industry, for example) but leave after less than a year's employment. Their crossing and recrossing of the border may cause substantial fluctuations in the number of people in the U.S. illegally. Moreover, since seasonal workers do not intend to live in the U.S., they present problems quite different from those of permanent residents.

DATE OF ESTIMATE	ESTIMATE (MILLIONS)	SOURCE	ILLEGAL-RESIDENT GROUP
1970	1.6	HOWARD GOLDBERG (GEORGETOWN UNIVERSITY)	ALL MEXICANS
1973	3.9	FREDERICK LANCASTER AND CLARISE SCHEUREN (SOCIAL SECURITY ADMINISTRATION)	ALL NATIONALITIES, AGES 18 TO 44
1975	8.2	LESKO ASSOCIATES	ALL NATIONALITIES, ALL AGES
1975	3–4	J. GREGORY ROBINSON (BUREAU OF THE CENSUS)	WHITE MALES, AGES 20 TO 44
1976	6	LEONARD F. CHAPMAN (INS COMMISSIONER)	ALL NATIONALITIES, ALL AGES
1977	0.5–1.2	MEXICAN GOVERNMENT	ALL MEXICANS
1977	0.7–2.2	BUREAU OF THE CENSUS	ALL MEXICANS
1978	3–6	LEONEL CASTILLO (INS COMMISSIONER)	ALL NATIONALITIES, ALL AGES

ESTIMATES OF ILLEGAL RESIDENTS in the U.S. made during the 1970's vary so widely that as a group they are useless for policymaking. The estimates are not directly comparable; they apply to different groups of migrants. Goldberg's estimate was based on Mexican census records. He found a deficit in the Mexican population, which he attributed to migration to the U.S. Lancaster and Scheuren compared Federal records (presumed to include employed illegal migrants) with data extrapolated from the 1970 census (presumed not to include them). The estimate made by the consulting firm of Lesko Associates was prepared at the request of the INS. It is an average of guesses made by experts. Robinson's estimate is based on death certificates (presumed to include migrants) together with 1970 census data. Robinson found that the death rate among white males as indicated by the death certificates and census data was abnormally high; because the illegal migrants were included in the death records but not in those of the census (which provided the population base for the death rate) the apparent increase was presumed to be due to their presence. Chapman's estimate was based on guesses made by INS officials in the field. The estimate by the Mexican government was based on a survey made at the border; reanalysis of the data by demographers at the Bureau of the Census gave a higher number. The estimate by Castillo is based on his assessment of other estimates. The information in the table comes from an article by Census Bureau demographers. After reviewing these estimates and others they concluded that there were probably between 3.5 and five million illegal residents of the U.S. in 1978, of whom between 1.5 and 2.5 million were Mexican.

How fast is the number of illegal residents growing? Answers to this question are even less firmly grounded than estimates of the total. Michael Teitelbaum, formerly staff director of the House Select Committee on Population, has estimated that about 500,000 illegal migrants take up residence in the U.S. each year and that the number is growing. A similar estimate appeared in a study of immigration by Leon Bouvier of the Population Reference Bureau and has been adopted in projections made by President Reagan's task force on illegal immigration. No empirical basis for the estimate was presented in any of these accounts. A survey conducted in 1975 by the INS at a few points on the Mexican border also concluded that there were about 500,000 successful illegal entries made in that year by those possessing fraudulent documents. Because the sample in that survey was very small, and because of the absence of empirical grounding for the other estimates, it is difficult to have much confidence in the number.

Although it is not possible to determine precisely how many people enter the U.S. illegally each year or how many of them take up permanent residence, the number of apprehensions by the INS of aliens without valid immigration documents provides an indication of the general trend. The apparent increase in the number of illegal entrants is closely related to changes in U.S. immigration law and policy. Indeed, the current flow of illegal migrants is in part a result of U.S. immigration policy. The number of natives of other countries who want to live or work in the U.S. may or may not have changed; changes in immigration policy, however, have greatly reduced the number of Latin Americans allowed to enter. In consequence many who might formerly have come to the U.S. legally now resort to illegal entry.

The history of U.S. immigration laws can be correlated with changes in the perceived needs of the labor market. Throughout much of the 19th century there were few limitations: the "huddled masses" were accepted, although individuals could be classified as undesirables and denied entry by reason of poverty, disease or criminality. In that era large numbers of Europeans and Asians came to the U.S., many of them finding work in developing heavy industries such as the railroads, mining and steelmaking. The first ethnic or national group whose immigration was curtailed were the Chinese, who were excluded in 1882. Over the next 50 years a mood of pronounced xenophobia entered American political life. By 1917 all immigrants were required to pass a literacy test, and in the 1920's a quota system was introduced. Immigration from countries outside the Western Hemisphere was limited to 150,000 people per year, allocated to the various countries in proportion to the ethnic composition of the U.S. population in 1920; most of those allowed in under this system were Europeans. Asians and Africans were largely excluded.

Until quite recently migration from the countries of the Western Hemisphere was governed by a system of regulations separate from the quotas that applied to immigrants coming from Europe, Asia and Africa. Nevertheless, the relation between labor needs and immigration policy is clearly evident, and it is particularly marked in the case of Mexican migrants. Early in the 1900's Mexican workers were recruited for the mines and railroads of the Southwest. The enforcement of immigration regulations was then quite informal, and crossing the border was a simple matter. During World War I increased labor needs in the U.S. drew more Mexican workers into the Southwestern states. In the 1920's, in accord with the general revision of immigration laws, enforcement was made stricter in an attempt to reduce the flow of Mexicans. In the

1930's a concerted effort was made to repatriate Mexicans from the Southwest because of unemployment in the U.S.

During World War II, however, official policy was reversed because of the labor shortage caused by the war. Beginning in 1942 U.S. employers were encouraged to hire Mexican workers to do seasonal agricultural work. The terms of employment were spelled out by a series of agreements between the governments of the two countries; the agreements were collectively known as the Bracero Program, from the Spanish word for agricultural laborer. By the late 1950's about 400,000 Mexican temporary workers were entering the U.S. each year under the program.

Living conditions provided for the braceros by U.S. employers were poor. A series of journalistic accounts exposed the conditions; notable among them was a television documentary titled "Harvest of Shame," produced by Edward R. Murrow. Moreover, the presence of the Mexican laborers had the effect of lowering wages for indigenous farm workers and impeding their efforts to unionize. In 1964, after criticism from labor, church and ethnic groups, Congress ended the program.

At about the same time there was a major change in U.S. immigration policy. Amendments to the immigration law in 1965 abolished national quotas. In their place limits were established on the number of people in various categories (such as spouses and children of previously admitted immigrants, and adult children and brothers and sisters of U.S. citizens) who could enter the U.S. Each category was assigned a priority. What is more, for the first time a limit was set on the number of people who would be admitted to the U.S. from the other countries of the Western Hemisphere. Under the new regulations no limit was placed on the entry of spouses, parents and children of those who had attained U.S. citizenship. Only 120,000 visas, however, were reserved for new immigrants from the Western Hemisphere without immediate relatives who were U.S. citizens. New immigrants were generally required to prove that they had a job offer at the wage rate prevailing in the area where they intended to live.

The new policy made it much more difficult for Mexicans and other Latin Americans without close relatives who were already U.S. citizens to immigrate legally. For Mexicans the possibility of entering as temporary workers ended with the Bracero Program. The result was a large immediate increase in illegal migration, as reflected in the number of people apprehended by the INS.

NUMBER OF MIGRANTS TAKEN INTO CUSTODY by the INS for lack of valid immigration documents shows the influence of Federal policy on illegal migration. Beginning in 1942 a series of agreements with the Mexican government known as the Bracero Program allowed Mexican workers to enter the Southwest to pick crops. In 1953 and 1954 an attempt was made to apprehend all illegal agricultural workers, return them to Mexico and enter them in the official program; about two million were returned in the two years. After that the number of apprehensions fell rapidly and remained small until 1964, when the Bracero Program was ended by Congress. In 1965 limits were imposed for the first time on immigration from countries of the Western Hemisphere. The continuing economic pressures on natives of those countries to leave, coupled with limitations on legal entry, caused a sharp increase in illegal migration as shown by apprehension statistics. One estimate that has been widely accepted (although it has no empirical basis) is that 500,000 illegal migrants per year are added to the U.S. population.

In 1953 and 1954, while the Bracero Program was in effect, an effort had been made to return illegal residents of the Southwestern states to Mexico and to enter them in the official program. About two million workers were returned in the two years. Thereafter the number of apprehensions fell to about 40,000 per year, then it again rose slightly. Between 1964 and 1970, however, the number increased from 87,000 to 345,000. By 1977 about a million apprehensions were made per year for violations of immigration law. The influence of the changes in policy is suggested by the fact that the proportion of Mexicans among those who were apprehended rose from 51 percent in 1964 to 92 percent in 1977.

It is notable that the increase in the proportion of Mexicans among illegal migrants in the 1960's and 1970's was accompanied by a long-term increase in the proportion of legal immigrants coming from Latin-American countries. Until the 1930's most legal immigrants were Europeans. Now the two largest groups of legal immigrants are Latin Americans and Asians, who make up about 80 percent of the total. The 1965 changes in the immigration laws made it more difficult for Latin Americans to enter the U.S. legally unless they have relatives here. As a result the proportion of Latin Americans among legal immigrants has not risen as rapidly since 1965 as it did before; the proportion of Asians has increased faster in compensation. The statistics for both legal and illegal immigrants, however, suggest the strong economic and political pressures that impel Latin Americans to migrate to the U.S.

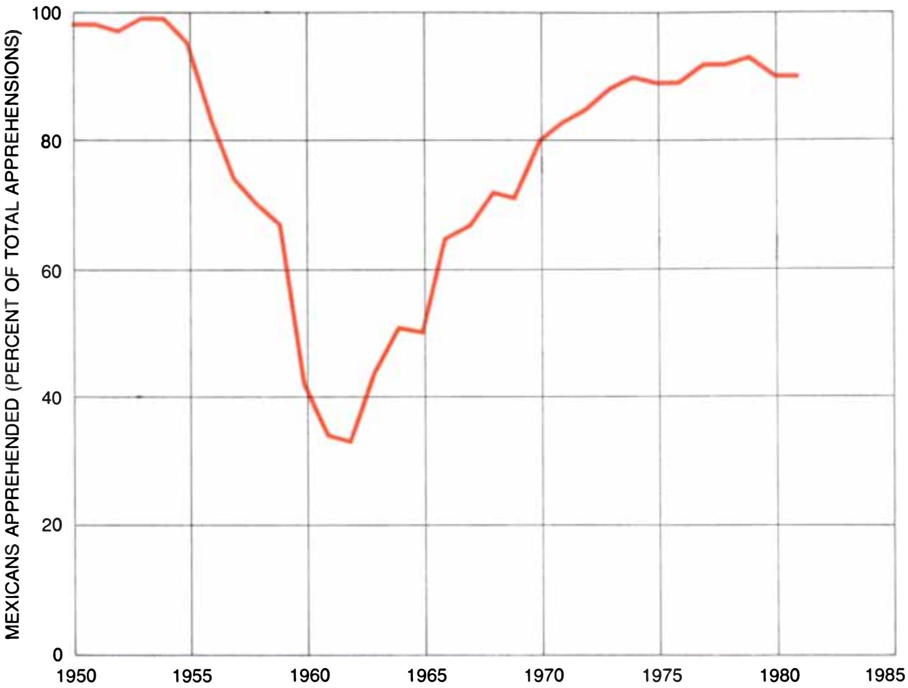
The recent arrival of a large number of refugees has contributed to the increase in the proportion of immigrants from Asia and Latin America. Since 1975 about 550,000 Indochinese refugees have been admitted. Their admission is continuing: under current plans about 100,000 will be allowed to enter this year. Between 1965 and 1980 about 450,000 Cubans were admitted as political refugees. The "Mariel boat lift," an exodus permitted by the Cuban government, swelled the ranks of the Cuban migrants. In addition many Haitians have sought refuge in the past three years; more recently refugees from El Salvador have begun to seek asylum in this country.

The U.S. has traditionally adopted a generous official attitude toward political refugees. The weakening economy, fear of illegal migration and the suspicion that recent refugees have left their countries for economic rather than political reasons, however, have combined to yield a skeptical official attitude. Instead of granting the Cubans and Haitians refugee status immediately, President Carter created a new category, that of entrant; 132,000 Cubans and Haitians were given the new status. An entrant has only a temporary status; entrants do not have the privileges of refugees. While some entrants have been resettled, the indecisiveness of official policy has led to many being held for long periods in camps designed for temporary occupancy; considerable resentment and frustration have resulted.

What is the effect of illegal migration on U.S. society? The consequences appear in at least three areas: employment, public services and national cohesion. The most direct effect is that on the employment of indigenous workers. Illegal migrants probably do displace some U.S. citizens from jobs. Labor economists do not take seriously, however, the hypothesis that each illegal resident who is hired takes a job from a citizen. Without reliable knowledge of the cities and economic regions where illegal migrants have settled it is not possible to determine how many workers are displaced or what industries are most affected.

A more serious adverse consequence is that the presence of illegal migrants in the work force holds down wages and impedes efforts to improve working conditions. Illegal residents compose a docile segment of the working population, and one that has proved to be difficult to organize. One reason for the docility, of course, is the fear of deportation, which makes the illegal worker vulnerable to intimidation and cuts him off from governmental means of redress. In addition many migrants intend to work in the U.S. for a short time and then return to their own countries with their savings. Hence they feel they have no stake in the improvement of working conditions in the U.S.

The economic effects of illegal migration are not all deleterious, however. Illegal migrants can help to revitalize failing industries; they have done so in the garment industry in New York and Los Angeles. To say this is not to condone the exploitation of such workers. The inhumanities of the sweatshop system, which can include not only low wages but also overlong hours of work, unsafe conditions and reliance on the labor of children, have been known for at least a century. Nevertheless, when illegal migrants work in an industry that would collapse without them, they are not competing for jobs with citizens. The availability of cheap labor can also make it possible for new businesses to be established, as has happened in the case of nursing homes and ethnic restaurants in the Eastern states. Again the jobs held by the migrants would not exist without them. Moreover, the expansion of the economy can create jobs in ancillary industries such as trucking, food services and accounting; by this means the total number of jobs available to U.S. citizens may actually be increased as a result of illegal migration.



PERCENTAGE OF MEXICANS among those apprehended by the INS reflects both the end of the Bracero Program and the change in the immigration law in 1965. At the height of the Bracero Program about 400,000 Mexican agricultural workers were admitted per year. When the program ended, some continued to enter illegally. When the immigration law was changed, 120,000 visas were reserved for new immigrants from the Western Hemisphere. Under these rules it became much more difficult for Mexicans to immigrate legally unless they had relatives in the U.S. The result of the concurrent changes was that the proportion of Mexicans among illegal migrants who were apprehended increased sharply. The increase came at at time when the total number of illegal migrants, as indicated by apprehension statistics, was rising.

One way of evaluating the effect of illegal migrants on the economy is to imagine they were all to be deported immediately. It is not clear that if this happened there would be an increase in the number of jobs available to indigenous workers. Some companies, deprived of cheap labor, would close; others would move their plants overseas; still others would substitute machinery for human labor, which was the response of farmers in the Southwest at the end of the Bracero Program. Of course some companies would hire citizens to replace the lost workers, thereby making jobs available, but it is not known how many illegal migrants are employed in such companies. Tending to compensate for such gains would be the loss of employment in the ancillary industries as the economy contracted.

In the absence of such a large-scale social experiment the data are scanty, but on balance it seems likely that the U.S. economy as a whole is not harmed by illegal migration. Instead there is a redistribution of costs and benefits. Those who benefit are employers, consumers and workers in industries related to those where migrants work; U.S. citizens who lose their jobs, or whose wages and working conditions deteriorate, suffer. These victims are likely to be concentrated among groups that are already placed at an economic disadvantage. What of the migrants themselves? One might say they have voted with their feet by coming here. Presumably they believe they are better off economically than they would have been if they had stayed home. Precisely because of their illegal status, however, their views are rarely represented in the deliberations that determine their fate.

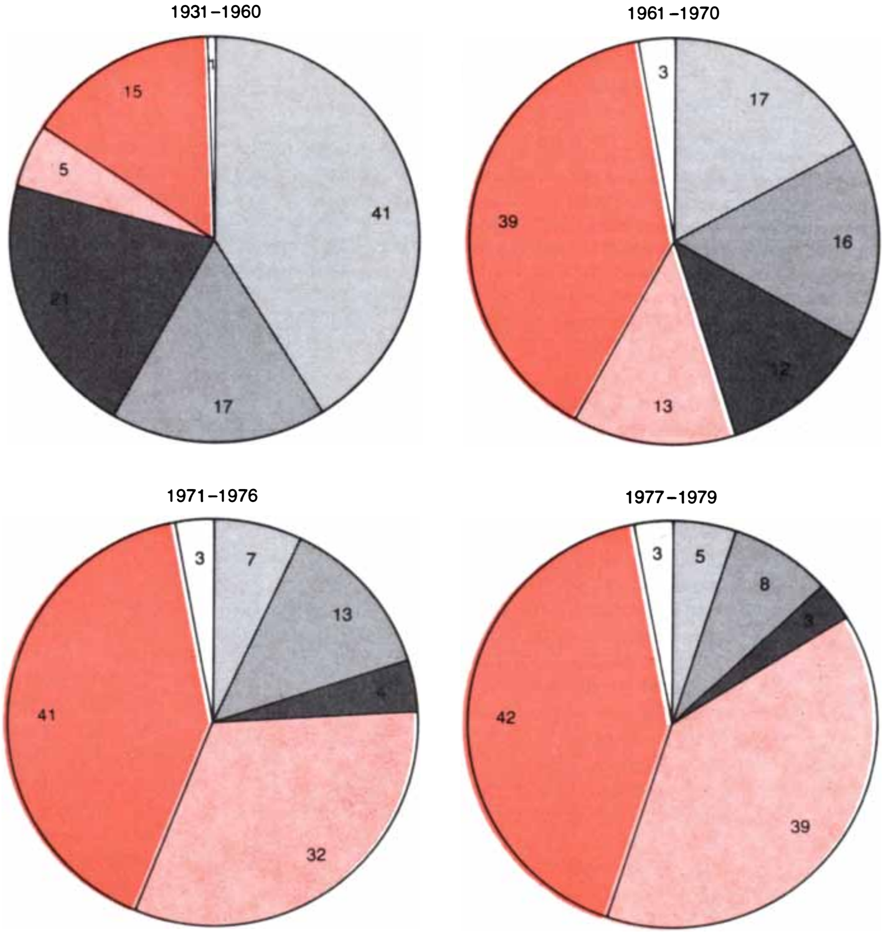
In illegal migrants' use of social services there is a similar redistribution of advantages and disadvantages. Many studies have found that illegal migrants pay more in Federal taxes than they consume in government services. From two-thirds to three-fourths of all illegal residents pay the Federal witholding tax on income and contribute to the Social Security fund. Perhaps only a few percent of them make direct demands on services subsidized by the Federal Government. For example, the illegal resident is not likely to apply for welfare or Social Security benefits. In some recent studies, however, there is an indication that illegal migrants are making increased use of public services as families are reunited and the proportion of young males is reduced.

In some parts of the U.S., however, the effect of illegal migration on certain government services may be quite disadvantageous. Illegal residents are most likely to pay Federal taxes, because these are deducted from payroll checks; the services they are most likely to utilize, on the other hand, are those financed by state and local taxes. Public schools and hospitals in the Southwest, for example, may be overburdened by illegal residents whose payment of local taxes is outweighed by their use of services financed by such taxes.

From the foregoing it appears that the U.S. economy as a whole is not greatly damaged by the activities of illegal migrants and may even benefit from them, although the costs and benefits are unevenly distributed. The balance of gains and losses for the migrants themselves remains an open question, but there is little doubt that the economies of their home countries benefit. It would be shortsighted, however, to consider the movements of people across national borders an economic transaction only.

For many U.S. citizens the existence of an underclass of exploited workers within North American society is deeply troubling, regardless of whether those workers burden the economy or contribute to it. To eliminate the underclass by deporting or excluding illegal foreign workers, and thereby imposing further hardships on them, is hardly a humane or practical solution. It is often argued that the richer nations of the world have an obligation to the poorer ones, and the bond is notably strong in the case of close neighbors such as the U.S. and Mexico. It has been suggested that allowing migrant workers access to jobs in the U.S. (and treating them fairly while they are here) may be an effective way to meet a part of that obligation; the costs and benefits of such a solution, however, are far from clear. As a practical matter, immigration policy cannot be considered apart from foreign policy; in formulating it the sensibilities of other nations must be taken into account.

Even if illegal migration is considered exclusively in terms of its effects on U.S. society, what is likely to prove most disruptive is not the presence of the migrants but their status as illegal aliens. Some of the things valued most highly in U.S. society are personal freedom, equality of economic opportunity and cultural tolerance. To some degree social cohesion in this country depends on the extent to which people not only respect these values but also enjoy the benefits of them. The existence of a large group of illegal residents who are denied such benefits seems a formula for social and political disaster. The fact that many of the illegal migrants are members of a single ethnic group (Hispanics) and are concentrated in one region (the Southwest) intensifies the problem. Societies have tremendous resilience, but to ignore a situation that threatens the basic values of a political culture is to tempt fate.



LEGAL IMMIGRANTS' geographic origins have shifted sharply toward Latin America and Asia in recent decades. Most 19th-century immigrants to the U.S. were from northern and western Europe. Those from southern and eastern Europe predominated between 1900 and 1930. During the 1930's and 1940's European political refugees made up a large portion of immigrants. In the 1920's immigration quotas were established for each country of Europe. Asians and Africans were for the most part excluded; immigrants from the Western Hemisphere, however, were exempt from the national quotas. Hence the quota system discouraged immigration from Asia and encouraged it from Latin America. When the quotas were abolished in 1965 in favor of preferences unrelated to nationality, the proportion of Asians was increased and that of Latin Americans reduced. Large numbers of refugees have contributed to the increase in the proportion of immigrants from Asia or Latin America. In 1980 the two groups made up about 80 percent of the 800,000 legal immigrants.

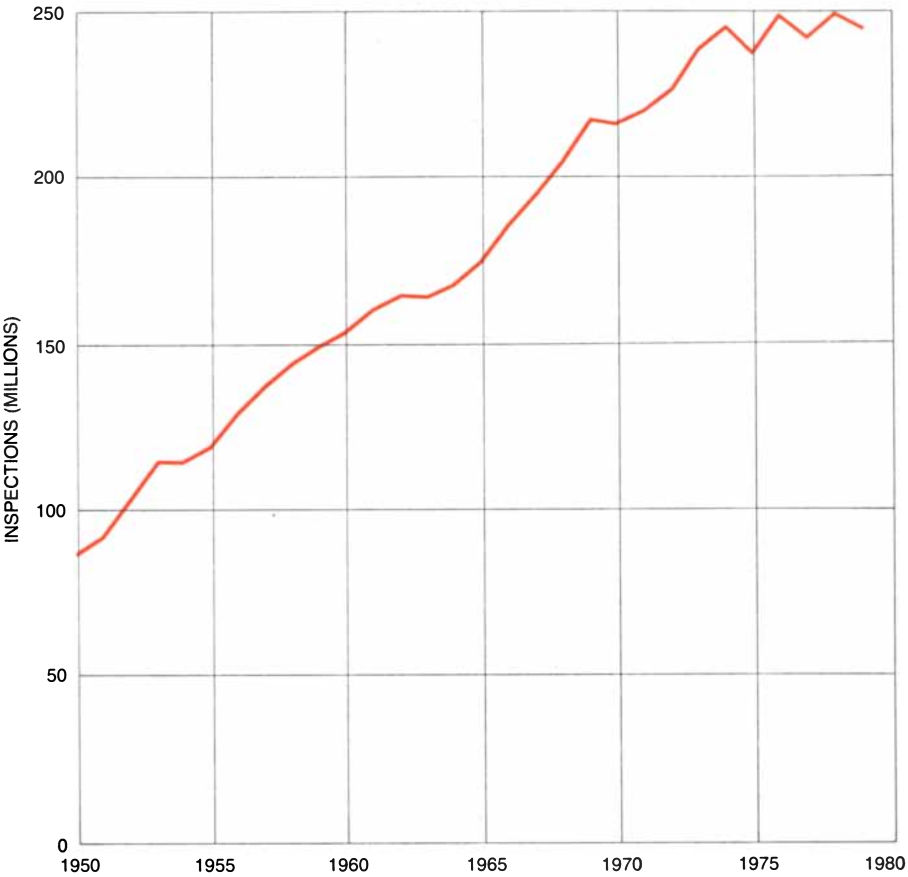
How should the problem of illegal migration be solved? The one sure solution is to establish a more equitable distribution of wealth among nations. Such a fundamental change will clearly be slow in coming. In the meantime there is strong feeling within the U.S. that some remedial measures are needed. The proposals that have been recently considered have been aimed mainly at methods of excluding new migrants; concessions might be granted to those already in the country illegally.

One proposal is to make it harder for migrants to get and hold jobs. Higher wages in the U.S. are presumed to attract most migrants. If access to jobs could be controlled, the motivation to enter the U.S. illegally might be reduced. The immigration law makes it a felony to transport or harbor an illegal migrant, but employing one is specifically excluded from the category of acts subject to prosecution. Because of the exception, which is called the Texas Proviso, it is against the law for an illegal migrant to take a job but it is legal for his employer to hire him. Although several bills designed to eliminate the Texas Proviso have been introduced into Congress, none has passed.

President Carter, President Reagan and the Select Commission on Immigration and Refugee Policy all advised the adoption of sanctions on hiring. No single proposal, however, has elicited wide support. A major difficulty is that in order to enforce sanctions a reliable method of identifying legal workers must be available to employers. A work-authorization card and a system that enables an employer to check the authorization number by telephone have been suggested. The idea of a national identification card or internal passport is not likely to be well received in the U.S., and so other proposals would rely on existing documents. The Social Security numbering system might serve if the procedure for issuing numbers were made more rigorous. The measures offered by the Carter and Reagan administrations call for identifying legal workers by means of at least two existing documents.

One incentive for employers to hire illegal migrants is that the migrants will accept lower wages and poorer health and safety conditions than are allowed by law. It has been said that if current laws on working conditions were more strictly enforced, the motivation to hire illegal residents would be reduced.

The second proposal to limit illegal migration concerns the INS. If the nation is indeed determined to control illegal migration, the budget and personnel of the INS are widely seen to be inadequate to meet its responsibilities. From 1970 through 1978 admissions of holders of temporary visas more than tripled; apprehensions of illegal migrants nearly quadrupled; man-hours devoted to enforcement by the INS, on the other hand, only slightly more than doubled. The agency now makes about 250 million inspections of people entering the U.S. at land borders and airports each year, a total larger than the U.S. population. Almost every official group that has studied illegal migration has recommended that the budget of the INS be increased and that the resources of the agency be better managed.



NUMBER OF INSPECTIONS made by INS agents at the Canadian and Mexican borders and at airports has increased greatly since 1950. The agency now makes about 250 million inspections per year. The number of inspections has increased much faster than the staff or budget of the agency. Almost every official body that has studied illegal migration has advocated increasing the staff of the INS. Most official consideration of the issue has focused on stopping the flow of illegal migrants by stricter enforcement of immigration regulations or by making it harder for migrants to get jobs. Neither solution, however, is aimed at the underlying cause of illegal migration: the large difference in wealth between the U.S. and the developing countries.

If some means of ending the flow of illegal migrants were devised and adopted, what would become of those already living here illegally? Most official studies have concluded that mass deportation is neither feasible nor humane. The alternative is the granting of amnesty to illegal residents. Presidents Carter and Reagan have advocated programs in which two kinds of status could be conferred: permanent residency for those who have lived for a long period in the U.S. and have developed strong ties here, and temporary status for those who have been living here for a shorter time. Those who were given temporary status might later be allowed to become permanent residents.

In President Reagan's proposal those who entered before 1970 would be eligible for permanent residency. Those who entered in the period from 1970 through 1979 would be eligible for three-year visas that could be renewed and would allow employment but would not allow the migrant to bring his family to the U.S. After 10 years of residence the migrant could apply for immigrant status. Those who entered in 1980 or later would be deported. The Select Commission proposed a simpler amnesty program. Those living in the U.S. on January 1, 1980, who had been residents for a period to be determined by Congress would be eligible for immigrant status. All others would be deported.

For a program of amnesty to succeed its terms must be attractive enough to encourage a large majority of the illegal residents to register. The experience of countries that tried amnesty programs in the 1970's, including Canada, France and Great Britain, is not encouraging; the number of migrants who registered was far smaller than the number estimated to be living in the country. Some contend that the terms of the Reagan amnesty are so rigorous that they offer little incentive to register.

The Reagan Administration has acknowledged that migrant workers do have a legitimate place in the U.S. economy. It has been recognized that there is a need for them to do seasonal work; they are also needed to do permanent work that is difficult and not well paid and to take jobs in regions where there is a labor shortage. Some maintain that a temporary-worker program should be created to meet these needs and to help control the flow of illegal migrants by channeling them into an official program. President Reagan has proposed a small experimental program: for two years 50,000 people per year would be allowed to take jobs in states identified by the Department of Labor as having labor shortages.

The Select Commission, however, recommended against creating any new temporary-worker programs. The committee also argued against expanding the one program now in existence, designated H-2 and administered by the Department of Justice, under which about 25,000 workers per year, mainly from various Caribbean countries, are admitted to Eastern states to pick crops. The committee cited the problems of overseeing a large temporary-worker program, the dangers of migrants staying in the U.S. illegally and the effect of temporary workers in forestalling the mechanization and wage increases needed to make the work attractive to citizens.

Illegal migration is a complicated problem. Its solution will require coordinated policy measures. Moreover, the experience with migration from Mexico proves that making entry illegal does not stop it. The economic, political and family pressures that cause migration will have to be considered if the problem is to be solved.

The Search for Intermediate Vector Bosons

In theory these massive elementary particles are required to serve as the carriers of the weak nuclear force. They should be detected soon in the aftermath of collisions between protons and antiprotons

by David B. Cline, Carlo Rubbia and Simon van der Meer

One of the major achievements of modern physics has been the development over the past 15 years or so of a new class of unified theories to describe the forces acting between elementary particles. Until these theories were introduced the four observable forces of nature seemed to be quite independent of one another. The electromagnetic force governs the interactions of electrically charged particles; the weak nuclear force is responsible for such processes as the beta decay of a radioactive atomic nucleus; the strong nuclear force holds the nucleus together, and gravity holds the universe together. The most successful of the new theories establishes a link between electromagnetism and the weak force, suggesting that they are merely different manifestations of a single underlying force.

The unified electroweak theory is about to be put to a decisive experimental test. A crucial prediction of the theory is the existence of three massive particles called intermediate vector bosons (also known as weakons). The world's first particle accelerator with enough energy to create such particles has recently been completed at the European Organization for Nuclear Research (CERN) in Geneva. The accelerator was originally built to drive high-energy protons against a fixed target, but it has now been converted to a new mode of operation, in which protons and antiprotons collide head on. The colliding-beam machine has just completed its preliminary runs and is scheduled to resume operating next month with an elaborate new particle-detection system in place. If the intermediate vector bosons exist and if they have the properties attributed to them by the electroweak theory, they should be detected soon. They are currently the most prized trophies in all physics, and their discovery would culminate a search that began more than 40 years ago.

According to the prevailing view of the interactions of elementary particles, a force is transmitted between two particles by the exchange of a third, intermediary particle. Such a description is the essence of a quantum field theory. The idea of a field extending throughout space is needed to explain how particles can act on one another at a distance; it is a quantum field because it is embodied in discrete units, namely the intermediary particles. In electromagnetic and weak interactions the exchanged particle is a member of the family called the vector bosons. This term refers to a classification of particles according to one of their most basic properties: spin angular momentum. A boson (named for the Indian physicist S. N. Bose) is a particle whose spin, when measured in fundamental units, is an integer, such as 0, 1 or 2. "Vector" designates a boson whose spin value is equal to 1.

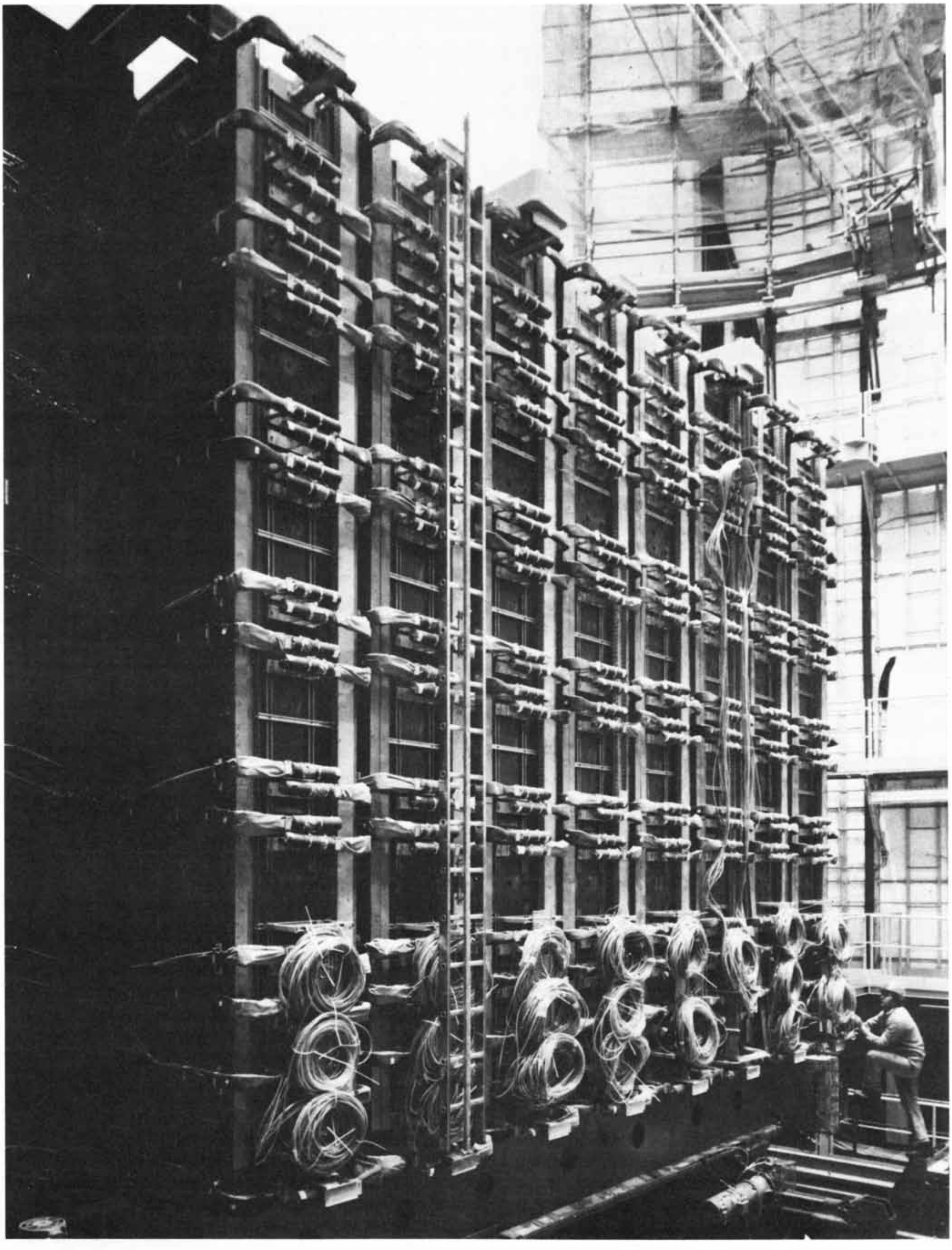
In the case of electromagnetism the exchanged vector boson is the photon, the massless and chargeless "wave packet" of electromagnetic energy that functions as the quantum of the electromagnetic field. Photons are easy to observe experimentally (in the form of light, for example), and from the study of their properties physicists have constructed the remarkably precise and comprehensive theory called quantum electrodynamics, or QED, which is the quantum field theory of the electromagnetic force.

The corresponding force carrier in weak interactions is the intermediate vector boson (intermediate simply because of its mediating role between particles). The existence of such a particle was first suggested in 1935 by the Japanese physicist Hideki Yukawa, who at the time was seeking a unified explanation of the two newly discovered nuclear forces: the strong and the weak. Yukawa noted that the range of a force should be inversely proportional to the mass of the particle that transmits it. For example, the range of the electromagnetic force is infinite, in accord with the masslessness of the photon. On the other hand, the two nuclear forces have only a limited range, and so Yukawa suggested they should be carried by particles with mass.

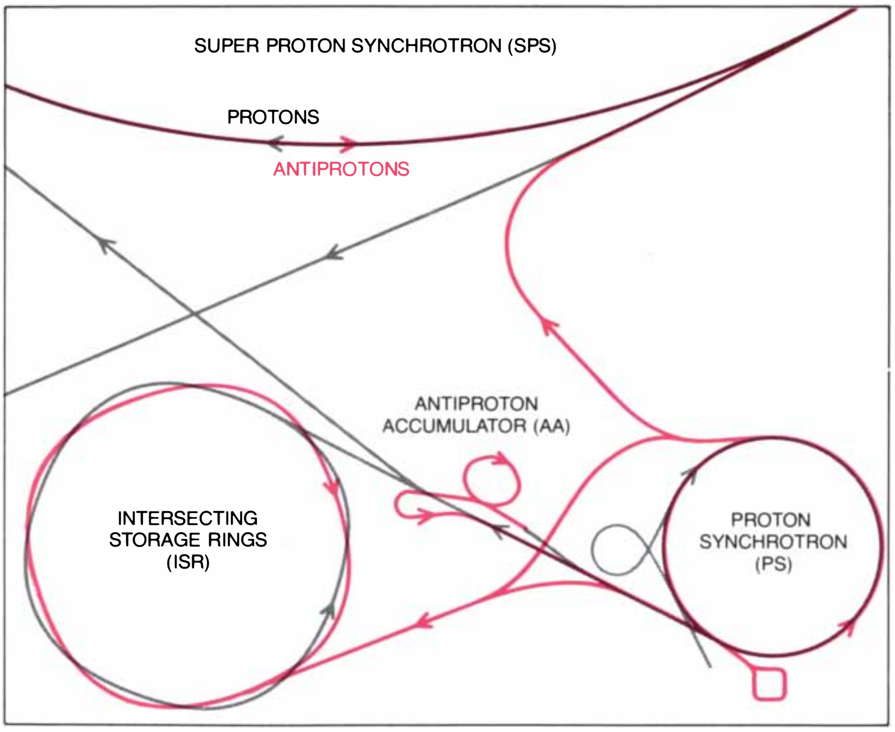
Specifically, Yukawa postulated the existence of a moderately heavy particle, later named the meson, the exchange of which gives rise to the strong attractive force between the proton and the neutron. The first particle of this type to be correctly identified, the pi meson (or pion), was discovered in 1947 in the shower of secondary particles generated by the collision of a cosmic-ray particle with an atom in the atmosphere; large numbers of mesons can now be made at will with particle accelerators.

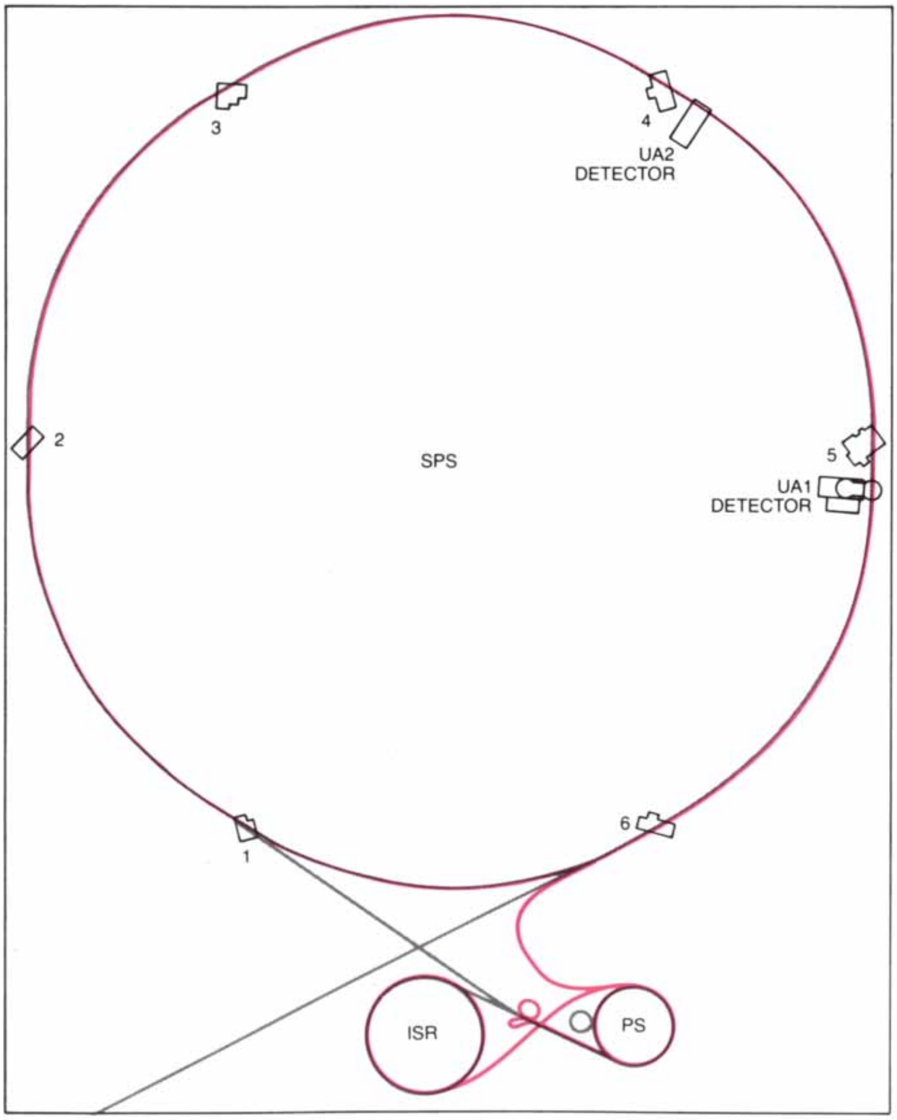
The particles of the nucleus, including the mesons, are now thought to be composed of the more fundamental constituents known as quarks. The quarks are bound together by the strong force, but in this context the force has a form quite different from the one observed between protons and neutrons. The force between quarks is thought to be transmitted by the family of eight massless vector bosons called gluons. A property with the arbitrary name "color" is assigned to the quarks and the gluons and plays the same role in strong interactions as electric charge does in electromagnetic interactions. In recognition of this analogy the quantum field theory of the strong force is called quantum chromodynamics, or QCD.

The weak nuclear force has a still shorter range than the strong force that acts between protons and neutrons. The intermediate vector bosons of the weak force can therefore be expected to have a mass larger than that of the pi meson. Early attempts to detect the intermediary particles associated with the weak force were unsuccessful, presumably because the bosons' larger mass put them out of reach of the existing particle accelerators. Until the advent of the uni-



NEW PARTICLE DETECTOR at the European Organization for Nuclear Research (CERN) in Geneva was designed and built by a team of more than 100 physicists from 11 institutions in Europe and the U.S. The first observation of an intermediate vector boson is expected to be made with the aid of this device. The large, multipurpose detector, named the UA1 (for Underground Area One), is seen here in its "garage" adjacent to the largest particle accelerator at the CERN site: the Super Proton Synchrotron (SPS), which was recently converted into a proton-antiproton colliding-beam machine. When the detector is ready for operation, it is rolled to the left on rails into the path of the colliding beams. Electronic equipment covers the outside of the apparatus, obscuring the central detection chambers.





COUNTERROTATING BEAMS of protons and antiprotons are generated at CERN as the end products of a sequence of events in several interconnected accelerator rings. First a beam of protons (*gray*) is accelerated to an energy of 26 GeV (billion electron volts) in the Proton Synchrotron (PS). The protons are then directed at a metal target, producing (among other things) a few antiprotons with an energy of 3.5 GeV. The antiprotons (*color*) are collected and transferred to the Antiproton Accumulator (AA), where they are combined with previously injected antiprotons and concentrated in dense "bunches" numbering several hundred billion particles each. The antiproton bunches are next sent back to the PS ring, where they are accelerated to 26 GeV. The 26-GeV antiprotons are injected into the SPS ring, where protons of the same energy are already circulating in the opposite direction. The two beams are finally accelerated to 270 GeV each in the larger ring. The overall site plan (*bottom*) shows the location of the new particle detectors, which are placed in two long straight sections of the SPS ring where the counterrotating beams are brought into collision. The other rings shown are used for a variety of other experiments.

fied electroweak theory in the late 1960's and early 1970's, however, there was no good estimate of the mass of the weak-force particles.

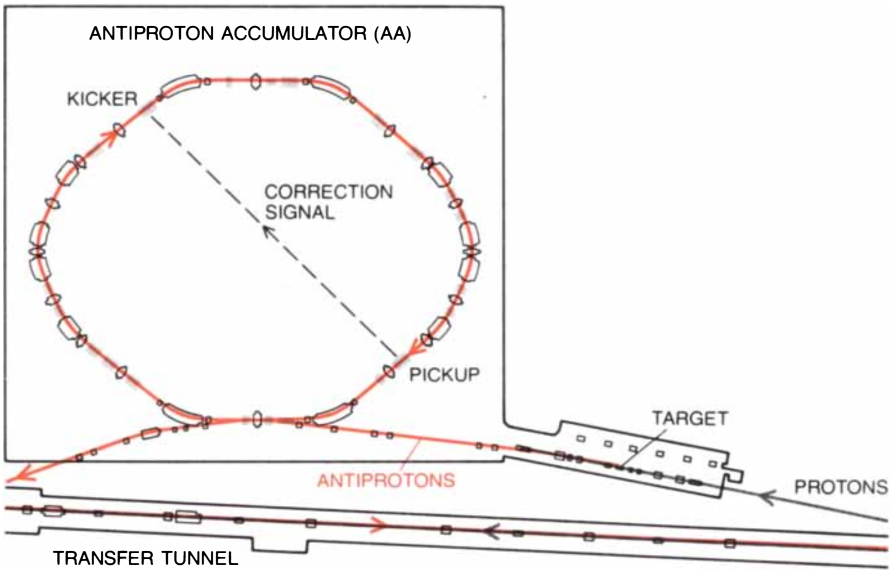
The electroweak theory was developed independently by Steven Weinberg of Harvard University and Abdus Salam of the International Center for Theoretical Physics in Trieste, with major contributions by Sheldon Lee Glashow of Harvard and others. The theory, which can now be considered the "standard" account of electromagnetic and weak interactions, for the first time made specific and testable predictions about the properties of the intermediate vector bosons, including their mass. Furthermore, the theory required that there be three such particles, with electric charges of +1 (the W^+), −1 (the W^-) and zero (the Z^0). The best present estimate of the mass of the intermediate vector bosons, expressed in terms of their equivalent energy, is 79.5 GeV for the W^+ and W^- and 90 GeV for the Z^0. (The abbreviation GeV stands for giga-electron volts, or billions of electron volts; for comparison the mass of the proton is equivalent to a little less than one GeV.)

The idea at the heart of the standard theory is that electromagnetism and the weak force both stem from a single and more fundamental property of nature. At exceedingly high energy (high enough for W and Z particles to be made as readily as photons) events mediated by the two forces would be indistinguishable. This theoretical unification is accomplished by assigning the photon and the intermediate vector bosons to the same family of four particles. At energies accessible today there is no question that electromagnetic events are quite different from weak ones; moreover, the photon and the W and Z particles seem to be unlikely siblings, since the first is massless and the other three are among the heaviest particles supposed to exist. The discrepancy is explained in the standard theory by the notion of a broken symmetry, which distinguishes between the forces as the energy is lowered in much the same way as a substance separates into different phases as the temperature is lowered.

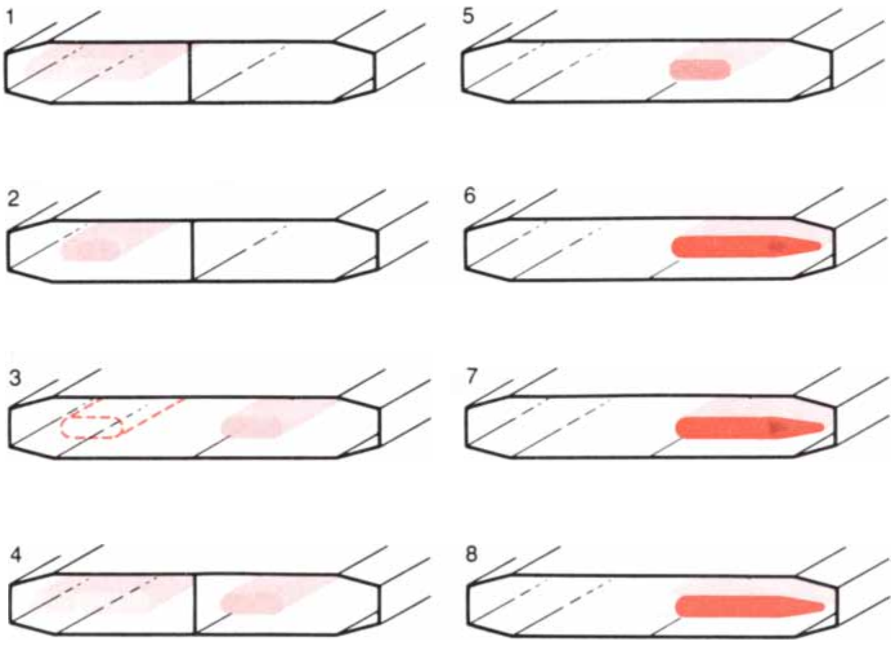
One approach to understanding the unified electroweak theory begins with an imaginary primordial state in which the photon and the intermediate vector bosons were all equally massless. It was the breaking of a symmetry of nature that endowed the W^+, the W^- and the Z^0 with large masses while leaving the photon massless. A mechanism for discriminating in this way among the carriers of the forces was first discussed in 1964 by Peter Higgs of the University of Edinburgh. Curiously, the Higgs mechanism is able to supply masses for the W and Z particles only by postulating still another massive particle, which has come to be called the Higgs boson. It is being sought along with the intermediate vector bosons.

The electroweak theory received important experimental support in 1973 from a discovery made at CERN and at the Fermi National Accelerator Laboratory (Fermilab) near Chicago. Up to then all known weak interactions of matter entailed an exchange of electric charge. For example, a proton might give up its charge of +1 to a neutrino (a massless particle with no charge). As a result the proton becomes a neutron and the neutrino is converted into a positron, or antielectron. All such events can be accounted for by the exchange of the charged intermediate vector bosons W^+ and W^-. The 1973 experiments revealed weak interactions in which the particles maintain the same charges they had before the event, as they do in electromagnetic interactions. A weak interaction of this type can be explained only by the exchange of a neutral intermediate vector boson (the Z^0 particle) or, in an equivalent description, by the operation of a neutral weak current [see "The Detection of Neutral Weak Currents," by David B. Cline, Alfred K. Mann and Carlo Rubbia; SCIENTIFIC AMERICAN, December, 1974]. In 1979 Weinberg, Salam and Glashow were awarded the Nobel prize in physics "for their contributions to the theory of the unified weak and electromagnetic interaction between elementary particles, including ... the prediction of the weak neutral current."

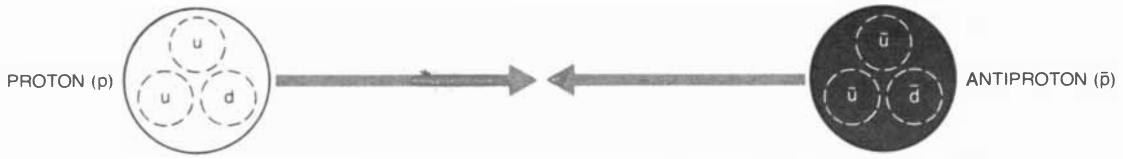
Once the existence of neutral weak currents had been fully confirmed it was only natural to try to find a way to detect the Z^0 as well as the W^+ and W^-. The task of creating particles with such a large mass, however, remained daunt-



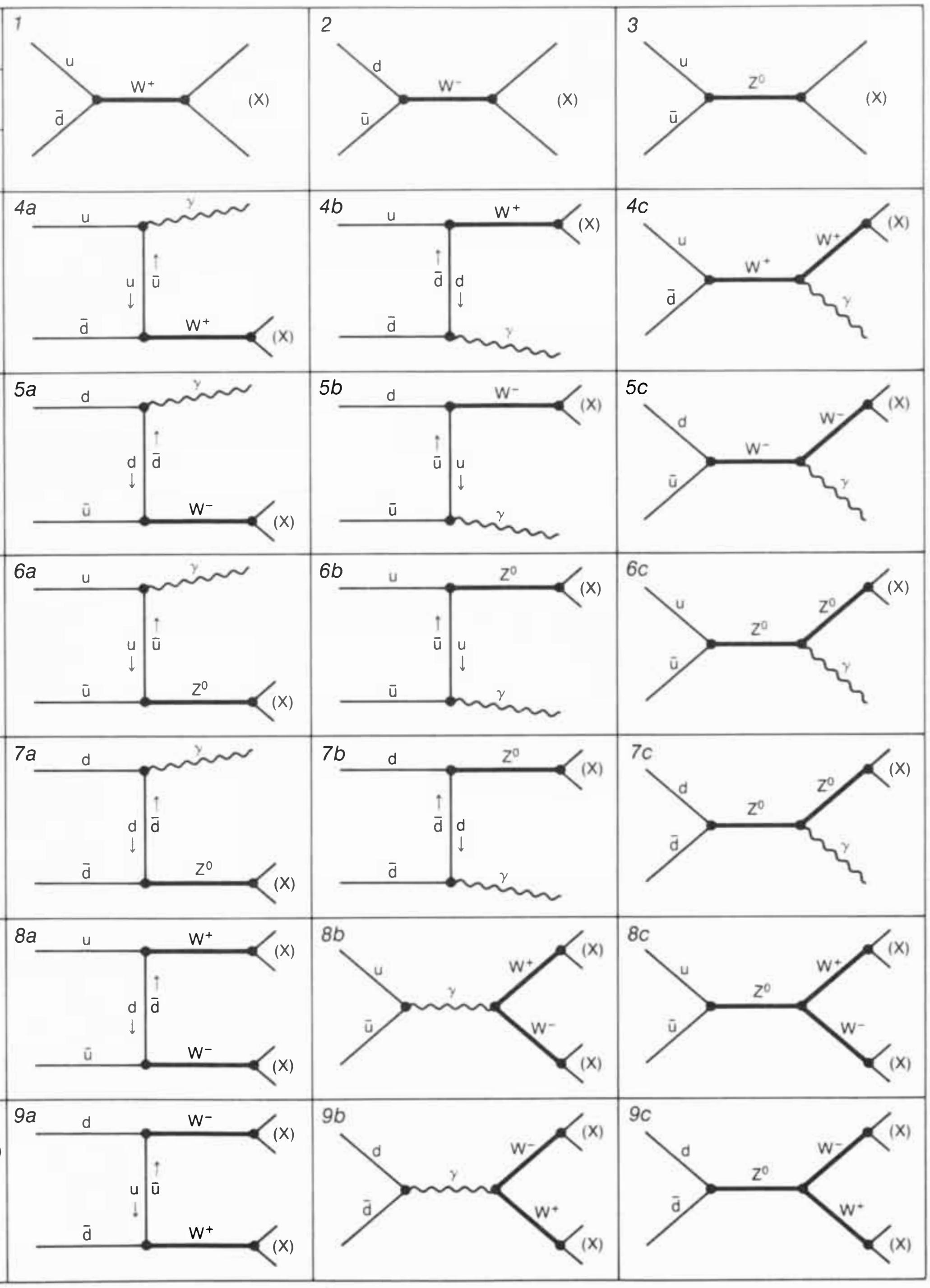
ANTIPROTON ACCUMULATOR performs two functions essential to the CERN colliding-beam experiments: it "stacks" the successively injected bunches of antiprotons, and it "cools" them by a statistical process known as stochastic cooling. For cooling the ring incorporates a number of linked "pickup" and "kicker" devices (*gray shapes*) as well as the beam-bending and beam-focusing magnets (*white shapes*) found in any storage ring. In stochastic cooling a pickup at one section of the storage ring senses the average deviation of the particles from the ideal orbit; a correction signal is then sent across the ring to a kicker on the other side, arriving just in time to nudge the particles back toward the ideal orbit. One such link is indicated.



STACKING AND COOLING of antiprotons in the AA ring are illustrated in this sequence of cross-sectional diagrams. First a bunch of about 20 million antiprotons is injected into the ring and made to circulate on the outside of the wide-aperture vacuum chamber (*1*). During the injection this space is shielded from the rest of the chamber by a mechanically operated metal shutter. The injected particles are precooled by the stochastic method for two seconds, reducing their random motion by a factor of 10 in both the longitudinal and the transverse directions (*2*). The shutter is then lowered and the precooled antiprotons are magnetically shifted into the stack position in the main body of the chamber (*3*). The shutter is raised again and a second bunch of antiprotons is injected 2.4 seconds after the first one (*4*). The second bunch is subjected to the same procedure, ending up in the stack after being precooled (*5*). About an hour later, when 1,500 bunches have been injected and some 30 billion antiprotons have been precooled and stacked, a dense core begins to form in the stack (*6*). After 40 hours, when 60,000 bunches have been injected, approximately a trillion antiprotons are orbiting in the stack, most of them concentrated in the core (*7*). Magnetic fields are then actuated to extract the core, providing about 600 billion antiprotons for the colliding-beam experiments. The residue of some 400 billion antiprotons remains stacked in the AA ring and is available to start the formation of the next core (*8*). After another 24 hours the second core of 600 billion antiprotons will be cooled and ready for injection. In each case the inside of the ring is to the right.



1. $u + \bar{d} \rightarrow W^+ + (X)$
   $(2/3) + (1/3) = (1)$
2. $d + \bar{u} \rightarrow W^- + (X)$
   $(-1/3) + (-2/3) = (-1)$
3. $u + \bar{u} \rightarrow Z^0 + (X)$
   $(2/3) + (-2/3) = (0)$
4. $u + \bar{d} \rightarrow W^+ + \gamma + (X)$
   $(2/3) + (1/3) = (1)$
5. $d + \bar{u} \rightarrow W^- + \gamma + (X)$
   $(-1/3) + (-2/3) = (-1)$
6. $u + \bar{u} \rightarrow Z^0 + \gamma + (X)$
   $(2/3) + (-2/3) = (0)$
7. $d + \bar{d} \rightarrow Z^0 + \gamma + (X)$
   $(-1/3) + (1/3) = (0)$
8. $u + \bar{u} \rightarrow W^+ + W^- + (X)$
   $(2/3) + (-2/3) = (1) + (-1)$
   $(0) = (0)$
9. $d + \bar{d} \rightarrow W^+ + W^- + (X)$
   $(-1/3) + (1/3) = (1) + (-1)$
   $(0) = (0)$



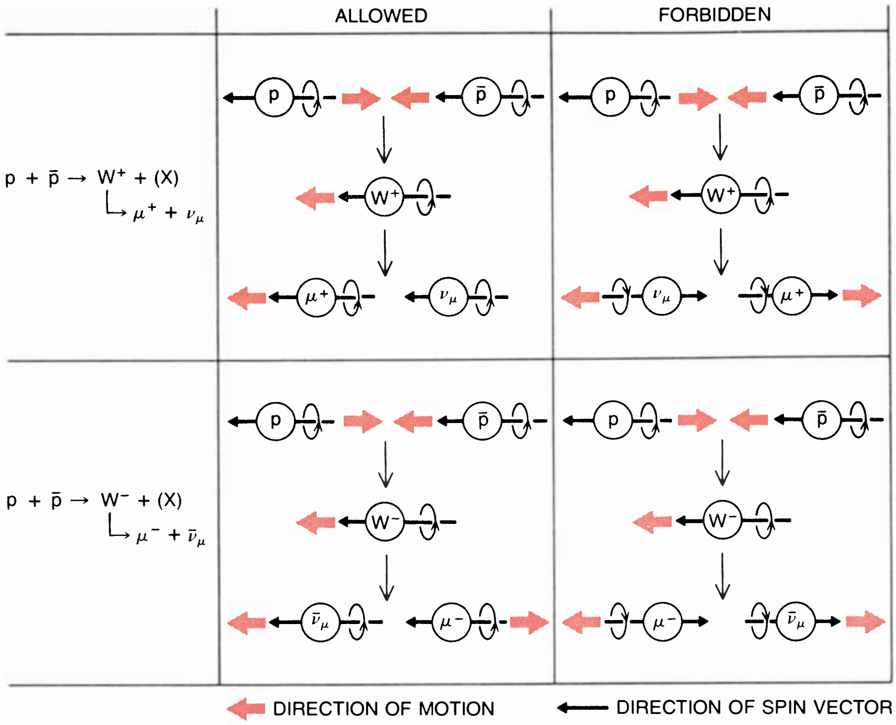
NEW PARTICLES are expected to appear in the aftermath of high-energy proton-antiproton collisions when a quark from the proton's structure interacts with an antiquark from the antiproton's structure. As the idealized drawing at the top indicates, the proton is presumed to consist of two "up" quarks (labeled *u*) and one "down" quark (*d*), whereas the antiproton has two "antiup" antiquarks ($\bar{u}$) and one "antidown" antiquark ($\bar{d}$). Quark-antiquark interactions that might contribute to the production of intermediate vector bosons are shown in symbols at the left and diagrammatically at the right. Three intermediate vector bosons are postulated: two charged particles (designated W^+ and W^-) and one electrically neutral particle (Z^0). All three are expected to be extremely short-lived and hence to be detectable only by virtue of their decay products, represented generically here by the symbol *X*. The first row shows the three most likely processes. The next four rows show processes in which intermediate vector bosons are created in association with energetic photons (γ). The last two rows show processes in which pairs of intermediate vector bosons are formed. (Where particles are shown moving in both directions along a vertical line the interaction can proceed by either mechanism.) The numbers in parentheses below the symbols at the left give the electric charge of each particle; the equations show that charge is conserved in all cases. Quarks are fractionally charged. Photons have no charge.

ing. The largest particle accelerators at the time were machines in which a single beam of protons is raised to high energy and then directed onto a fixed target. In the ensuing collision of a beam particle with a target particle most of the energy released goes into moving the two-particle system rather than demolishing it; only a small fraction of the beam energy is made available for the creation of new particles. The only chance of observing an intermediate vector boson, it seemed clear, would be in a colliding-beam machine, where the accelerated particles meet head on, transforming essentially all their energy into new particles.

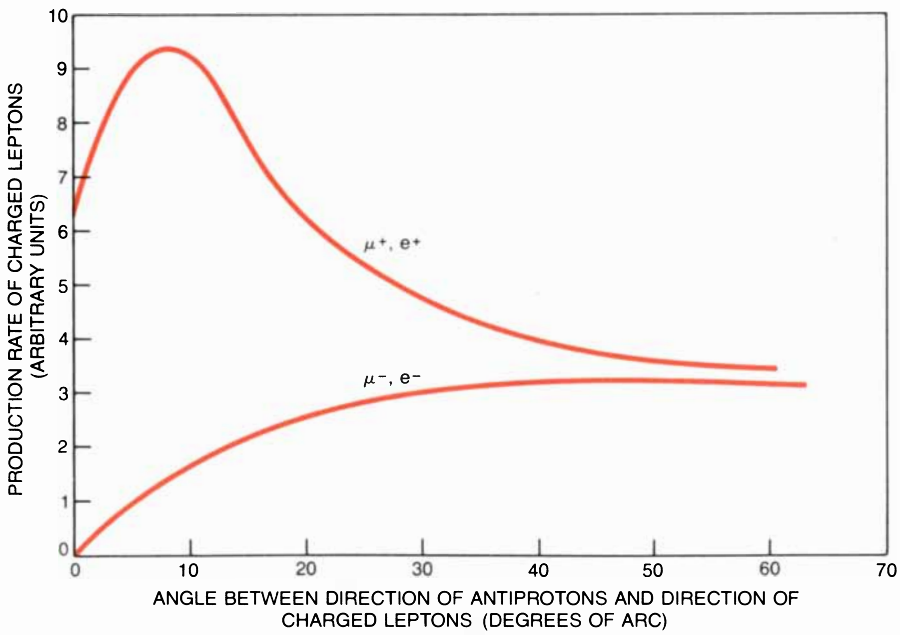
Storage rings for electrons and positrons had already been in operation for several years. The great advantage of employing electrons and positrons is that a single ring of magnets and radio-frequency cavities can simultaneously accelerate a particle and its antiparticle in opposite directions, so that counter-rotating beams are formed in a single doughnut-shaped vacuum chamber. On the other hand, because electrons and positrons are very light they rapidly dissipate their energy when they are made to follow the curved path of the storage ring. It did not seem feasible then to build an electron-positron machine large enough to reach the energy of the intermediate vector bosons. The plan instead was to build storage rings in which protons would collide head on with other protons; two interlaced rings are needed to arrange such collisions. The first of these proton-proton machines were not scheduled to begin operating until the mid-1980's or later.

Then in 1976 two of us (Cline and Rubbia), together with Peter M. McIntyre, came up with an alternative idea. Instead of building an entirely new colliding-beam machine, we proposed, it would be feasible (and much cheaper) to convert an existing fixed-target proton accelerator into a colliding-beam machine by arranging to generate a counterrotating beam of antiprotons in the same annular space occupied by the original proton beam. Our suggestion was well received, and after a thorough review of the problems likely to be encountered in such a project it was decided to build proton-antiproton machines at two of the world's largest proton accelerators: the Super Proton Synchrotron (SPS) at CERN, which began operating at a peak energy of 400 GeV in 1976, and a more advanced version of a comparable machine at Fermilab that was then still in the planning stage.

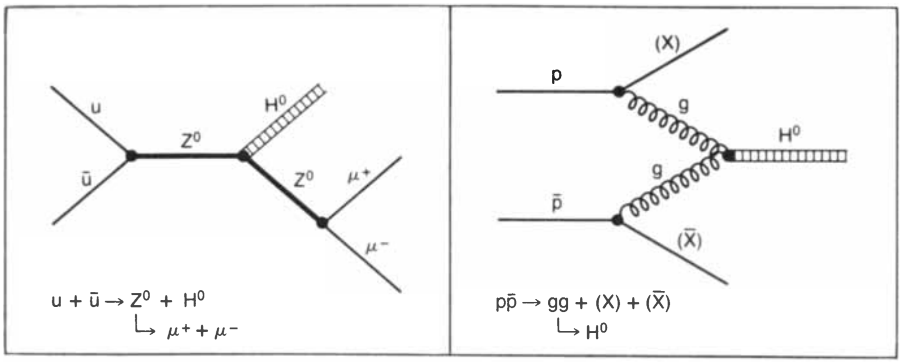
The CERN conversion was in many ways easier to accomplish, and it was completed by last summer, under the direction of Roy Billinge and one of us (van der Meer); the first proton-antiproton collisions at the designed peak energy of 270 GeV per beam were observed



DIRECTION OF SPIN of the particles involved in the production and decay of intermediate vector bosons has an important effect on the direction of motion of the decay products. The result is an asymmetry in the rate at which charged leptons are expected to be observed in different directions with respect to the incoming beams. (Leptons are particles such as electrons and muons that participate in weak interactions but not in strong ones.) For example, when a charged intermediate vector boson (W^+ or W^-) decays to form a muon (μ^+ or μ^-) and a muon-type neutrino (ν_μ or $\bar{\nu}_\mu$), most of the positively charged muons leave the collision in the direction of the incoming antiproton beam, whereas most of the negatively charged muons leave in the direction of the incoming proton beam (here left and right respectively). The observation of this effect, which is unique to interactions mediated by the weak force, would be taken as strong evidence for the fleeting existence of charged bosons among the decay products.



ASYMMETRY IS PREDICTED in the angular distribution of charged leptons emerging from proton-antiproton collisions in which charged intermediate vector bosons are created. Here the theoretical production rate of the charged leptons is given at various angles in relation to the direction of the antiproton beam. The effect is strongest in the forward direction (that is, at small angles with respect to the antiproton direction). Electrons (e^-) and positrons (e^+) as well as muons are expected to contribute to the effect, although the electrons and positrons interact with the materials of the detector more readily than muons do and hence are detected less frequently. The data points for the two curves were calculated for proton-antiproton collisions at a total energy of 2,000 GeV, the goal of the colliding-beam machine now under construction at the Fermi National Accelerator Laboratory (Fermilab) near Chicago.



HIGGS BOSON might also make its first appearance in the colliding-beam experiments at CERN or Fermilab. The discovery of this massive, uncharged particle (designated H^0) is considered the ultimate test of the "standard" unified theory linking electromagnetic interactions and weak interactions. Two processes that might lead to the production of Higgs bosons are illustrated. At the left a Higgs boson is created in association with a neutral intermediate vector boson. At the right a Higgs boson arises from the fusion of two gluons emitted during a grazing collision between a proton and an antiproton. (Gluons are the intermediary particles of the strong force that is thought to hold the quarks together inside the particles of the nucleus.)

in July. By late December, when the machine was shut down for the Christmas holiday, more than 250,000 such collisions had been recorded; because of the comparatively low rate at which intermediate vector bosons are expected to be produced in proton-antiproton collisions, however, it was not surprising that none were detected in these early runs. This situation is expected to change dramatically in the next round of experiments, in which the particle intensity of the beams, and hence the collision rate, will be increased by an order of magnitude or more.

The big proton-antiproton colliding-beam machine at Fermilab is still under construction and is scheduled to begin operating in 1985. Because it was originally designed to accelerate a single beam of protons to an energy of 1 TeV, or a trillion electron volts, it was named the Tevatron. In its reincarnation as a colliding-beam machine it is expected to be able to produce collisions with a total energy of 2 TeV (2,000 GeV), as opposed to 540 GeV for the CERN machine. When the Fermilab machine is completed, it will have the further distinction of being the first large accelerator to employ a ring of superconducting magnets.

How does one go about generating opposed beams of matter and antimatter in a storage ring? The hardest part in the two present instances is to accumulate a dense enough "bunch" of antiprotons to ensure a large number of collisions with the counterrotating protons. Unlike protons, antiprotons are not readily available from any natural source; they must themselves be created in high-energy collisions. A beam of high-energy protons is directed at a metal target, and antiprotons created in collisions with the target atoms are steered magnetically into a specially designed storage ring. The process is extremely inefficient; on the average one comparatively low-energy antiproton is produced for every million or so high-energy protons striking the target. To put this production rate in perspective, it has been calculated that in order to obtain a useful number of proton-antiproton collisions in the colliding-beam machine at CERN one must collect bunches of antiprotons (and protons) each made up of at least 100 billion particles. Successive bunches of antiprotons are collected and "stacked" every 2.4 seconds; at this rate it takes about 24 hours to accumulate a few hundred billion antiprotons for the colliding-beam experiments at CERN.

Creating enough antiparticles is not the only problem. As the antiprotons emerge from the target they have a range of velocities and directions. Viewed in their own frame of reference the antiprotons form a gas, and their random motions are indicative of a temperature. If the temperature is too high, some of the particles will strike the walls of the accelerator and the beam will be dissipated. Therefore some method is needed to "cool" the antiproton beam (that is, to reduce its random motions) in order to keep it as concentrated as possible before it enters the accelerator ring.

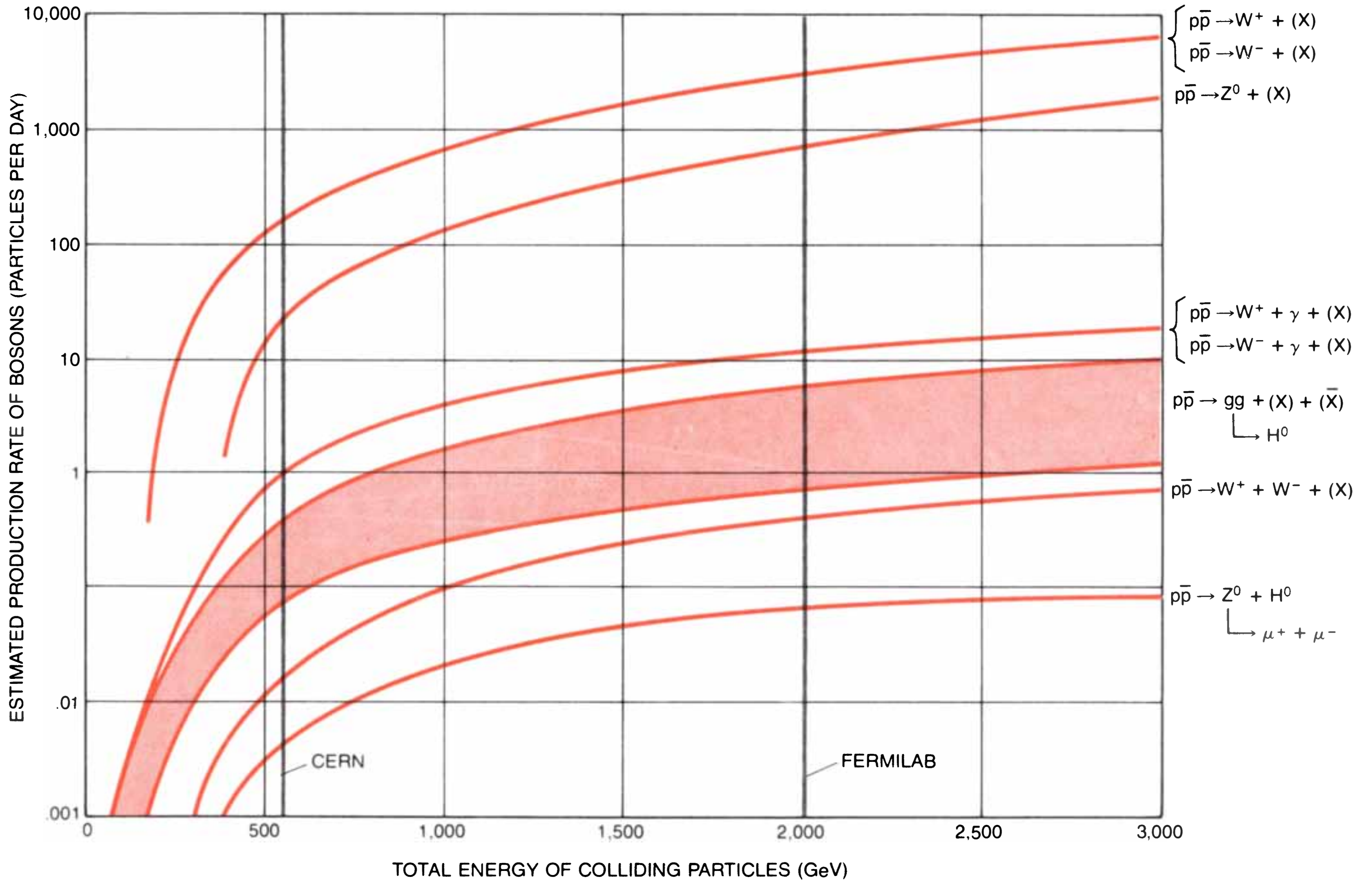
One such beam-cooling technique, called electron cooling, was first proposed more than a decade ago by Gersh I. Budker of the Institute of Nuclear Physics at Novosibirsk in the U.S.S.R. Basically it operates by mixing a "cool" beam of electrons (one in which all the particles have the same velocity and direction) with the "hot" antiproton beam for a short distance. In the process some of the random thermal energy of the antiprotons is transferred to the electrons. Mixing the antiproton beam repeatedly with fresh electron beams can cool the antiprotons significantly, provided their energy is not too high to start with. As it happens, the CERN scheme calls for an antiproton beam whose energy is initially too high to be cooled effectively by this method, and it is no longer being studied for this purpose. Electron cooling is still under consideration, however, for a role in the Fermilab project.

Another beam-cooling method, better suited to the requirements of the CERN proton-antiproton machine, was invented in 1968 by one of us (van der Meer). This method, called stochastic cooling (because it relies on a statistical process), utilizes a "pickup," or sensing device, in one section of a storage ring to measure the average deviation of the particles from the ideal orbit. The measurement is converted into a correction signal, which is relayed across the ring to a "kicker" device on the other side. The kicker applies an electric field to its section of the ring in time to nudge the center of mass of the passing particles back toward the ideal orbit. Although the particles move with nearly the speed of light, the correction signal can arrive in time because it takes the shorter path across a chord of the cooling ring.

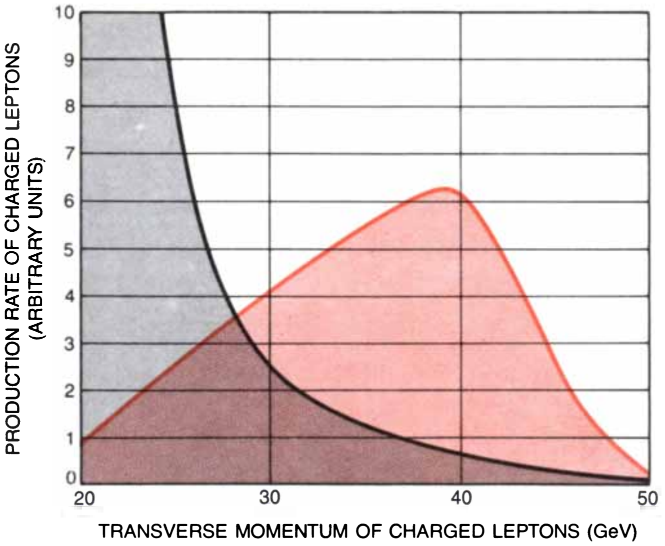
Both beam-cooling techniques have been tested successfully in the past few years at Novosibirsk, CERN and Fermilab. As a result there is every reason to believe full-scale antiproton collector rings such as the ones at CERN and Fermilab will work as planned. Beam cooling is becoming a routine part of accelerator technology.

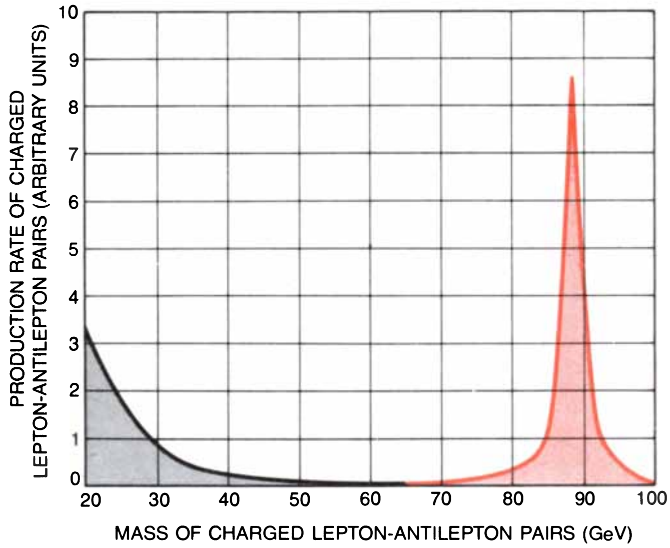
For the experiments at CERN the particles are directed through a complex sequence of interconnected beam-manipulating devices. First a beam of protons is accelerated to an energy of 26 GeV in the Proton Synchrotron (PS), the original accelerator ring at CERN, completed in 1959. The proton beam is then directed at a copper target, producing a spray of particles, including a small number of antiprotons with an energy of 3.5 GeV. The antiprotons are collected and transferred to a wide-aperture storage ring called the Antiproton Accumulator (AA), where they are first precooled by the stochastic method and then moved to a slightly smaller orbit, where they are stacked with the previously injected bunches and subjected to further cooling. After a few hundred billion antiprotons have been collected they are sent back to the PS ring, where they are accelerated to 26 GeV before being injected into the SPS. Meanwhile protons at 26 GeV from the PS ring are injected into the SPS ring in the opposite direction. The counterrotating beams are finally accelerated to 270 GeV each in the SPS ring. The beams collide at two interaction sites, where the large particle detectors are placed [*see illustration on page 50*]. The interactions are so rare that the beam lifetime of several hours is not affected by them.

At present the Fermilab plan calls for

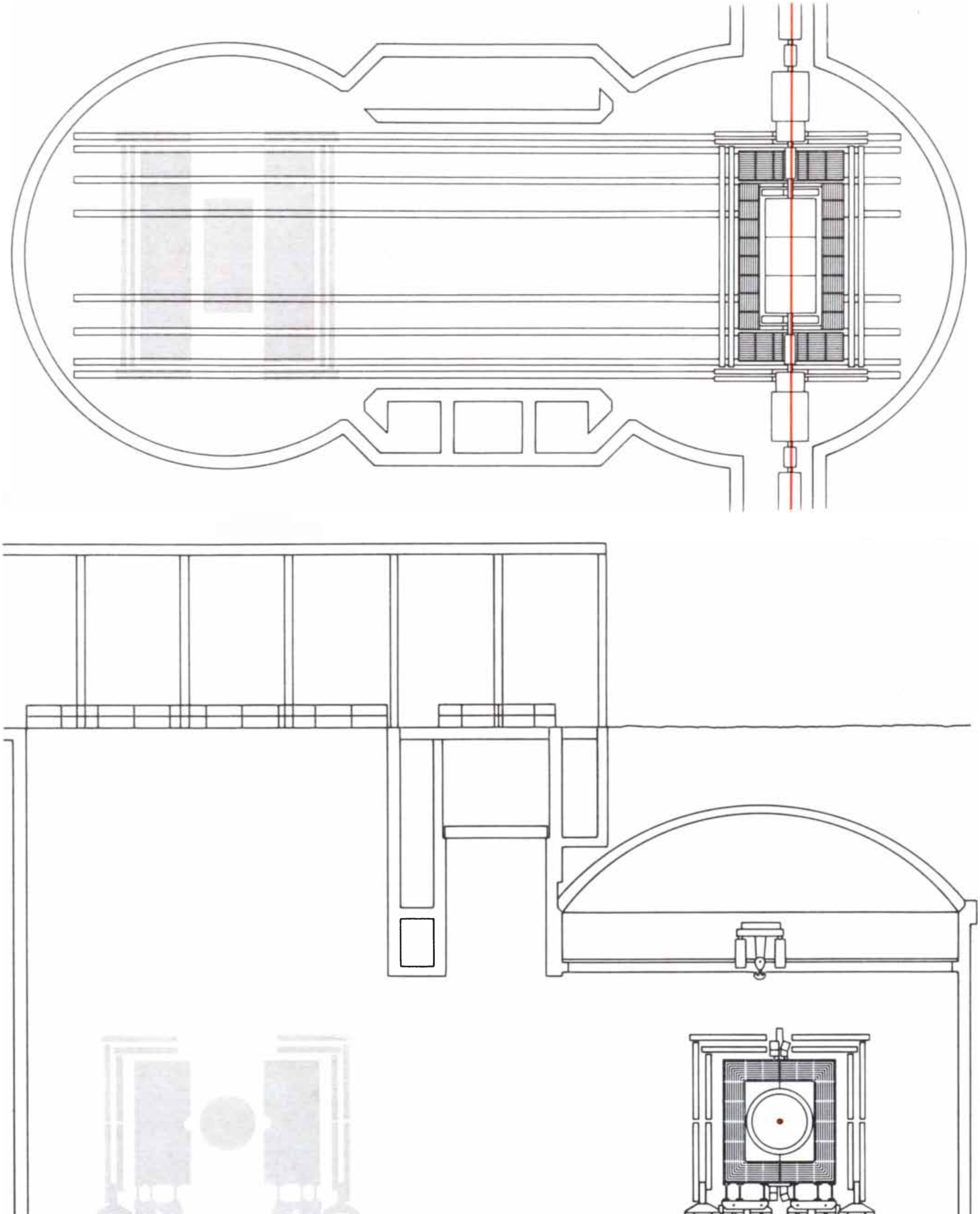


ESTIMATED PRODUCTION RATES of intermediate vector bosons and Higgs bosons by various collision processes vary as a function of the total energy of the colliding beams. The curves indicate the rate calculated for each process for one day of operation at the designed collision rate of the CERN and Fermilab colliding-beam machines. (The actual detected rate will be lower because of various experimental background effects.) The two vertical gray lines indicate the designed total energy of the proton-antiproton collisions at CERN and Fermilab. The processes that result in the production of a charged intermediate vector boson accompanied by an energetic photon are expected to serve as sensitive indicators of the magnetic properties of the W^+ and W^- particles. The process in which a Higgs boson is produced by the fusion of two gluons has a less predictable rate than the other processes and so is represented by a range of values.



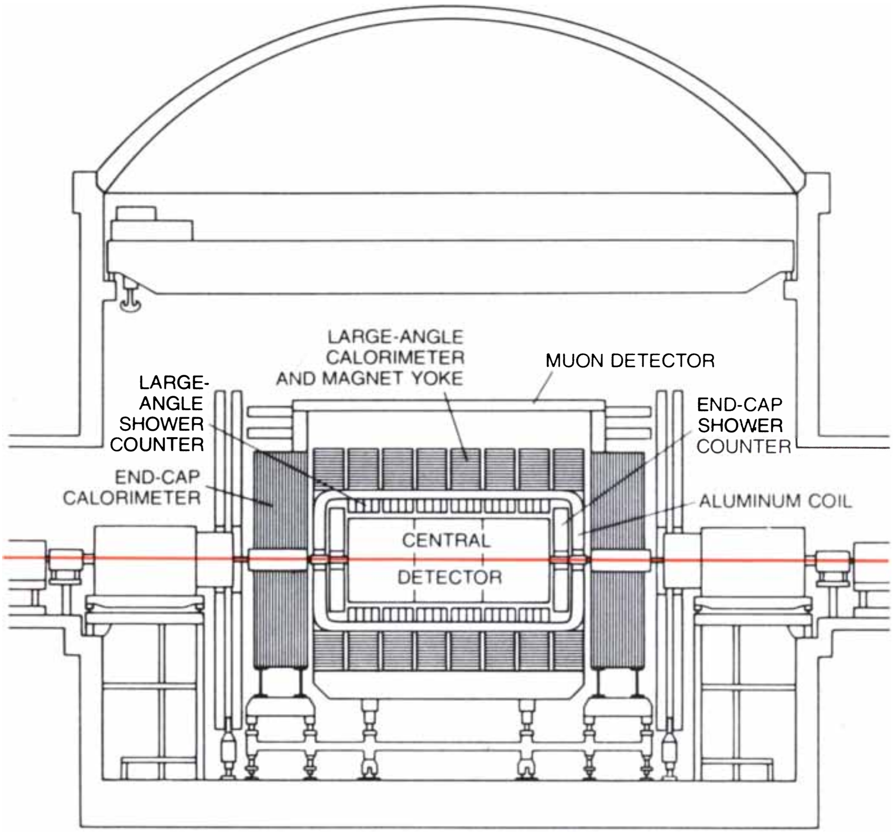


TELLTALE SIGNALS of the production of intermediate vector bosons (*color*) are expected to stand out above the background "noise" (*gray*), particularly at large angles with respect to the beam axis. The graph at the left shows the calculated mass spectrum of the charged leptons that would be emitted with a large transverse momentum from proton-antiproton collisions in which charged intermediate vector bosons are created. The peak in the signal is predicted to appear at about half the estimated mass of a charged intermediate vector boson. The background rate of leptons expected from other sources related to the collision process has no peak and is lower than the rate expected from the decay of the W^+ and W^- particles. The graph at the right shows the calculated mass spectrum for the decay of a neutral intermediate vector boson into a pair of charged leptons. The spectrum in this case has a peak near the predicted mass of the Z^0 particle (90 GeV). The negligible background noise makes this process the one in which the Z^0 is most likely to be discovered.



UA1 DETECTOR IS INSTALLED in a cavernous experimental area approximately 25 meters underground. In both the plan view (*upper diagram*) and the corresponding end view (*lower diagram*) the detector is drawn in black line when it is in its operating position straddling the beam pipe of the SPS and in gray tone when it is parked in its garage. The UA1 is by far the largest detector ever built for a colliding-beam experiment: it is 10 meters long by five meters wide and weighs 2,000 tons. It was built at a cost of roughly $20 million.

a more intense antiproton source than the one in operation at CERN. Protons with a higher energy will be used to make the antiprotons at Fermilab. Several alternative schemes are being investigated. One would rely on a combination of stochastic cooling and electron cooling to accumulate 100 billion antiprotons in less than an hour. The antiprotons would be injected into the Tevatron ring, accelerated to an energy of 1,000 GeV and made to collide with counterrotating 1,000-GeV protons in two experimental areas. The first area, where construction is scheduled to start soon, is designed to house a very large detector.



SIDE VIEW of the UA1 detector shows it in place in the SPS beam line. Experimental components for various purposes, including the search for intermediate vector bosons, are labeled.

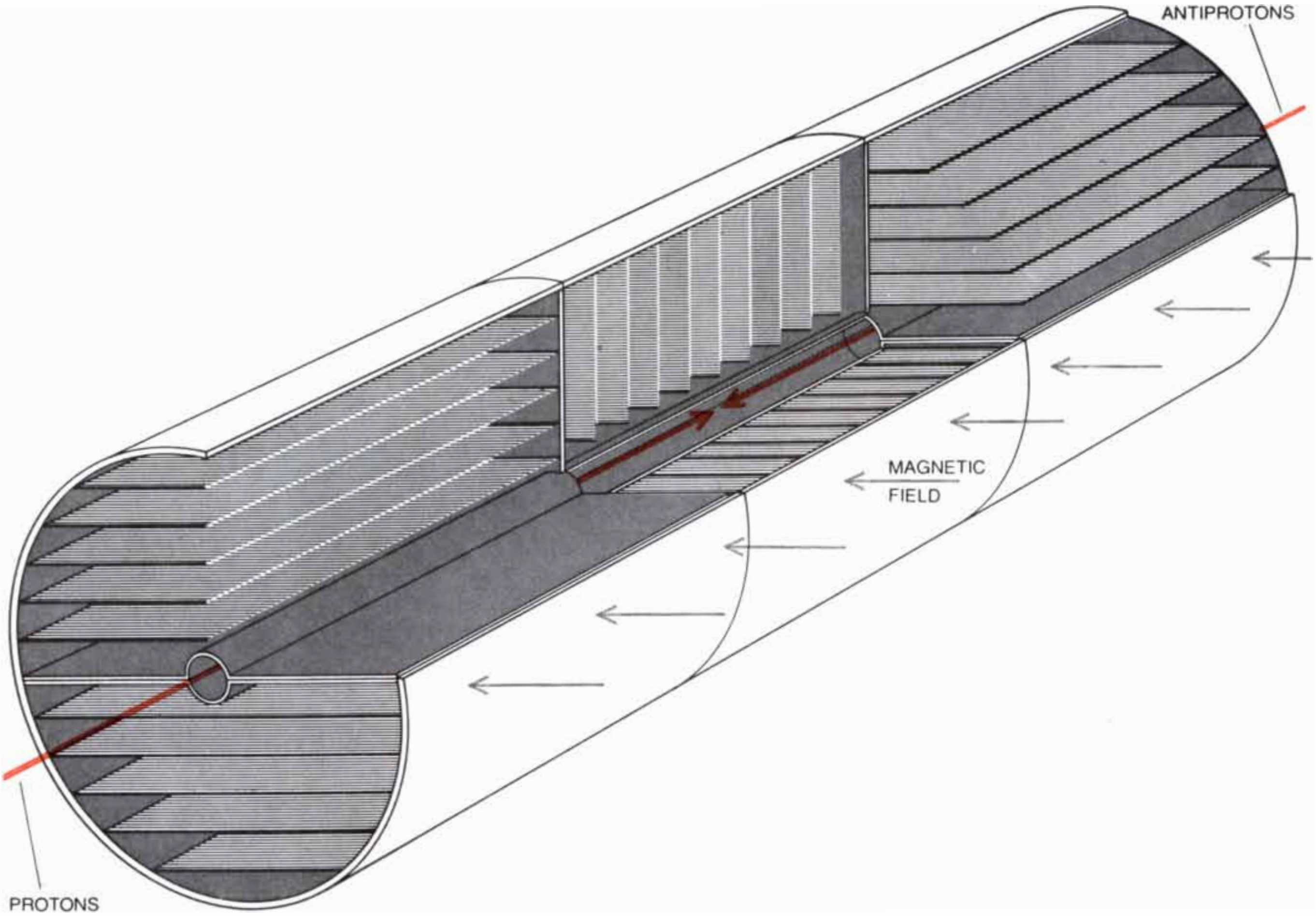
According to the electroweak theory, intermediate vector bosons can be created in proton-antiproton collisions by a variety of mechanisms in which a quark from the proton's structure interacts with an antiquark from the antiproton's structure. Both the proton and the antiproton are assumed to be made up of three constituent particles; in the whimsical nomenclature of the QCD theory the proton is said to harbor two "up" quarks (labeled u) and one "down" quark (d), whereas the antiproton has two "antiup" antiquarks ($\bar{u}$) and one "antidown" antiquark ($\bar{d}$). When a quark and an antiquark collide, they annihilate each other, creating a burst of energy that can rematerialize as new particles, including intermediate vector bosons. In some cases a single intermediate vector boson is expected to appear (accompanied by other kinds of particles); in other cases a pair of intermediate vector bosons is predicted [*see illustration on page 52*].

The expected production rate of W^+, W^- and Z^0 particles in proton-antiproton collisions varies as a function of an experimental parameter called the luminosity, which is defined as the number of high-energy particles per square centimeter per second passing through the cross section of the interaction region. The designed luminosity of the CERN machine, assuming the injection of 600 billion antiprotons per bunch, is 10^{30} particles per square centimeter per second. Given the same number of antiprotons per bunch, the Fermilab machine should attain a luminosity of 4×10^{30} particles per square centimeter per second (owing to its higher energy and hence smaller beam size). It can be calculated that at such luminosities the production rate of W^+, W^- and Z^0 particles, both singly and in pairs, should be high enough for them to be detected fairly often, perhaps as often as thousands of times per day [*see top illustration on page 55*]. Just before the CERN machine was shut down in December it briefly reached a luminosity of 10^{28} particles per square centimeter per second, putting it at the threshold of where one would expect to catch the first glimpse of an intermediate vector boson. (The initial operating experience with the CERN collider has also shown that proton-antiproton collisions could in principle be obtained at luminosities of 10^{31} particles per square centimeter per second or higher, provided enough antiprotons are available.)

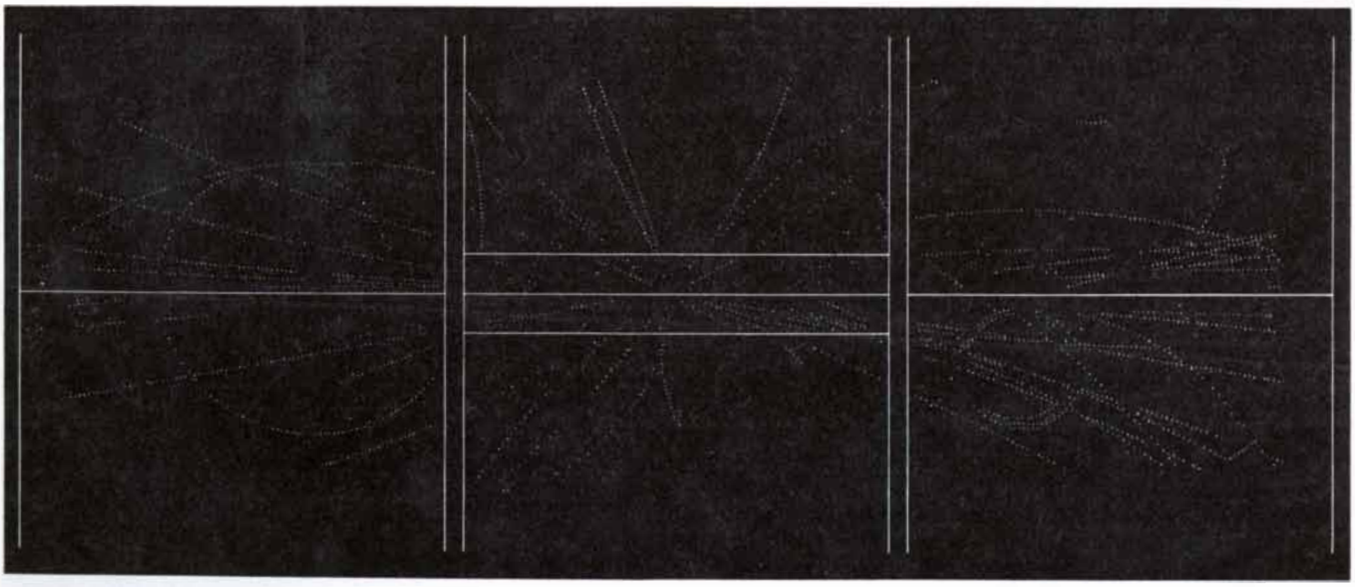
How will intermediate vector bosons produced in such collisions make their presence known? The lifetime of the particles is expected to be exceedingly short. In approximately 10^{-20} second they should decay spontaneously to form a variety of other particles, mainly quark-antiquark pairs and lepton-antilepton pairs. (Leptons are particles that respond to the weak nuclear force but not to the strong one.) Charged leptons such as electrons and muons can be detected by various means. In general the object is to detect the charged leptons from the decay of the intermediate vector bosons and to compare their production rate or other properties with the values expected from other causes arising in the collision process. The signal, in this case the angular distribution of leptons from the decay of W^+, W^- or Z^0 particles, is expected to stand out above the background "noise" of leptons from other sources, particularly at large angles with respect to the beam axis [*see bottom illustration on page 55*].

One unmistakable indication of the presence of intermediate vector bosons would be the appearance of a marked asymmetry in the rate at which leptons are detected in the forward and backward directions (measured arbitrarily with respect to the direction of the antiproton beam). Leptons that arise directly from strong or electromagnetic interactions of the beam particles should be completely symmetrical. According to the electroweak theory, however, decaying intermediate vector bosons should emit positively charged leptons predominantly in the forward direction and negatively charged ones predominantly in the backward direction. The expected lepton asymmetry, which is unique to events mediated by the weak force, results from the spins of the particles involved in the production and decay of intermediate vector bosons [*see illustrations on page 53*]. The observation of this effect will be taken as strong evidence that the long-sought intermediate vector bosons have finally been discovered. Their other properties can then be measured.

The ultimate test of the correctness of the electroweak theory would be the observation in the debris of proton-antiproton collisions of the Higgs boson. The discovery of this particle would demonstrate not only that electromagnetism and the weak force are unified but also that the unification is of the kind prescribed by the standard electroweak theory. A full discussion of



CENTRAL DETECTION SYSTEM of the UA1 consists of three cylindrical "drift" chambers, each containing an array of closely spaced wires and a gas at low pressure. In all three chambers the wires are strung horizontally. In the central chamber the horizontal wires are arranged in vertical planes; in the two flanking chambers they are arranged in horizontal planes. A charged particle passing through the chamber ionizes molecules of the gas, which then drift to the wires, depositing their charge. The pattern of charges appearing on many wires is recorded electrically and can later be analyzed by a computer to reconstruct the trajectory of the particle on the face of a cathode-ray tube. The chambers are approximately three meters in diameter. The wires are spaced about three millimeters apart.



VISUAL RECORD of a proton-antiproton collision that took place late last year in the central detection system of the UA1 was made by photographing a computer-generated display. The event is the same one that appears on the cover of this issue of SCIENTIFIC AMERICAN. More than 250,000 events of this type have been registered so far in the computer. Millions more will be recorded when the search for intermediate vector bosons resumes at CERN next month. The applied magnetic field bends the paths of the charged particles.

the experimental technique required to detect the Higgs boson is beyond the scope of this article. Proton-antiproton collisions could give rise to Higgs bosons, however, and the estimated rate of production is high enough for them to be discovered by means of their characteristic decay products in either the CERN or the Fermilab colliding-beam machines.

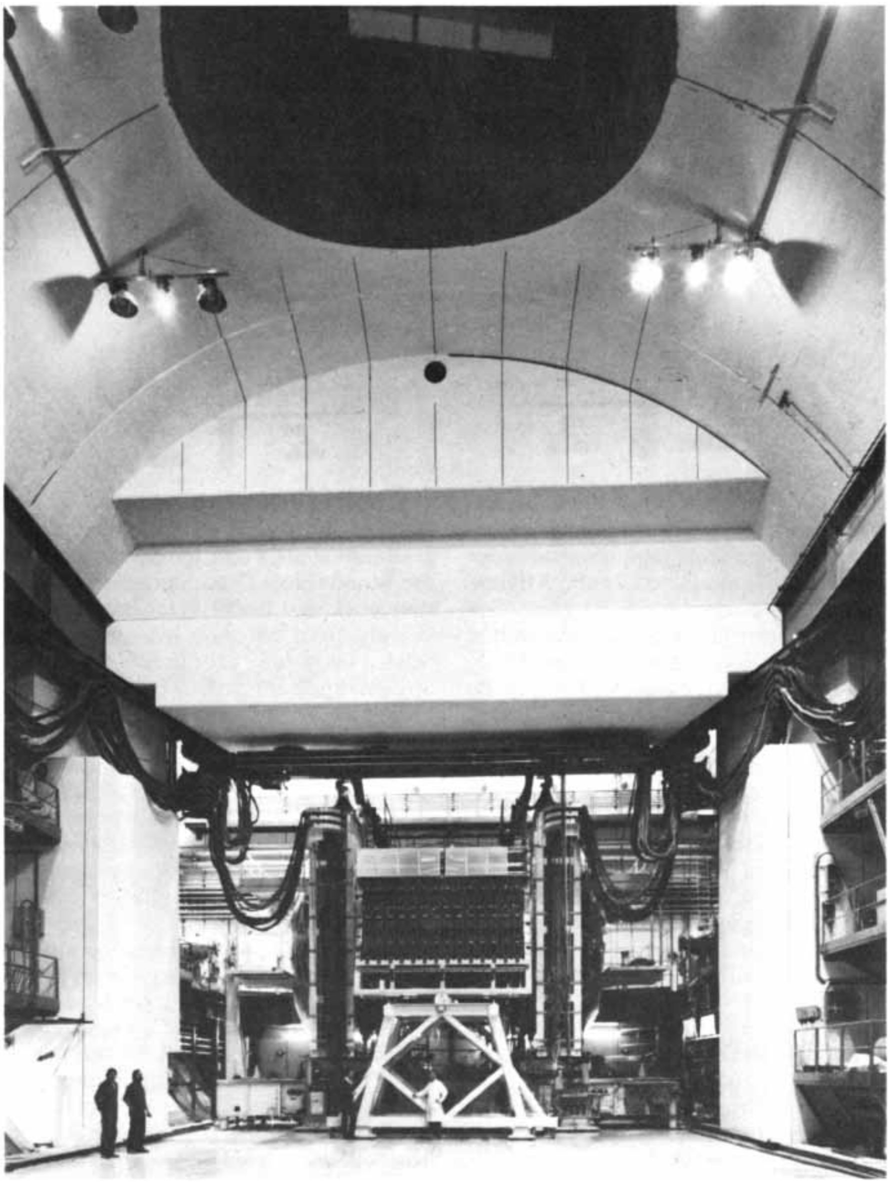
Several large detectors have been designed to search for the decay products of intermediate vector bosons and Higgs bosons. One of these devices, named the UA1, is now finished and ready for next month's resumption of the search at CERN. The detector is the result of a collaborative effort by a team of more than 100 physicists from 11 institutions in Europe and the U.S.: the University of Aachen, the Annecy Particle Physics Laboratory, the University of Birmingham, CERN, Queen Mary College (London), the Collège de France (Paris), the University of California at Riverside, the University of Rome, the Rutherford Laboratory, the Saclay Nuclear Research Center and the University of Vienna. It is 10 meters long by five meters wide, and its total weight is 2,000 tons. The underground hall in which it is installed is large enough for the detector to be rolled back into a "garage" when it is not in place in the path of the colliding beams.

The UA1 detector is a multipurpose device, designed to sense many kinds of particles and to collect information over a wide solid angle surrounding the point where the beams collide. It measures the energy of the particles by several means, including the curvature of their paths in a magnetic field. A large dipole magnet applies the main magnetic field horizontally throughout a volume of 85 cubic meters.

Inside the magnet, surrounding the beam tube, there are three "drift chambers," each containing an array of closely spaced wires and a gas at low pressure. An electrically charged particle passing through the chamber ionizes molecules of the gas; the ions then drift to the wires, where they deposit their charge. From the pattern of charges appearing on many wires the trajectory of the particle can be reconstructed. The central drift chamber has its wires arranged in vertical planes; the two flanking chambers have their wires arranged in horizontal planes. Signals from particles crossing the wire planes can be processed by a computer to yield an image of the decay products on the face of a cathode-ray tube [*see illustrations on opposite page*].

Surrounding the three drift chambers are various other detectors. Just outside the innermost detector is a lead calorimeter, a device that measures the amount of energy deposited in it by a charged particle such as an electron. The calorimeter is enclosed in turn by a series of iron plates interleaved with scintillation counters to measure the energy of heavier particles such as pions by means of their interactions with iron atoms in the plates. Finally, on the outside of the apparatus there are several large chambers for the detection of muons that pass through both the lead and the iron plates.

Another large detector, the UA2, is designed more specifically to search for intermediate vector bosons. It has no magnetic field but instead relies on a large array of calorimeters similar to those in the UA1 detector to measure the energy and direction of the emerging particles [*see illustration above*]. Detectors comparable to the UA1 and the UA2 are planned at Fermilab. If intermediate vector bosons exist, we believe these detectors will be adequate to discover them and to investigate their properties, thereby confirming the unified electroweak theory. If Higgs bosons exist, they might also be detected, thereby providing further support for the theory. Of course, it is still possible that the electroweak theory is incorrect and that none of these particles is real. One way or the other, the answer should be known soon.



ANOTHER LARGE DETECTOR, the UA2, has recently been completed at CERN. It was designed more specifically to search for intermediate vector bosons. Unlike the UA1, it has no magnetic field. It is installed in a second experimental area, built more than 60 meters underground, some distance away from the UA1. The detector is shown in place in the SPS beam line; when it is not in operation, it is rolled back into the large work area in the foreground. The large cylindrical shaft at the top is used to lower bulky components from the surface.

Rings of the Gulf Stream

The Gulf Stream and other currents give rise to eddies that trap water whose physical and biological properties differ from those of the surrounding sea. Therefore the oceans are shifting mosaics

by Peter H. Wiebe

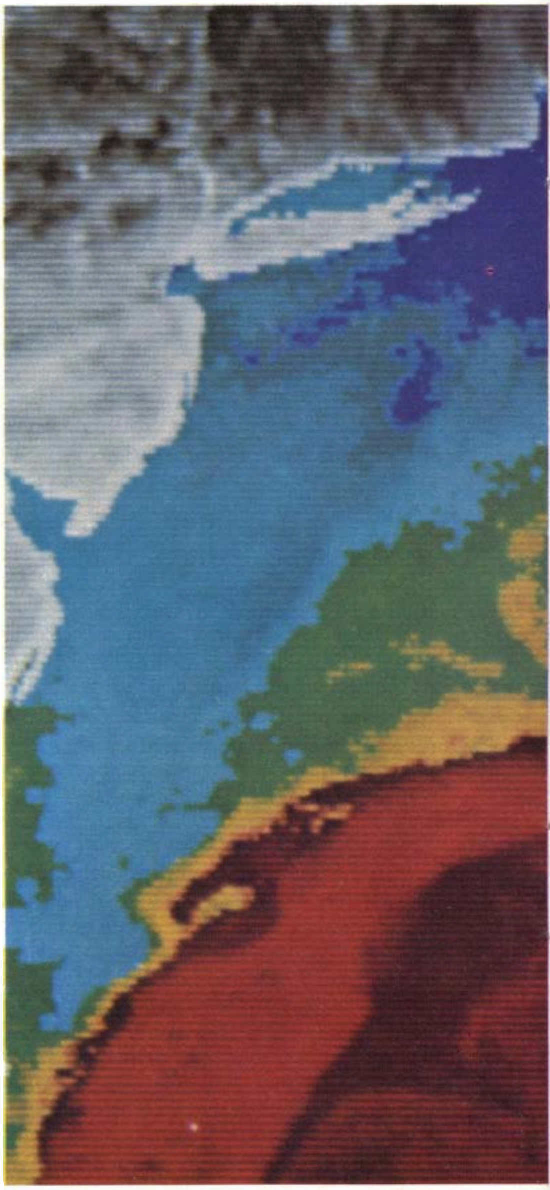
The view that large areas in the oceans are uniform in their physical, chemical and biological properties is well entrenched. In the Atlantic Ocean, for example, the Sargasso Sea has long been taken to be essentially a stagnant pool with a uniform profile of temperature with respect to depth in the sea and a uniform paucity of life. Today this view is giving way. New research methods involving arrays of ships, instruments deployed from aircraft and remote measurements made from aircraft and satellites reveal patterns of variation on a variety of scales of space and time unsuspected 20 years ago. It emerges that the large-scale processes of ocean current and world climate that create and maintain the distribution of marine plants and animals on scales of thousands of kilometers also perturb these ecological systems by introducing eddies in the oceans on scales of hundreds of kilometers.

The most dramatic of these eddies are closed rings as much as 300 kilometers in diameter that break off from currents such as the Gulf Stream, which flows across the Atlantic to the north of the Sargasso Sea. As each ring forms, a column of water is appropriated from one side of the current and moved to the other side, a realm where the temperature, salinity, oxygen content and distribution of life are quite different. A number of rings can occupy an ocean basin at any one time. As a result the Sargasso Sea and other regions are mosaics.

My own interest in oceanic eddies was sparked shortly after I had finished my graduate studies and joined the staff of the Woods Hole Oceanographic Institution in 1969. I heard Peter Saunders of Woods Hole describe his attempts to track a large eddy produced by the Gulf Stream, and I immediately realized that the persistence of such an eddy (they can last for months or even a year or two) must have biological consequences. I also found that no one had yet tried to study the relation of eddies to various forms of life. It took my colleagues and me another three years to get to an eddy and begin to study its biology.

The term ring in reference to closed eddies that detach themselves from the Gulf Stream results from a suggestion made by Frederick C. Fuglister of Woods Hole, who pioneered the study of their physics beginning in the late 1950's. The term signifies that a ring of water detached from the Gulf Stream flows around a core of water that is distinct from both the ring current and the water in which the ring with its core is embedded. Oceanographers had encountered such rings on cruises south of the Gulf Stream in the 1930's and 1940's, but it was not until Operation Cabot, a survey of the Gulf Stream undertaken with six ships in 1950 by Woods Hole in collaboration with other institutions, that a ring was actually recorded in the act of forming.

The process of formation begins when the Gulf Stream starts to meander as it flows northeastward and away from the North American continent at Cape Hatteras, N.C. If the meander becomes a loop, it will pinch off from the stream, much as an oxbow lake forms from a meander in a river. The result will be a



FOUR RINGS occupy the image at the right, which shows in false color the pattern of intensity of the infrared radiation emitted from the surface of the northwestern Atlantic Ocean last September 12–15. The pattern indicates the surface seawater's temperature. Two rings are south of the Gulf Stream, the red current that snakes across the image from the southwest to the northeast. They are the two bronze-colored oval patches west of the small white cluster of clouds. In their case an annulus of water detached from the Gulf Stream surrounds a core of cold water appropriated from the Slope Water, which appears in shades of blue. Meanwhile two rings are north of the Gulf Stream. One of them has a green core partially surrounded by yellow; it lies due south of Cape Cod. The other one, which looks more complete, is toward the eastern edge of the image. It formed two months before the image was made, but its clockwise rotation is now pulling a red and yellow streamer of Gulf Stream water three-fourths of the way around itself. For these rings a Gulf Stream annulus surrounds a core of warm water appropriated from the Sargasso Sea south of the stream. The data for the image were recorded by an infrared sensor aboard the satellite NOAA-7 and were processed with the aid of a computer by Robert H. Evans, Otis B. Brown and James W. Brown of the University of Miami.

ring of swift-moving Gulf Stream water around a core of seawater of different origin. A ring that forms to the north of the stream will have a core of water appropriated from the south, that is, the Sargasso Sea. Such water is warm; the Sargasso Sea is notable for an extended layer in which the water temperature is 18 degrees Celsius, plus or minus one degree. Typically the layer begins at a depth of about 200 meters and ends at about 600 meters. Conversely, a ring that forms to the south of the Gulf Stream will have a core of water appropriated from the north, specifically the Slope Water regime, which extends from Cape Hatteras to the Grand Banks off Newfoundland and lies between the Gulf Stream and the continental shelf of North America. Such water is cold; it has a characteristic pattern that makes its temperature at a given depth some 10 degrees C. colder than the water of the Sargasso Sea.

The newly formed rings are big. Cold-core rings generally have a diameter of 150 to 300 kilometers. This gives them a surface area approximately equal to that of Massachusetts, New Hampshire, Rhode Island and Vermont combined. Moreover, they extend to the sea floor, which lies at a depth of 4,000 to 5,000 meters. Warm-core rings are generally somewhat smaller: they are 100 to 200 kilometers in diameter, and their vertical extent is often detectable only down to 1,500 meters. They have been studied much less than cold-core rings, however, and further measurements may change our understanding of them.

Each ring, warm-core or cold-core, includes a Gulf Stream remnant whose annular width can be as great as 100 kilometers, the width of the Gulf Stream itself. The sustained rotational velocity of the water in this annulus can be as high as three knots, or 1.5 meters per second, at a distance of 30 to 40 kilometers from the center of the ring. Buoys equipped with a radio transmitter and tracked by a satellite loop around a young ring in a period of two to five days. The rotational velocity decreases toward the center of the ring, and it decreases with distance in the water surrounding the ring. It also decreases with depth. Nevertheless, velocities greater than .5 meter per second have been measured at depths of 400 to 500 meters below the strong currents at the surface.

The anomalous conditions entailed by the presence of a cold-core ring in the Sargasso Sea are most easily appreciated in charts that show the vertical temperature structure of the sea, and in particular the 15-degree isotherm: the curve that marks the depth of water whose temperature is 15 degrees C. In the Sargasso Sea the 15-degree isotherm typically lies 600 to 700 meters below the surface. Near the southern margin of the Gulf Stream it rises gradually. Then in the stream itself its rise is quite dramatic. In fact, the gradient of temperature across the Gulf Stream is one of the steepest anywhere in the open oceans of the world. The gradient is so steep that the northern margin of the

stream is called the Cold Wall. Along the Cold Wall the 15-degree isotherm is at a depth of 100 meters. In the Slope Water to the north of the stream it lies still closer to the surface.

The temperature structure of a newly formed cold-core ring shows similarly steep gradients. Consider the cold-core ring designated Bob, which formed in February, 1977, and was tracked and studied throughout that year from its birth to its death. At the center of Ring Bob the isotherms for the temperatures from 10 through 16 degrees C. were elevated as much as 600 meters above their normal depths in the Sargasso Sea. The density of the seawater, whose pattern essentially parallels that of the temperature, showed similar elevations.

The increased height of the dense seawater meant that Ring Bob embodied a huge reservoir of potential energy. It also meant that Ring Bob exerted pressure on the surrounding water. The pressure was augmented by the centrifugal force exerted by the rotation of the ring. In addition it was modulated because the surface of the sea in a ring deviates from being gravitationally level. (Measurements made from satellites indicate that the surface at the center of a ring is depressed by about .5 meter.) The outward pressure exerted by a ring is balanced largely by the Coriolis effect, a force that acts on a ring as it migrates in the ocean. The Coriolis effect arises because the earth is rotating. More broadly, the migration of a ring and its eventual decay are thought to be mostly due to imbalances among pressure, centrifugal force and the Coriolis force. Other forces such as friction amount to only about 1 percent of these primary forces.

The distribution of salinity in the sea gives a further elucidating glimpse into the structure of rings. In particular the salinity in the upper few hundred meters of a cold-core ring can be one part per 1,000 lower than the salinity of water at the same temperature (but greater depth) in the surrounding Sargasso Sea. This anomaly, which is quite striking for water in the open sea, can persist for 18 months or even longer; therefore it constitutes evidence that a substantial part of a ring remains intact over a period of time. It further shows that the ring is quite different from a waveform in which cold water temporarily wells up in the ocean.

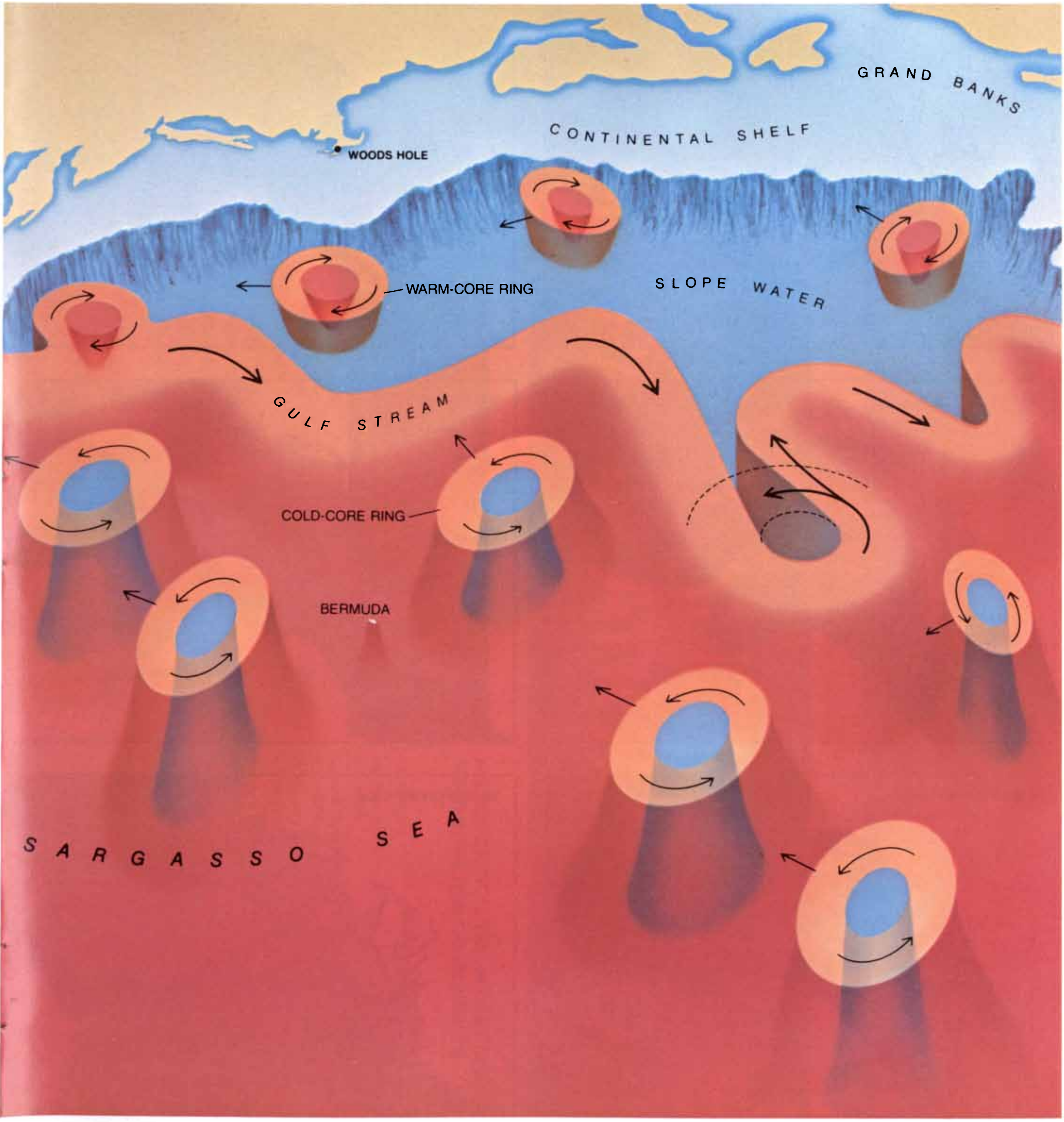
The distribution of oxygen dissolved in the ocean is also diagnostic. Gulf Stream water from the Caribbean has a low concentration of oxygen for water temperatures between eight and 18 degrees C.; Slope Water and the cores of young cold-core rings have a high concentration, and Sargasso Sea water and the cores of young warm-core rings have an intermediate concentration. It follows from these differences that high-resolution measurements of the oxygen concentration along the flanks of rings make it possible to classify the seawater according to its origin. The measurements, which were made by Dana R. Kester of the University of Rhode Island, clearly show that water from the flanks of a cold-core ring and water from the surrounding Sargasso Sea are interleaved. The rate of exchange in these interleavings has not yet been determined. In fact, it is not yet clear whether the interleavings represent exchanges of water. Instead they might be artifacts that persist from the time the ring formed.

How many rings form per year? How

TYPICAL DISTRIBUTION OF RINGS in the northwestern Atlantic is depicted in a schematic drawing based on an analysis of ocean temperature and other data by Philip L. Richardson of the Woods Hole Oceanographic Institution. In general cold-core rings, which occupy the Sargasso Sea, have a diameter as great as 300 kilometers. According to theoretical modlong do they exist? Where do they go, and how do they evolve?

Fuglister estimated that the number of rings arising on each side of the Gulf Stream in a year varies between five and eight. His method was indirect; he based his calculation on two quantities. They were first the difference in volume between the inflow of Gulf Stream water at Cape Hatteras and the outflow of Gulf Stream water at the Grand Banks over the course of a year, and second the variation in the surface area of the Slope Water caused by deviations of the Gulf Stream from its average position.

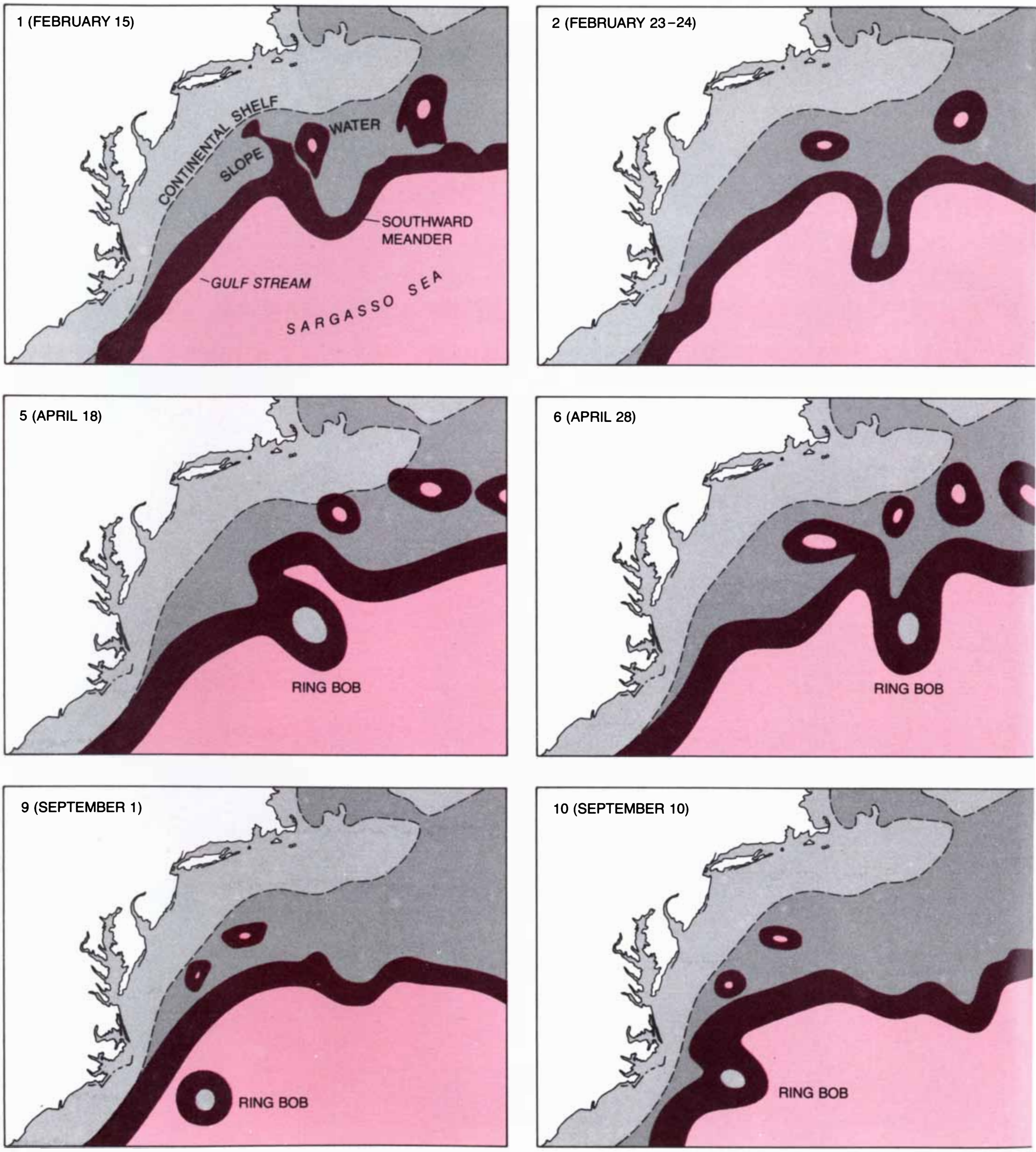
A census of warm-core rings done by J. L. Chamberlin and his colleagues at the National Marine Fisheries Service in Narragansett, R.I., corroborates Fuglister's estimate on the north side of the stream. The census recorded from seven to nine warm-core rings per year for 1975 through 1978. At times during the year no rings were present; at other



els, they can extend downward to the sea floor at a depth of 4,000 to 5,000 meters. Warm-core rings, which occupy the Slope Water, are smaller and shallower. The rings arise in the eastern part of the region shown and move erratically toward the southwest, becoming weaker over several months. At any one time warm-core rings can cover as much as 40 percent of the Slope Water and cold-core rings can cover as much as 15 percent of the Sargasso Sea. The ocean's vertical scale in the drawing has been exaggerated by a factor of about 100.

times six or seven rings were. When many rings were present, they occupied as much as 40 percent of the Slope Water's surface area. No similar census has been attempted south of the stream, partly because the cost of large-scale surveys made from ships and aircraft is high. In addition the tracking of rings by satellites is compromised much of the time by cloud cover. (The satellite observations are made at the infrared wavelengths strongly emitted by warm bodies. The resulting images therefore show the pattern of temperature distribution on the surface of the sea, including the characteristic pattern of rings.)

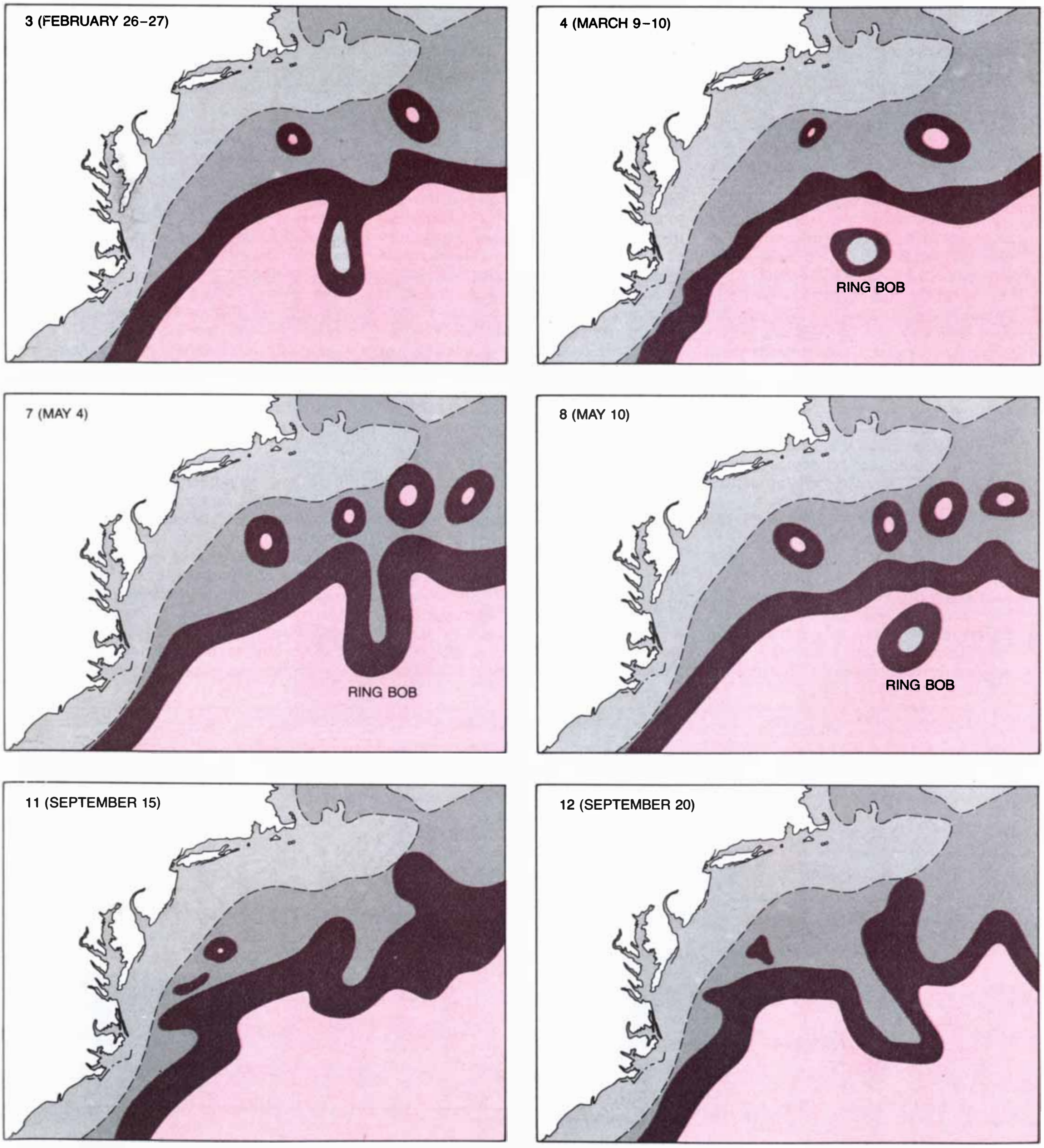
A direct count of cold-core rings awaits the advent of microwave sensors aboard satellites later in this decade. Such sensors will detect the ocean's radiation, and hence its temperature, at



HISTORY OF A RING is exemplified by that of the cold-core ring designated Bob, which formed in mid-February, 1977. First the Gulf Stream meandered southward (*1*). The meander became a loop (*2*); the loop then pinched itself off (*3*) and detached itself from the stream (*4*). Thus it became a ring. Five weeks later it rejoined the stream, moving from the west limb to the east limb of a meander that be-

wavelengths better able to penetrate clouds. They will also measure the height of the sea, a further indicator of the presence of rings. For now Philip L. Richardson of Woods Hole has analyzed records of the temperature pattern in the Sargasso Sea and also the records of a survey of the Sargasso Sea made with ships and satellites in the spring of 1975. He concludes that at any one time about 10 cold-core rings are present.

The youngest and most distinct of those rings are found in a region where rings preferentially arise just south of the Gulf Stream between 75 and 60 degrees west longitude. Older rings are found to the south and the west of that region. In the Sargasso Sea south of 30 degrees north latitude and east of 65 degrees west longitude no rings have ever been detected. This region, however, has



came a warm-core ring (*5, 6*). In May it became detached again (*7, 8*). For several weeks it was more or less stationary. By September, however, it was approaching the Gulf Stream off the coast of North Carolina (*9*). There it coalesced into the stream and disappeared (*10, 11, 12*). The history was inferred by Richardson from infrared images. The closed curves north of the Gulf Stream are warm-core rings.

not been well surveyed. Overall it seems fair to say that cold-core rings occupy between 10 and 15 percent of the Sargasso Sea's surface area.

The number of rings at any one time obviously is determined in part by how long rings persist. This too is an aspect of the life history of rings that can at present be treated in only a qualified manner. Richardson and his colleagues have placed buoys in a number of rings and have tracked the buoys by satellite for as long as eight months. In none of these observations, however, was the entire life of the ring followed. Rings identified in successive infrared images have been known to persist for two years. Moreover, a calculation based on the amount of potential energy contained in a young ring and an estimate of the rate at which such energy is dissipated suggests a lifetime of two to four years. Coalescence into the Gulf Stream often cuts short the life of a ring; thus Richardson estimates that the average life span of a cold-core ring is one to 1.5 years. The probable maximum is three years.

The rings move in complex ways. In addition to their swirling they migrate through the ocean; indeed, the translational movement of the water composing a ring is predicted for depths as great as 1,500 to 2,000 meters in a model devised by Glenn R. Flierl of the Massachusetts Institute of Technology. Once rings are free of the Gulf Stream both the warm- and the cold-core ones tend to move toward the west or the southwest at a mean velocity on the order of five centimeters per second (four kilometers per day). The rate of translation varies: typically there are periods when the ring moves at a velocity higher than 10 centimeters per second and other periods when the ring is stationary.

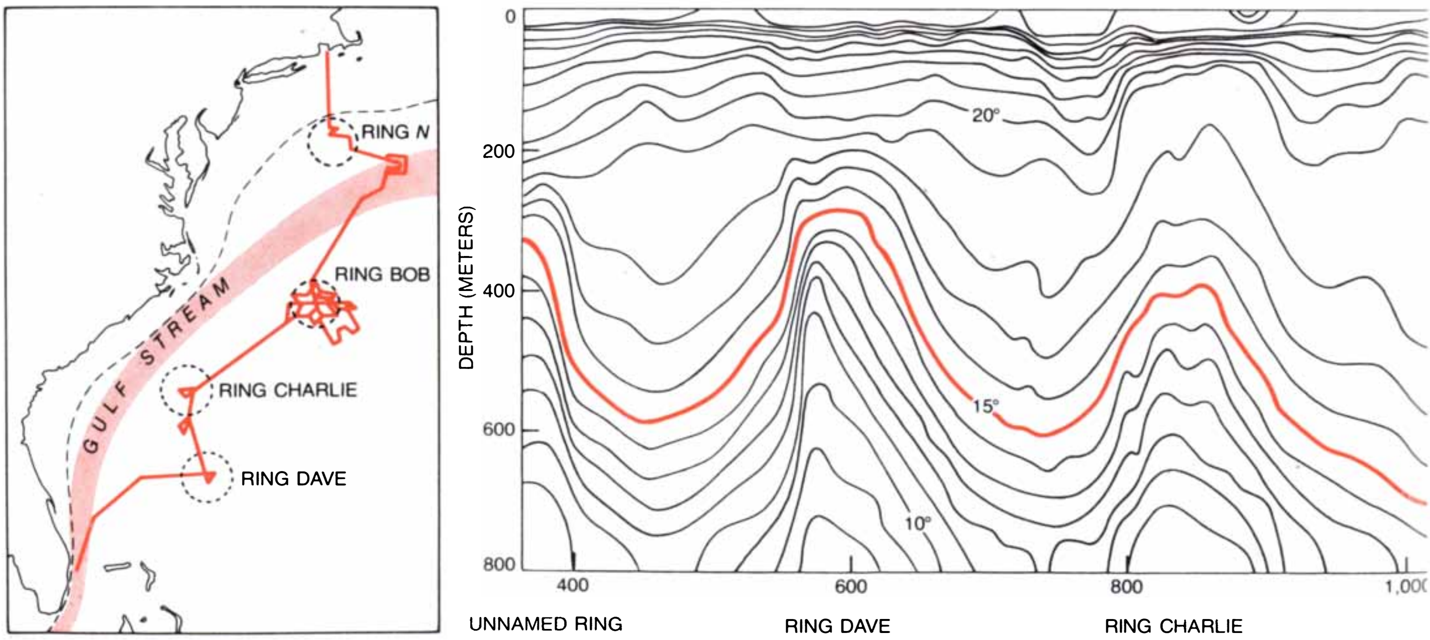
Some rings stay away from the Gulf Stream as they move toward the southwest along an "eddy alley." Eventually they enter a region north of the Bahamas and east of Florida. It is a ring graveyard, a place where rings coalesce with the Gulf Stream and disappear. It seems, however, that most rings interact with the Gulf Stream repeatedly. Such rings become attached to the stream, which draws them back toward the northeast at a velocity as high as 25 centimeters per second. Then, if their coalescence is not complete, they split off, sometimes greatly modified. In some instances they are smaller but more energetic than they were before the interaction. Each such cycle takes from one to three months. With repeated interactions with the Gulf Stream the trajectory of a cold-core ring across the Sargasso Sea can be a set of large clockwise loops.

During the periods when a ring is isolated from the Gulf Stream it gradually decays. For a cold-core ring the process can be detected in the warming of the water in the core, in the increasing salinity of this water and in the deepening of the zone where the oxygen content of the water is at a minimum. (In the northern Sargasso Sea the zone of minimum oxygen lies between 700 and 1,000 meters. In the Slope Water that forms the core of a young cold-core ring it lies between 100 and 200 meters.) The best single indicator of the physical decay of a cold-core ring is the rate at which the isotherms sink. Repeated measurements of the vertical temperature structure in a number of cold-core rings suggest that the rate of sinking is from .3 to .4 meter per day. In general the smaller values are measured in older rings. The sinking isotherms represent a collapsing dome of cold water. They also signal the diminution of the ring's kinetic energy and of the volume of water it transports through the sea.

The physical decay of a ring is accompanied by a profound transformation of its biology. The reason is straightforward. The ecology of the cold-water life in the Slope Water is quite different from that of the warm-water life in the Sargasso Sea. A young ring sustains this difference, but as the ring decays, the environment that favors the existence of the forms of life inside it begins to disappear.

Basically the water of the Sargasso Sea (and the core of a young warm-core ring) supports a large number of tropical and subtropical species. Their populations, however, are relatively small and tend to be similar in size. In contrast, the Slope Water (and the core of a young cold-core ring) is much like the subarctic seas to its north. It supports a relatively small number of species, of which a relative few are dominant. Its biomass, or total content of living matter, is much larger than that of warmer waters.

Here I shall focus on the planktonic forms of marine life. The term embraces the plants (phytoplankton) and animals (zooplankton) that drift in the open sea. In general phytoplankton are single-cell plants that vary in size from five to 75 micrometers. They are the primary producers of the sea: they make organic molecules out of inorganic ones by means of photosynthesis. Thus they contain chlorophyll, the substance that mediates photosynthesis. Zooplankton can be somewhat larger than phytoplankton but generally are no more than four centimeters long. They vary from



TEMPERATURE STRUCTURE of the northwestern Atlantic was recorded on a zigzag course from Miami, Fla., to Woods Hole, Mass., that took the research ship *Endeavor* through the core of five rings, including Ring Bob, in August, 1977. In Ring Bob, a cold-core ring,

single-cell organisms such as protozoans to more complex organisms such as crustaceans and the larvae of many species of fish. They live on phytoplankton and on one another in a highly complex set of relations.

With all of this in mind consider Ring Bob again. When my colleagues and I first studied it in April, 1977, the distribution of chlorophyll suggested striking differences between the concentration of phytoplankton at the center of the ring and that in the surrounding water. In the cold Slope Water core of the ring chlorophyll (and presumably phytoplankton) was highly concentrated; we measured 3.0 micrograms of chlorophyll per liter of seawater. The distribution was largely confined to the uppermost 30 meters of the sea. With distance from the center the concentration decreased, and the maximum concentration came to lie well below the surface at 60 to 70 meters. The pattern was typical of the differences between the Sargasso Sea and the Slope Water.

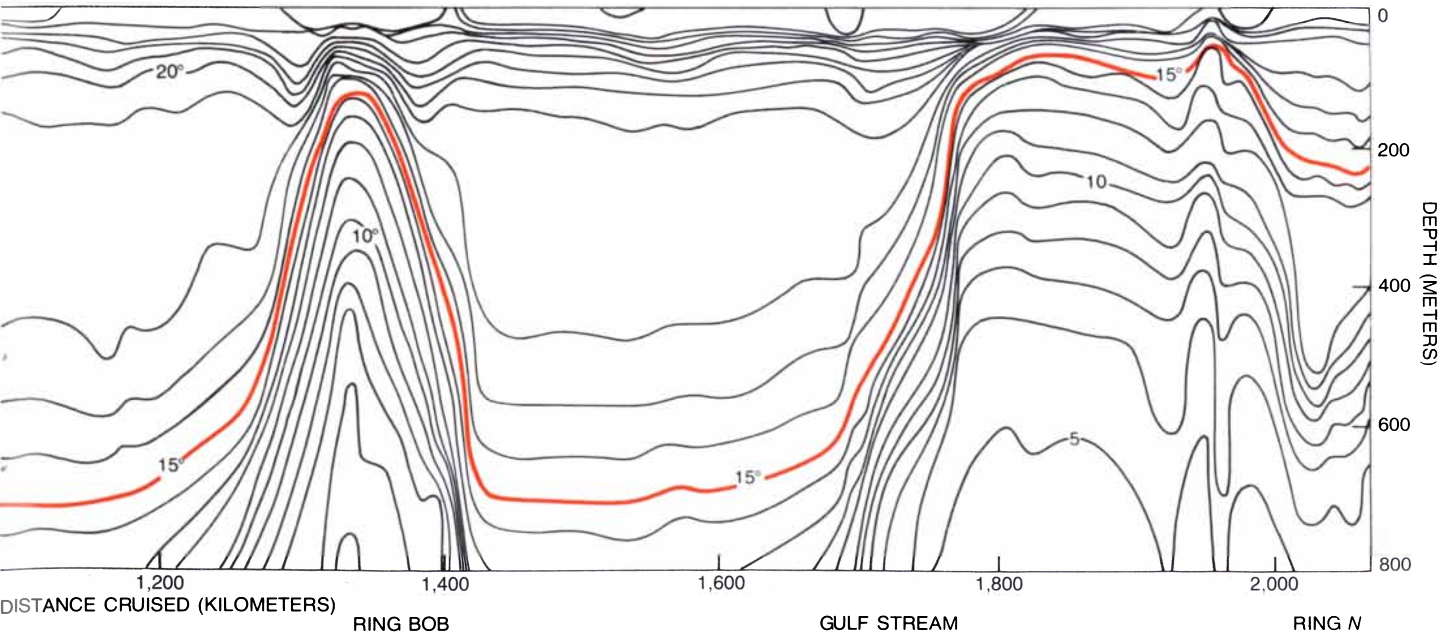
The distribution of zooplankton was somewhat more complex. We had expected to find high concentrations at the center of the ring and a gradual decline at its flanks, so that the concentration at the periphery would match the concentration typical of the Sargasso Sea. The concentration indeed proved to be high at the center. In the Gulf Stream remnant that encircled the core, however, we found a zone in which zooplankton was even less abundant than it was in the surrounding ocean. Apparently the Gulf Stream at the time Ring Bob formed (two months before we arrived in it) had a concentration of zooplankton lower than that of the Sargasso Sea, and this scarcity of life persisted.

The patterns of overall abundance for phytoplankton and zooplankton in April, 1977, reflect patterns in the abundance of particular species. For example, warm-water species such as *Limacina inflata,* a marine snail that lives in the tropical and subtropical waters of all the oceans of the earth, were found in abundance only outside the young Ring Bob. Conversely, species of zooplankton that live in the cold Slope Water were confined to the core of the ring. Such species included *L. retroversa,* another marine snail. They also included *Nematoscelis megalops,* a small, shrimplike crustacean, and *Pareuchaeta norvegica,* a small crustacean notable for its lack of compound (that is, image-forming) eyes. The differing vertical distributions we measured for these cold-water species in the core of Ring Bob in April, 1977, reflected the depths they normally inhabit in the Slope Water.

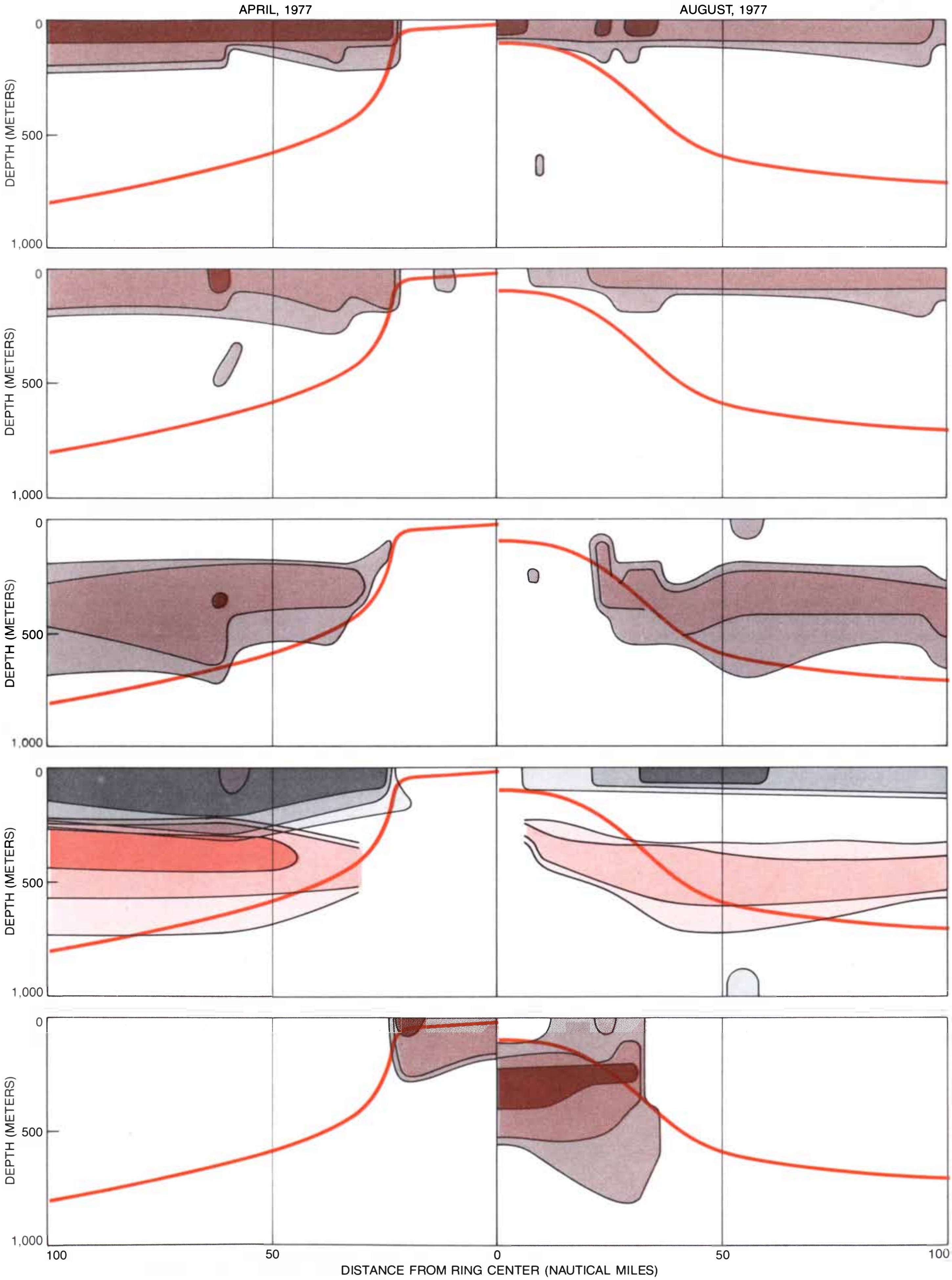
On our return to Ring Bob in August, 1977, we found what we have found in other decaying cold-core rings: the species of the ring's cold-water planktonic community had declined in number. If the ring were to continue in its isolation from the Gulf Stream, they would no doubt disappear from it altogether. Our studies show that the time required for extinction varies with the species; it ranges from a few months to as long as 1.5 years. Our studies also show that the diminution in cold-core plankton gives rise to a curious distribution of life in the water column. Often the top 200 meters of a decaying cold-core ring has significantly less biomass than the top 200 meters of the surrounding Sargasso Sea. The next 800 meters is significantly richer than the surrounding Sargasso Sea. It is rich enough for the water column in the ring to retain a greater abundance of zooplankton than the surrounding Sargasso Sea for at least a year after it forms.

Two processes that operate simultaneously appear to be responsible for this curious distribution. In the first place the stirring of the sea by winds and the heating of the sea by the sun affect the uppermost part of the sea. Thus the decay of the conditions in a cold-core ring that favor cold-water life is fastest near the surface. For one to three months after a cold-core ring forms, the uppermost part of its core may remain little different from Slope Water. Then a marked transformation begins. The abundance of cold-water phytoplankton decreases and the size of the organisms diminishes. In Ring Bob the change was dramatic: between April and August, 1977, the concentration of chlorophyll in the core decreased by a factor of eight. It is little wonder that herbivores such as *L. retroversa* that live near the surface seem to be more immediately susceptible to the decay of their cold-water environment in an aging cold-core ring than the deeper-living omnivores such as *N. megalops* and the deeper-living carnivores such as *P. norvegica.*

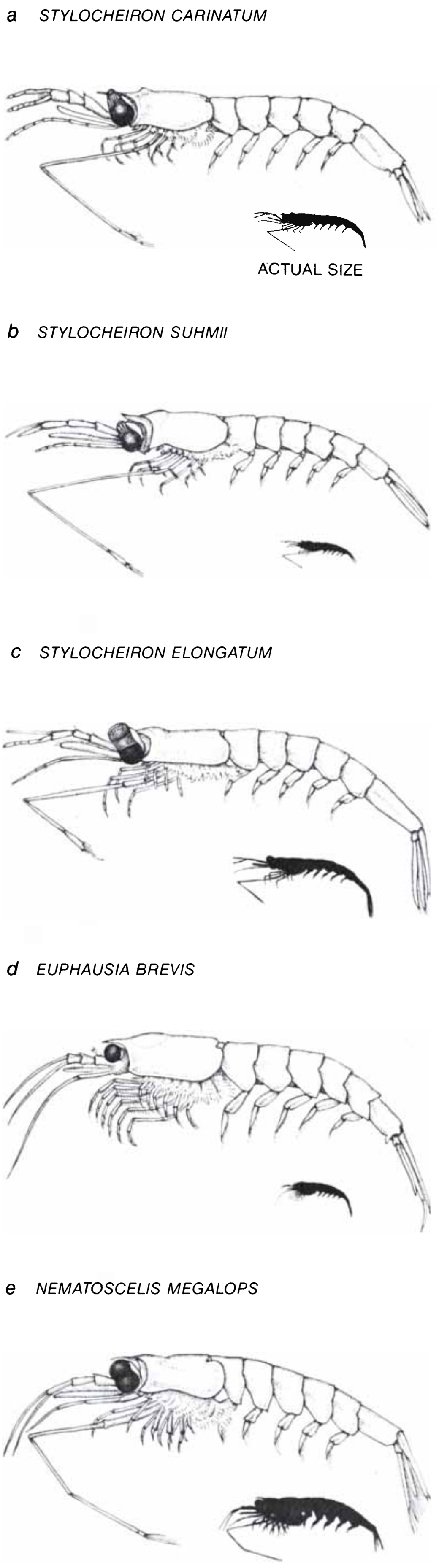
The second reason for the concentration of planktonic life well below the surface in a decaying cold-core ring is that a number of Slope Water species appear to move downward in the water column in an attempt to stay in their preferred environment. The shrimplike species *N. megalops* demonstrates this shift quite well. It is endemic to the Slope Water; in our samplings of the northwestern Atlantic we have never found it south of the Gulf Stream except in cold-core rings. In the Slope Water the bulk of the population of *N. mega-*



the isotherm marking the depth of seawater whose temperature is 15 degrees Celsius (*color*) was elevated almost 600 meters above its level in the surrounding Sargasso Sea. Conversely, in Ring *N*, a warm-core ring, the isotherm was depressed 150 meters below its local level.



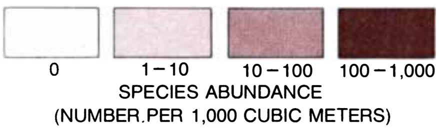
APRIL, 1977
AUGUST, 1977
DEPTH (METERS)
0
500
1,000
100
50
0
50
100
DISTANCE FROM RING CENTER (NAUTICAL MILES)



lops is above a depth of 300 meters both day and night.

In young cold-core rings the distribution of the population is the same. By the time the ring is six months old, however, a substantial number of the organisms live below 300 meters and individual organisms are found as deep as 800 meters. In still older rings most of the population live below 300 meters. The deepening vertical distribution of *N. megalops* in a decaying cold-core ring coincides with the sinking of isotherms and also the sinking of isohalines, which outline the salinity of the water. Indeed, the central 50 percent of the adult population of *N. megalops* appear to seek out water with a temperature within two degrees of 10 degrees C. in the Slope Water or in a ring.

The downward movement of *N. megalops* means that the animals become separated from the concentration of food closer to the surface; hence the downward movement coincides with a deterioration in the physiological and biochemical condition of the species. The physiological decline is shown by a reduction in the animals' respiration rate, the biochemical decline by an increase in their water content and a reduction in their content of lipids, carbon and nitrogen. Furthermore, we have found in older cold-core rings a reduction in the growth rate of individuals of *N. megalops* compared with their growth rate in the Slope Water. We have also found a reduction in the proportion of males to females. In sum, it appears that the individuals of *N. megalops* in a decaying cold-core ring are starving as food levels become inadequate for their growth and reproduction. The oldest cold-core ring we have sampled had an age of 17 months. In it we found no *N. megalops* individuals, even though the ring had held them earlier.



BIOLOGY IN A RING is transformed as the ring ages because conditions such as temperature, salinity and oxygen content change. Here the abundance of five small, shrimplike crustacean species in the core of Ring Bob and in the surrounding Sargasso Sea is shown for April, 1977 (when the ring was two months old), and August, 1977. In the four intervening months *Stylocheiron carinatum* (*a*), a warm-water species that lives near the surface, entered the ring. Indeed, it reached the center. Meanwhile *S. suhmii* (*b*), another warm-water species near the surface, penetrated to a lesser degree. *S. elongatum* (*c*), a warm-water species that lives well below the surface, penetrated hardly at all. *Euphausia brevis* (*d*), a warm-water species that migrates upward at night and downward during the day, showed equal penetration in day and night nettings. Color denotes its daytime abundance; black denotes its nighttime abundance. For the other species the day and night abundances are combined. *Nematoscelis megalops* (*e*) is a cold-water species. It was found only inside the ring. Between April and August it moved deeper in the ring; apparently it sought to maintain the conditions that favor its survival. The curve in each chart is the isotherm for 15-degree seawater.

Do other species starve too? It seems likely, and yet for every planktonic species there is the further possibility that the motion and decay of a ring can cause individuals to be cast out into the surrounding Sargasso Sea. The model devised by Flierl at M.I.T. strongly suggests that as a ring moves across the Sargasso Sea only part of the ring's water column is actually carried along. According to the model, the trapped part of the ring is largest at the surface and gets smaller with increasing depth until at a depth of 1,000 to 1,500 meters it vanishes. Thus individuals of a species such as *N. megalops* that move deeper in the water column of a decaying ring may end up below the trapped part of the ring, and the ring may leave them behind. Our own analysis of the distribution of cold-water species such as *N. megalops* and *P. norvegica* at the flanks of cold-core rings suggests that some individuals are indeed removed by physical processes.

As the various cold-water species begin to find the conditions in a decaying cold-core ring less suitable for their survival, some of the warm-water species in the surrounding Sargasso Sea begin to find the conditions attractive. Near the surface the cold-water phytoplankton give way to some of the smaller, less abundant forms typical of the Sargasso Sea. Meanwhile the cold-water zooplankton give way to warm-water zooplankton. In fact, some of the species of warm-water zooplankton in the Sargasso Sea exploit the changing conditions in cold-core rings of intermediate age to such an extent that they become more abundant there than they are elsewhere in the Sargasso. The warm-water sea snail *L. inflata* is an example. In April, 1977, its abundance in the core of Ring Bob was small. Four months later its abundance there had increased by a factor of 300. At the same time its abundance in the Sargasso Sea had decreased by a factor of .7, apparently as part of a normal seasonal cycle.

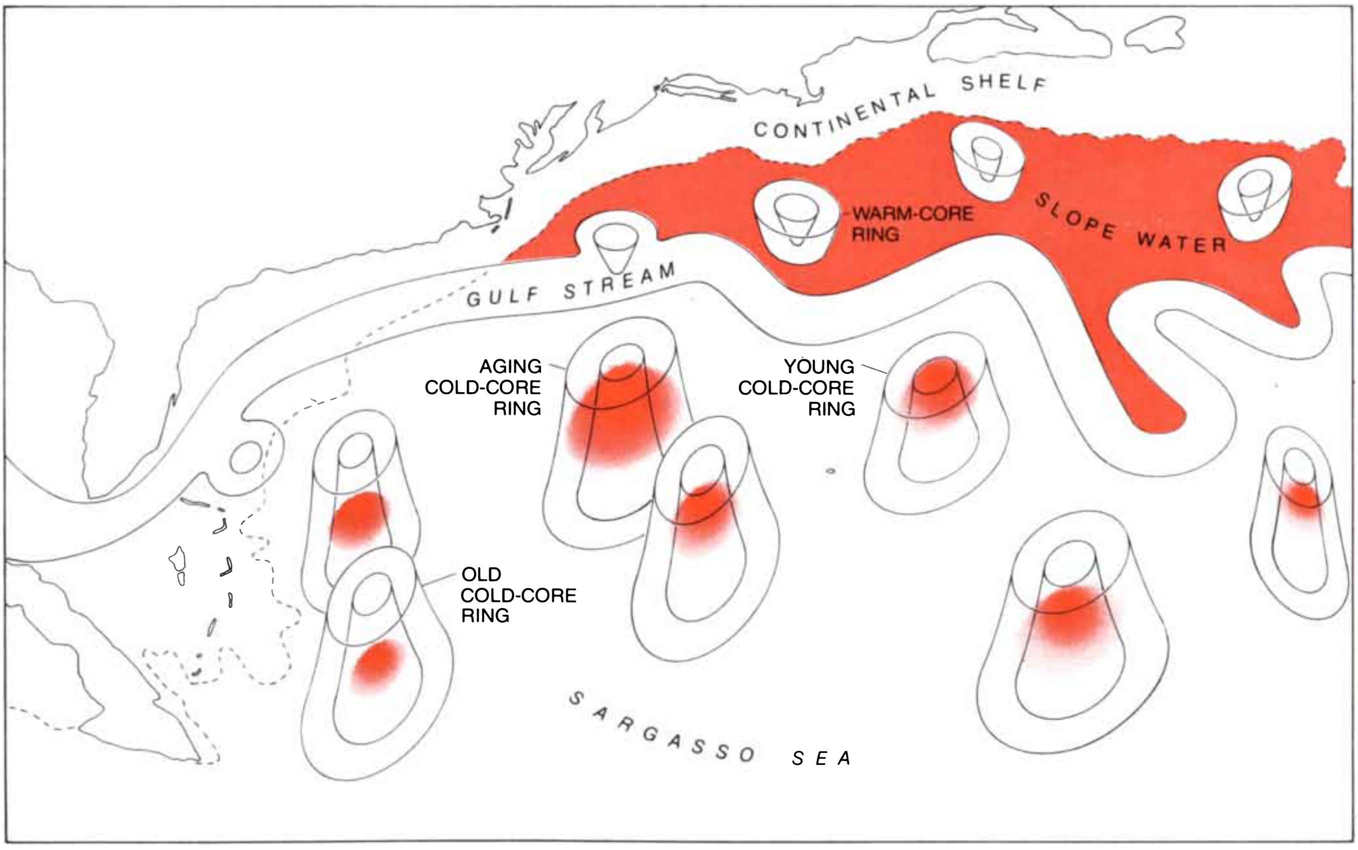
Other warm-water species enter a decaying cold-core ring more gradually. In this regard the zooplankton of the shrimplike genus *Stylocheiron* may be quite typical of the warm-water species that occupy the ecological niches left vacant in a decaying cold-core ring by the waning Slope Water species. Five species of *Stylocheiron* are often found in our samples from the northwestern Atlantic. Unlike some other shrimplike plankton, they show little or no tendency to move upward or downward diurnally. On the other hand, they exhibit a remarkable degree of vertical stratification. In the Sargasso Sea we have consistently found the center of the vertical distribution of *S. elongatum* to lie between 300 and 400 meters, that of *S. affine* to lie between 200 and 250 meters, that of *S. abbreviatum* to lie between 100 and 200 meters and that of *S. carinatum* and *S. suhmii* to lie between the surface and 100 meters.

In very young cold-core rings none of these species is found. Ring Bob, for example, lacked all five of them in April, 1977. Four months later one of them, *S. carinatum,* had reached the center. The others had not, including *S. suhmii,* which lives at much the same depth as *S. carinatum.* After another two or three months all of them had probably arrived there. Cold-core rings we have studied when their age was from six to nine months include the five species at depths much like those they occupy in the Sargasso Sea. Evidently the passage of half a year brings sufficient modification of the uppermost 300 meters of a cold-core ring for many warm-water species to conduct their normal life functions there.

Warm-core rings form by a process that is in essence the mirror image of the process that gives rise to cold-core rings. Can one therefore expect the biological transformations in a decaying warm-core ring to be the reverse of the transformations I have described for cold-core rings such as Ring Bob? The answer is uncertain; the biology of warm-core rings has not yet been studied in detail. Still, the physiological and biochemical status of a species of zooplankton depends on the quality and quantity of the available food and also on the conditions in the sea that affect the organism's rate of metabolism. The warm-water species expatriated from the Sargasso Sea in the core of a warm-core ring probably encounter an increasing abundance of food as the ring they are in decays, but they must endure the lower rate of metabolism that results from a decreasing water temperature.

I have tried to show in this article that rings deserve study in themselves. More than that, they offer an opportunity to study the mechanisms that regulate ocean ecologies. After all, the formation, decay and ultimate disappearance of a ring make it possible to witness what happens as a large expatriated population of marine life is subjected to changes in hydrography and biology beyond its normal experience. The same changes take place continuously along the margin of a species' home range. There, however, it is difficult to sort out which particular changes are important. In rings the various changes (in temperature, salinity and so on) appear to become uncoupled: they happen at different rates, so that an observer can hope to learn which ones are strongly related to an alteration in the abundance of a species. The opportunity is unprecedented in population biology. An equivalent terrestrial experiment would require that an area of land the size of Massachusetts, New Hampshire, Rhode Island and Vermont, with all its soil, forests, meadows, rivers, lakes and animals, be monitored over a two-year period as it was moved toward the southern U.S.

I have, of course, been discussing only the northwestern Atlantic. Rings are also found in the oceans traversed by currents such as the Kuroshio, the Agulhas, the Brazil, the East Australia and the Somalia. Other forms of eddies are found near the California current and in the equatorial seas. It is quite likely that wherever strong ocean currents flow they generate some form of eddies. It is quite likely as well that the eddies affect not only the hydrography of the sea but also the biology. New multidisciplinary studies of oceanic eddies are already in progress; others are being planned. One can expect that such studies will yield new perceptions into the forces that govern the oceans.



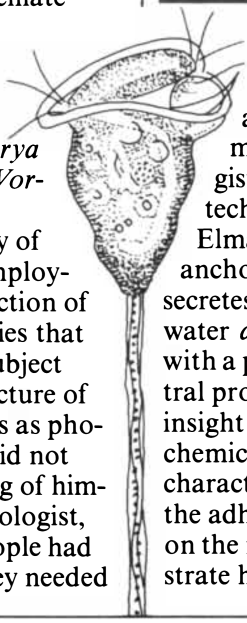
HYPOTHETICAL DISTRIBUTION OF N. MEGALOPS, indicated by shades of color, reflects the species' propensity for water whose temperature is about 10 degrees C. *N. megalops* is abundant north of the Gulf Stream. South of the stream its life is sustained only in young rings. In aging rings it seeks increasing depth. In old rings it becomes extinct. The distribution of *N. megalops* throws doubt on the earlier hypothesis that conditions are uniformly suitable or unsuitable for the survival of a given species throughout wide expanses of the sea.

Vorticella's secret

Is it worth knowing?

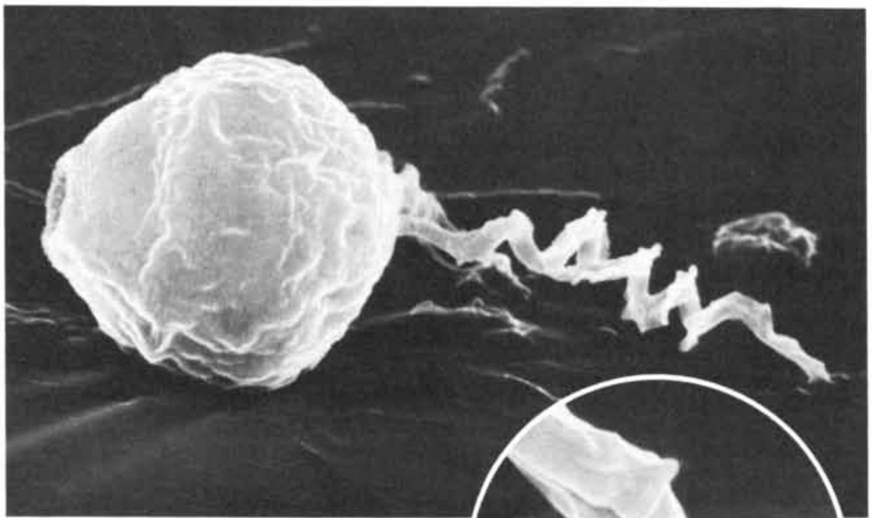
Jim Elman has convinced us that it would be smart to pay him to do what he would want to do even if he had no job. It is not a typical story.

At the edge of the Genesee River, just outside his office at the Kodak Research Laboratories, Elman engages in an activity he began as a high school sophomore. He is collecting material to start yet another culture of *Vorticella* sp., an ubiquitous ciliate protozoan usually pictured thus:



When Elman finished up a senior project on SEM studies of *Tokophrya* (a protozoan similar to *Vorticella)* at St. Michael's College of the University of Toronto, he accepted employment at Kodak in the section of the Research Laboratories that works on adhesives, a subject pertinent to the manufacture of such multilayer products as photographic film. Elman did not thereupon cease thinking of himself as a serious protozoologist, even if the adhesives people had to be made to realize they needed a serious protozoologist.

Elman solved this problem without—at least not immediately—diverting himself from his assigned duties. The adhesive disc on which *Vorticella* stands provided his solution. Most serious protozoologists concern themselves more with activity within



the cell itself than with the stalk that attracts attention at first sight of this anchored creature. But most serious protozoologists do not work for high-technology companies.

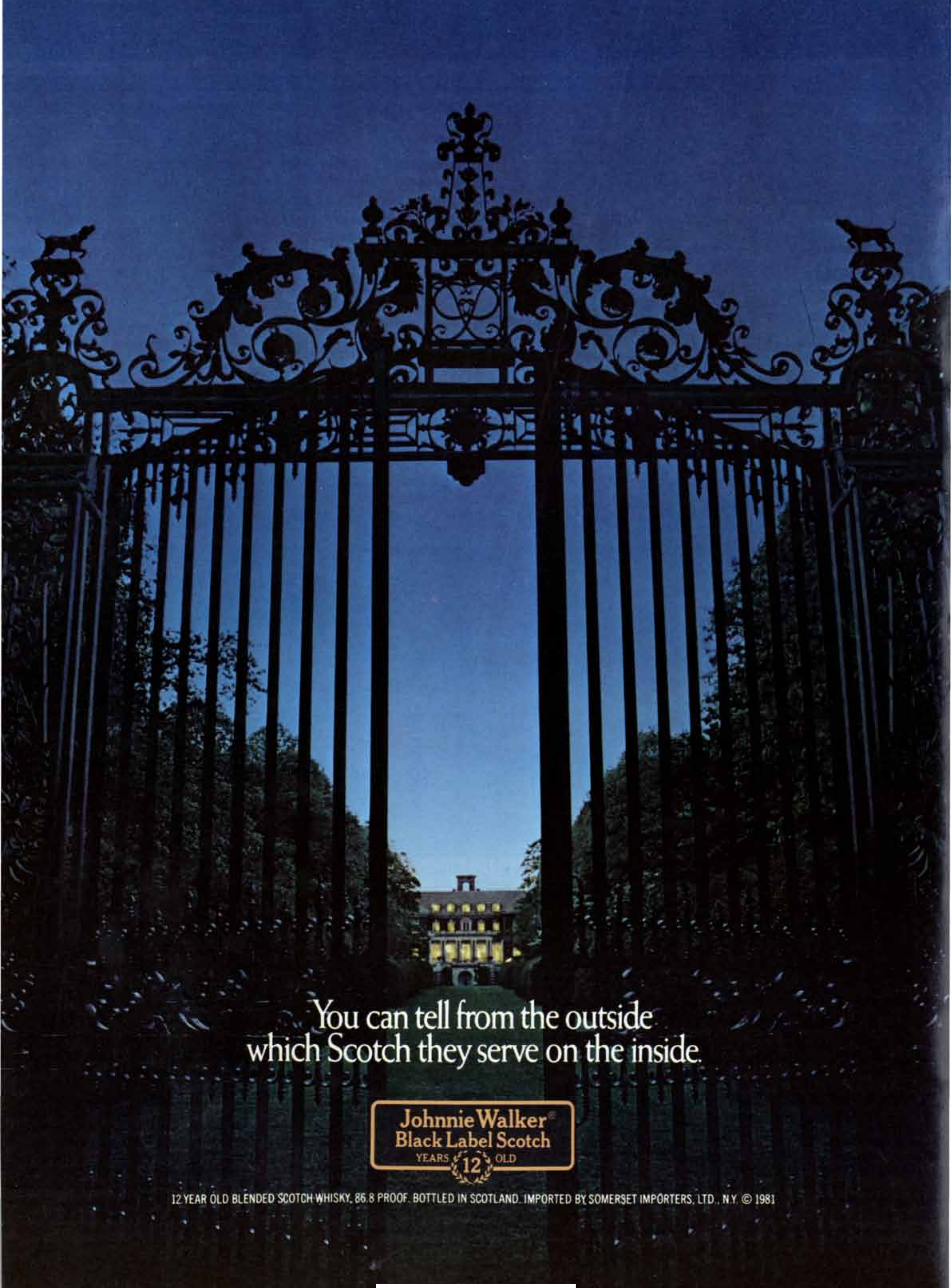
Elman showed us that for its anchor this one-celled creature secretes an instant-acting underwater *adhesive*. He then came up with a plan for two different spectral programs to provide some insight into what's going on there chemically. And he demonstrated characteristic configurations of the adhesive buildup, depending on the nature of the synthetic substrate he presented to *Vorticella*.

Additionally, Elman pointed out that the infrared spectrum of the adhesive used by the protozoan *Stentor* for anchoring *its* foot bears a remarkable resemblance to that of ionized polyglutamic acid.

Elman talks like an industrial research chemist because now he is one, in addition to being a protozoologist.

Past experience has shown us that the rare instances where high-grade curiosity pays off make science a winning game economically for both producer and consumer of products still far over the horizon.

Kodak


You can tell from the outside
which Scotch they serve on the inside.
Johnnie Walker
Black Label Scotch
YEARS 12 OLD
12 YEAR OLD BLENDED SCOTCH WHISKY, 86.8 PROOF. BOTTLED IN SCOTLAND. IMPORTED BY SOMERSET IMPORTERS, LTD., N.Y. © 1981

SCIENCE AND THE CITIZEN

Ground Swell Zero

Over the years the most vocal and persistent opponents of nuclear weapons have been drawn mainly from the scientific community. In recent months clergymen, physicians and citizens organized at the local political level have taken up the argument. They have questioned the morality of possessing nuclear arms, warned of the medical consequences of nuclear war and proposed negotiating a permanent moratorium on the manufacture of nuclear arms. In particular they have criticized the idea that a nuclear exchange could be limited to a single theater of war, that it could be "won" by destroying most of an enemy's retaliatory capacity, or indeed that it could lead to anything other than global holocaust.

The individuals and institutions in the current surge of protest are in general more conservative and closer to the social mainstream than those of the anti-Vietnam-war movement or earlier protests against nuclear weapons. Among Catholic clergymen, for example, opposition to the war in Vietnam came mainly from a small minority of priests who were at the edge of the church hierarchy. Recently, however, the bishops have become the focus of the opposition to nuclear weapons. At the National Catholic Bishops Conference in November, 240 out of 280 bishops endorsed a resolution calling for an end to the development of nuclear weapons. Some bishops have gone beyond this relatively cautious position: 29 have signed a statement characterizing the possession of nuclear weapons as morally unjustifiable.

The Catholic Church is only one of several churches whose leaders have recently spoken in opposition to the accumulation of nuclear arms. The General Assembly of the Presbyterian Church has said that halting the proliferation of nuclear weapons is the most important issue outside the church itself for the institution in the 1980's. Spokesmen for the American Baptist, Episcopal, United Methodist and Greek Orthodox churches have also criticized the policy of the Administration.

The opposition from physicians has focused on the catastrophic effects of a nuclear attack on the U.S. population. In December the American Medical Association passed a resolution that asked its members to apprise Government officials of those consequences. Also in December the Harvard Medical School announced a course on "The Health Aspects of Nuclear Weapons and War."

Herbert L. Abrams of Harvard, who is one of the lecturers in the course, and his colleague William E. Von Kaenel recently examined the effects of a nuclear attack in *The New England Journal of Medicine.* They employed a projection of the results of an attack on the U.S. developed by Federal officials for the purpose of civil defense. Designated CRP-2B, the projection assumes the detonation of 6,559 megatons over targets in the U.S. Some 86 million people would be killed immediately by explosions and heat; 34 million more would be severely injured. There would be 50 million additional deaths before the level of radiation had fallen sufficiently for people to leave underground shelters safely. Only about 60 million people would survive the "shelter" period relatively uninjured.

The damaging effects of nuclear explosions are obvious, but the epidemiological catastrophe that would follow is less readily imagined. Exposure to radiation weakens the protection afforded by the immune system. Crowding in underground shelters would provide ideal conditions for the transmission of pathogenic agents. The collapse of sanitation services would allow populations of disease-bearing organisms such as insects and rats to grow unchecked. Moreover, the available medical resources would be greatly reduced. Physicians are concentrated in cities. Their survival rate would therefore be lower than that of the population at large; only about 20 percent would survive an attack, so that some 79,000 physicians would remain to treat more than 30 million injured patients. Most of the country's diagnostic and therapeutic facilities would be damaged or destroyed. Under such conditions food poisoning, dysentery, gastroenteritis and influenza followed by pneumonia would each affect a substantial proportion of the remaining population, as would bubonic plague contracted from rodents. From 20 to 25 percent of the 60 million healthy survivors might die from infectious diseases.

The specialized perspectives of the clergy and physicians have been reinforced by political organizing at the grass roots. The program of much of the current activity has been provided by a resolution calling on the governments of the U.S. and the U.S.S.R. to halt the testing and deployment of new nuclear weapons. The moratorium could then be made permanent by the negotiation of a treaty. The resolution, similar to the one endorsed by the Catholic bishops, was written by a group called the National Freeze Committee, which intends to have it placed on state and local ballots.

So far 25 towns in New England have adopted the statement in town meetings or referendums. The legislatures of Oregon, Massachusetts and New York have adopted the freeze resolution in a nonbinding form. The committee is currently asking legislators in New Jersey and other states to have the resolution placed on the ballots in November. The most ambitious effort is in California. In order to have the resolution put on the ballot there some 346,000 signatures of registered voters must be gathered by April 23. If the signatures are collected and the resolution is passed, the governor of California will be required to write a letter to the president asking him to implement the goals of the resolution. According to the organizers of the movement, when enough local support has been amassed, the campaign will move to Washington, where an effort will be made to have Congress adopt the resolution.

Image of a Burster

Of the objects in space that defy the slow rhythms of the stars by emitting short pulses of intense electromagnetic radiation, the ones most recently discovered are the gamma-ray bursters. They broadcast gamma rays (the most energetic form of electromagnetic radiation) in bursts that typically last for a few seconds. The bursters were discovered with satellites carrying gamma-ray sensors in earth orbit, and when three or more satellites detect the same burst, its direction can be determined by triangulation. The network of satellites has determined the positions of several bursters to within a patch of the sky roughly one arc-minute across. In these patches, however, no object detectable at optical wavelengths has been identified as the one that emits the gamma rays.

Bradley E. Schaefer of the Massachusetts Institute of Technology now reports what he terms a "probable optical counterpart" of a gamma-ray burster. He describes his findings in a letter to *Nature.* Schaefer had decided to look for optical counterparts at three "precisely known burst positions." To do so he searched the archives of the Harvard College Observatory to see whether something unusual could be found on photographic plates that take in those parts of the sky. On a set of six plates made at the observatory's station in Bloemfontein, South Africa, on November 17, 1928, he succeeded. Each plate shows the same area of the sky: a part of the southern constellation Sculptor. Each one includes the patch of the sky some eight degrees from the south pole of the Milky Way galaxy in which a gamma-ray burst was detected half a century later, on November 19, 1978. On the fourth plate a tenth-magnitude star is present in the patch; the star does not appear on the other plates.

Each plate was made by a 45-minute

exposure; hence the presence of the star on only one plate implies that its optical brightness was short-lived. Indeed, an analysis of the trace the star left on the single plate suggests to Schaefer that its brightness lasted for less than 10 minutes. The images of the other stars on the plate are elongated by faint trails 17 arc-seconds long. Presumably the trails were caused by the telescope's imperfect tracking of the sky's rotation with respect to the earth. The evanescent star has no such trail.

Is the image merely a flaw in the photographic emulsion or a record of some event in the telescope? Schaefer thinks not. The evanescent star shares with the other stars on the plate an asymmetry that gives it a slight elongation toward the center of the plate. The asymmetry, called coma, results from the telescope's imperfect optics; thus it suggests "that the light that formed the burst image followed the same path as normal starlight." Schaefer notes that a meteor might conceivably have made the burst image. He calculates, however, that the probability of a meteoroid's traversing a patch of the sky one arc-minute in size over a period of 45 minutes is one in 100. The probability of its approaching the earth head on, so that it would leave only a dot of brightness on the plate, is one in 100 million. The probability of both events jointly is one in 10 billion.

In spite of Schaefer's deductions the precise duration of the optical burst in 1928 is not known. Schaefer proposes that the burst might plausibly have lasted for one second and momentarily given the burster the brightness of a third-magnitude star. It would have made a visible flash in the sky. Between bursts it must be much fainter. As a class of celestial objects the gamma-ray bursters are thought to be distributed in the disk of our galaxy. On that hypothesis the burster of November 19, 1978, can be no farther from the solar system than one kiloparsec, or 3,260 light-years; otherwise its location near the galactic south pole (and hence in a direction nearly perpendicular to the plane of the disk of the Milky Way) would place it well outside the disk.

The limit of one kiloparsec allows a further inference. Two photographic plates made recently by Martha H. Liller of the Harvard College Observatory with the four-meter telescope at the Cerro Tololo Inter-American Observatory seem to include a star of about the 23rd magnitude at the position of the burster of November, 1978. Suppose the star is one kiloparsec away. Its absolute or intrinsic brightness is then about the 13th magnitude. It cannot be a normal star; it must be a fainter object or a system of fainter objects, such as white dwarfs, neutron stars, black holes or—more prosaically—the coolest, least massive and therefore the dimmest stars, the ones called very late *M* main-sequence stars. Suppose too the star is the gamma-ray burster; then in 1928 it climbed from the 23rd to the third magnitude. In about one second, perhaps, it became 100 million times brighter.

Zigzag DNA

The double helix described in 1953 by Watson and Crick, the double helix that has become an icon of molecular biology, is but one form of DNA, albeit the form assumed by almost all the DNA of living things. It is called *B*-DNA. Under certain conditions DNA can take on other forms: the *A* form when it is dehydrated, the *C* form in certain solvents. In all three forms the DNA consists of two strands of nucleotides twisted around each other, with the backbones of the chains tracing a smooth, right-handed helix.

Several years ago Alexander Rich of the Massachusetts Institute of Technology and his colleagues discovered yet another form of DNA. It too is a double helix, but the strands are twisted the other way: the helix is left-handed. Moreover, the backbone follows a zigzag course, leading Rich to call the new form *Z*-DNA. After two years of investigation a team headed by Rich has now established that *Z*-DNA is not just a laboratory curiosity. It is present in the chromosomes of the fruit fly *Drosophila melanogaster* and presumably in the chromosomes of other organisms; it may have a role in the regulation of gene activity. The discovery is reported in *Nature* by Rich, his associates Alfred Nordheim and Achim Möller, Mary Lou Pardue of M.I.T. and Eileen M. Lafer and B. David Stollar of the Tufts University School of Medicine.

The original observation of *Z*-DNA came when Rich's group was trying to define the precise structure of ordinary *B*-DNA by X-ray crystallography. To do that they had to make good crystals of DNA. Each of the four nucleotides of DNA consists of a sugar molecule, a phosphate group and one of four bases: the purines adenine (A) and guanine (G) and the pyrimidines thymine (T) and cytosine (C). Bonds between complementary bases link the two strands; A pairs with T and G pairs with C. Uniformity of structure generally makes for better crystals, and so the group worked with a synthetic double helix just six nucleotides long. Each strand consisted of alternating C and G nucleotides: CGCGCG in one strand paired with GCGCGC in the complementary strand.

The molecule yielded good crystals, and eventually crystallographic analysis revealed the crystals' structure. The structure was not that of *B*-DNA, however, but that of *Z*-DNA. Soon the left-handed DNA was observed in other laboratories, but only under certain conditions. Apparently only chains of alternating purines and pyrimidines can form *Z*-DNA. Moreover, the DNA is stabilized in the *Z* conformation only in a solution with a high salt concentration or when the DNA is brominated or methylated, that is, when a bromine atom or methyl group (CH_3) is attached to a base. Could such an unstable form of DNA have a biological role?

Rich and his colleagues decided to look for biological *Z*-DNA by making antibodies to the *Z* form and then probing animal cells to see if the antibodies would bind to any cellular DNA. The immune system ordinarily makes antibodies only to foreign substances and not to a normal component of the self such as DNA. Nevertheless, Rich and his colleagues were able to raise antibodies by injecting rabbits with short brominated double helixes of alternating G and C nucleotides. The rabbits made antibodies that were highly specific for *Z*-DNA and did not react at all with the *B* form.

After purifying the rabbit antibody and demonstrating its specificity for *Z*-DNA the investigators were ready to probe a biological material for the presence of *Z*-DNA. They chose the polytene chromosomes of a larva of *D. melanogaster.* Polytene chromosomes are found in the salivary gland of certain insects, in cells that keep replicating their chromosomal DNA but do not divide. Instead of separating to provide the chromosomes for daughter cells, the replicated chromosomes line up next to one another in perfect register. More than 1,000 threadlike chromosomes, thus aligned, make up a single, thick polytene chromosome; the structure of the chromosome is thereby magnified and made visible under the light microscope. What is seen is a pattern, like that of a tiger's tail, of dark bands alternating with lighter "interbands." For each chromosome the banding pattern is quite constant from cell to cell and from one individual to another.

The workers treated polytene chromosomes with the rabbit anti-*Z* antibody and incubated the material so that the antibody could bind to any *Z*-DNA in the chromosomes; unbound antibody was washed away. They made the bound rabbit antibody visible by adding a goat antibody specific for the rabbit one; the goat antibody was conjugated to a fluorescent dye, so that it glowed under ultraviolet illumination. The anti-*Z* antibody could be seen to bind to the chromosomes in a distinct and reproducible segmented pattern. By comparing ultraviolet micrographs with phase-contrast micrographs that reveal the bands and interbands of polytene chromosomes, the investigators established that the fluorescent segments are the interbands and not the bands. There must therefore be *Z*-DNA in the interbands. It is conceivable that some *Z*-DNA is also present in the bands but is not visibly stained by the antibodies. At the very least the results show that *Z*-DNA is either more abundant or more accessible in the interbands.

What is the function of the *Z*-DNA in *Drosophila* chromosomes? Work with synthetic DNA's has demonstrated that purine-pyrimidine sequences can undergo transitions between the right-handed *B* conformation and the left-handed *Z* conformation depending on the salt concentration or on chemical modification of the bases. The reversibility suggests that *Z*-DNA has some regulatory role. For example, some nucleotide sequences can be stabilized in the *Z* form by being methylated. Methylation and demethylation are thought to be important in controlling the activity of genes. One possibility is that when certain control sites are stabilized in the *Z* conformation by methylation, a regulatory protein binds to the site and keeps a gene turned off; demethylation might switch the site to the *B* conformation, causing the regulatory molecule to let go. Rich and his colleagues are testing the proteins in *Drosophila* salivary-gland cells and other cells to see whether any of them bind to *Z*-DNA.

Most-wanted List

The study of elementary particles often takes the form of a dialogue between theory and experiment. As matters stand now the theorists have had their say: several new particles and phenomena have been predicted. Word from the experimentalists is eagerly awaited. Dozens of experiments intended to test the predictions are under way, but answers have been slow in coming.

The prediction in which most theorists have the greatest confidence is the existence of three particles called intermediate vector bosons, which are supposed to transmit one of the basic forces of nature. It has been more than 15 years since a theory incorporating three very massive intermediate vector bosons was proposed, but experiments at energies high enough to create them are only now becoming possible. An account of the experiments appears elsewhere in this issue of SCIENTIFIC AMERICAN (see "The Search for Intermediate Vector Bosons," by David B. Cline, Carlo Rubbia and Simon van der Meer, page 48).

Another family of particles has been on the experimentalists' agenda for about five years: it consists of particles that include in their structure a new kind of quark, designated top. Quarks are the constituent elements of the proton, the neutron and more than 100 other known particles. All ordinary matter can be constructed out of just two kinds of quark, called up and down. A second pair of quarks, named strange and charmed, are components of more massive particles created in accelerator experiments. In 1977 a still heavier particle was discovered; there is strong evidence that it incorporates a fifth quark type, which has been given the name bottom. Since the other four quarks come in pairs, physicists immediately began searching for the partner of the bottom quark.

Recent unified theories of the forces that act between elementary particles lend at least provisional support to the idea that there is a sixth quark, but they provide little guidance on where to look for it. The member of the top family that should be easiest to identify would be made up of a top quark bound to a top antiquark. Initial estimates were that its mass (expressed in energy units) might be about 27 GeV, or 27 billion electron volts. That energy region was subsequently searched but nothing new was found. About a year ago Sheldon Lee Glashow of Harvard University suggested instead that the mass of the top-antitop particle is 38 ± 2 GeV. A colliding-beam device at the German Electron Synchrotron in Hamburg has explored energies up to a little more than 36 GeV without finding evidence of the new particle. A program that will raise the maximum energy of the machine in small steps is to begin this summer, and within a few years the energy is expected to reach 44 GeV.

Other investigators have been searching for free quarks rather than quarks bound into composite particles. A free quark should be easy to recognize because it is expected to have a fractional value of electric charge, equal to one-third or two-thirds the magnitude of the electron's charge. By the mid-1970's, however, a consensus was developing among theorists that quarks could never be isolated. Then in 1977 William M. Fairbank and his colleagues at Stanford University reported finding evidence of fractional electric charges on small superconducting spheres of the metal niobium. In a further report last April, Fairbank presented eight more observations of fractional charge, clustered closely about values of plus one-third and minus one-third. Nevertheless, the question of whether or not free quarks exist in nature is far from being settled. Fairbank's observations of fractional charge might have some other explanation; moreover, it is not obvious how to reconcile his findings with those of numerous other quark searches. For example, a series of experiments done by Giacomo Morpurgo of the University of Genoa has examined a somewhat greater quantity of material (iron rather than niobium) but finds no evidence of fractional charges.

New methods are now being applied to the search for free quarks. Both Fairbank and Morpurgo measure the charge of a metal sphere by levitating it in a magnetic field; after several years' work they have been able to examine only a few milligrams of metal. A technique devised by Ray Hagstrom of the Argonne National Laboratory and Gregory Hirsch of the Lawrence Berkeley Laboratory promises to bring a larger quantity and variety of materials under study. Hagstrom and Hirsch apply an electrostatic force to a stream of droplets; each droplet is deflected from its path by an amount proportional to its electric charge, and so fractionally charged droplets should be separated from all the others. Thousands of droplets per second can be processed.

The foregoing quark searches aim to detect fractionally charged particles that happen to be loose in ordinary matter; experiments of another kind endeavor to create new quarks or to break up composite particles into their constituent quarks. Three curious events that may ultimately turn out to signal the presence of free quarks have lately been observed in one of these experiments, done by Antonio Zichichi of the European Organization for Nuclear Research (CERN) in Geneva. Zichichi examines the debris formed when high-energy neutrinos strike a lead target. Charged particles in the debris are detected when they ionize the atoms of a low-density gas. The extent of ionization is proportional to the square of the charge, and so a particle with a charge one-third that of the electron would cause one-ninth as much ionization. The three candidate events are consistent with this value, but their identification is still tentative.

The neutrino itself is at the center of another controversy. There are three types of neutrino, designated the electron type, the muon type and the tau type. They have no electric charge, and for many years it was assumed that they also have no mass. Now theoretical developments make it seem plausible that at least some of the neutrino types do have a small mass. Furthermore, if the types differ slightly in mass, a neutrino might periodically change its identity from one type to another.

In 1980 Frederick Reines and his colleagues at the University of California at Irvine announced that they had found evidence of such neutrino oscillations. They had set up a detector 11 meters from the core of a nuclear reactor at Savannah River, S.C., where electron-type neutrinos are emitted copiously in the course of nuclear fission. The detector was designed to distinguish between two processes. At low neutrino energies one process could be induced only by an electron-type neutrino, whereas the other process could be induced by a neutrino of any type. Oscillations that convert electron-type neutrinos into other types would show up as an enhancement of the latter process over its expected rate.

Reines and his colleagues interpreted their results as showing a significant enhancement, which implied that the neutrino types differ in mass by roughly one electron volt. Other workers disputed the interpretation, however, primarily because it depends on a calculation of the energy spectrum of the emitted neutrinos. High-energy electron-type neutrinos can give rise to both processes in the detector, and so they can mimic the effect of oscillations.

Last September another measurement of the neutrino flux from a reactor was reported by Felix H. Boehm of the California Institute of Technology and his co-workers from Cal Tech, the Institute of Nuclear Sciences in Grenoble and the Technical University of Munich. With a detector about nine meters from the reactor core, they found that their data were consistent with no oscillations. By assuming the maximum degree of "mixing" among the neutrino types they set an upper limit of about .4 electron volt on the mass difference.

The most direct way to detect neutrino oscillations is to compare the proportions of the various neutrino types at several distances from a single source of neutrinos. The results are then independent of the neutrino energy spectrum, and fluctuations in the number of each type can be perceived directly. Experiments of this kind are now in progress. Boehm and his colleagues have moved their detector to a more powerful reactor at Gösgen in Switzerland, where they will measure the neutrino flux at distances of 38 and 65 meters from the core. Reines' group has constructed a new detector at Savannah River that can be moved continuously over a range of distances from 15 to 50 meters. Thus it seems possible the question will be settled in the coming year.

Perhaps the grandest of the present undertakings in experimental physics is the worldwide watch for the decay of the proton, a constituent of matter that until a few years ago was thought to be absolutely stable. The new unified theories predict that the proton does decay, but the average lifetime is at least 10^{31} years, or some 10^{21} times the age of the universe. The only way to detect the decay is to stand watch over at least 10^{31} protons, waiting for one to disintegrate. The mass of 10^{31} protons is about 18 tons, but practical detectors are much larger still. Moreover, the detecting instruments must be thoroughly shielded from cosmic rays, or the rare signal of a decaying proton will be obscured by extraneous events. To meet these conditions physicists have been going deep underground in mines and tunnels, where they have constructed great instrumented vats of water and stacks of iron plates.

Of the dozen or so groups that are now at work on proton decay only one group has reported even preliminary findings. The group is a consortium of investigators composed of Saburo Miyaki of the University of Tokyo and others from the Tata Institute of Fundamental Research in Bombay and the University of Osaka. The detector is a matrix of iron bars and gas-filled particle counters with a total mass of 140 tons; it is installed 2,300 meters underground in the Kolar goldfields near Bangalore in southern India. A year ago Miyaki and his co-workers described three candidate proton-decay events, in which the tracks of charged particles seemed to emanate spontaneously from a point within the volume of the detector. In all three cases, however, at least one of the tracks extended to the edge of the detector, which complicates the interpretation of the events; indeed, the possibility cannot be excluded that the tracks were made by a particle that entered from the outside and collided with an atomic nucleus in the detector.

Miyaki and his colleagues have since reexamined the three events and concluded that two of them should be disqualified from consideration as possible proton decays. One candidate remains, but little can be learned from a single event. If the one event is genuine, additional proton decays should be seen in the Kolar detector at a rate of roughly one per year; this rate implies a proton lifetime of about 3×10^{31} years.

Larger detectors are about to go into operation. Some 1,000 tons of water serves as the active medium in a device being tested in the Silver King Mine in Utah by investigators from Harvard University, Purdue University and the University of Wisconsin. In the Morton Salt Mine near Cleveland workers from Irvine, the University of Michigan and the Brookhaven National Laboratory are filling a cavity with 10,000 tons of water. (Both of these detectors might have been ready sooner had it not been for leaks in the plastic bladders that hold the water.) If the proton lifetime is 3×10^{31} years, the 10,000-ton detector should register about a dozen events per year. This rate might be enough to confirm the fact of proton decay, but more events will be needed to elucidate the mechanisms of decay and to discriminate among theoretical models.

If the lifetime of the proton is much longer than 10^{33} years, building larger detectors may be fruitless. Conceivably a mass large enough to generate a useful decay rate might be assembled and monitored. As the size of the detector increases, however, so does the irreducible background of extraneous events. The signal would be lost in noise. Hence, among the several questions now pending in experimental physics, the question of proton decay is one that may be answered soon or not at all.

Moving the Mud

Civil engineering in the Netherlands is famed for large-scale undertakings to expel the sea from the land. In 16th- and 17th-century Amsterdam, however, the Dutch engineers confronted the opposite problem. The expanding wharf area, sheltered by walls of pilings from the choppy waters beyond, was no longer adequately scoured by the tides and began to fill up with mud from the Zuiderzee. In response a major dredging project was begun. The first Amsterdam *moddermolen,* or mud mill, started work at the end of the 16th century. By the mid-19th century, when steam dredges took over the work, the number of harbor mud mills had grown to seven. They were tended by hundreds of *modderschouwen,* or mud barges, which carried away the dredged silt and dumped it at sea.

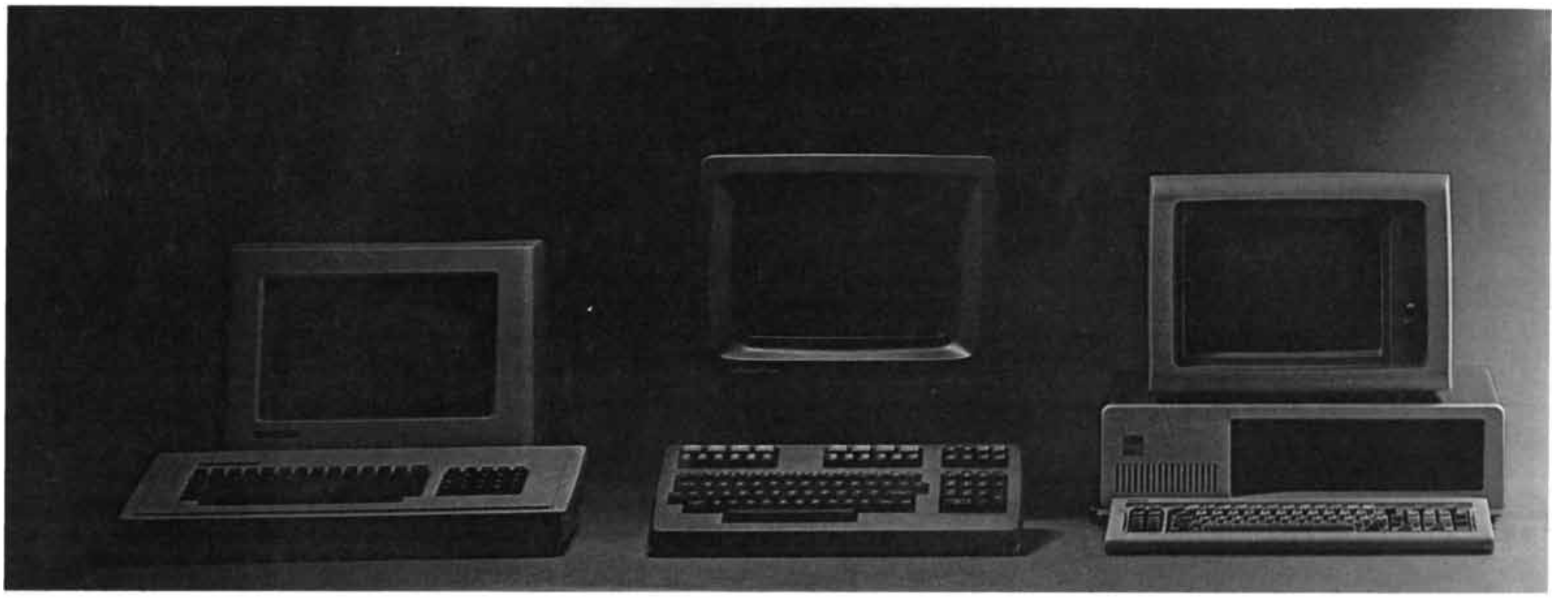
Inspired by the recent discovery and excavation of two 17th-century mud barges, the Dutch marine archaeologist Reinder Reinders searched the Amsterdam city archives for accounts of the harbor dredging. He discovered that by 1674 the two main wharf areas had become so heavily silted that the port authorities decided to take the dredging work out of private hands and do it themselves. As a result detailed records were kept from that year into the 1680's. Reinders has reported some of his findings in *Nautical Archaeology.*

To determine the annual amount of silting the port authorities had the harbor regularly sounded between 1678 and 1684. They found that the silt was accumulating at a rate of about 200,000 cubic meters per year. They concluded that under favorable winter weather conditions two large and two small mud mills could keep up with the accumulation. A large mud mill could dredge between 50,000 and 55,000 cubic meters in a good year and a small mud mill between 35,000 and 40,000 cubic meters. Critical to dredging performance was the availability of mud barges; the best results, they found, were achieved when 18 double barges attended each large mill and 20 single barges attended each small one. The authorities ordered the construction of 30 new mud barges.

The mud mills were worked by muscle power. Early in the 17th century human muscle supplied the power. Two men (each man on his own treadwheel) drove a mud-laden endless belt with a dredge head rigged at its lower end. The belt raised the mud out of the water and dropped it into a barge waiting at the high end of the belt. From the middle of the century on the large mud mills were powered by horses rather than men; two horses walked a circular path, turning a capstan, and a train of gears applied this motive force to the endless belt.

How did the bargemen dispose of the mud? Until the mid-19th century they ladled it over the side with special scoops. Then, at about the same time as steam power replaced muscle power, a new kind of barge was introduced that could be emptied by opening valves in the bottom. The mudmen must have cheered.

NOW THAT YOU FIRST GE TAKE A LOOK



Welcome to a tour of the biggest Apple.

The Apple® III Personal Computer—the most powerful machine in its class.

Because it's the only personal computer that lets you add up to 256K RAM, hang on a full complement of peripherals, and still have four expansion slots left for future growth. (Unlike some micros which become woefully "slot-bound" when upgraded to maximum memory.)

Because it's the only machine now using 64K RAM chips to keep 256K tidy on a single board.

And because it's the only machine that gives you the help of SOS.

THE MOST SOPHISTICATED OS.

SOS is the Apple III's Sophisticated Operating System, an elegant software interface that frees you from most system control tasks. It features a hierarchical file system, device- and user-level interrupt capabilities, a device-independent file system and memory management capability.

Since all Apple III languages use SOS, they share a common disk format. So Apple III programs can merge and communicate—a Pascal application program can directly access a BASIC text file, for example.

Xerox 820	Hewlett-Packard 125—Model 10	IBM Personal Computer	Apple III
Standard Memory			
64K	64K	64K	128K
*Maximum Memory when fully configured**			
64K	64K	192K	256K
Expandability			
No expansion slots	No expansion slots	No extra expansion slots in fully configured* 192K system	4 extra expansion slots in fully configured 256K system*
Diskette Storage (per drive)			
92K	256K	160K	140K
Mass Storage (per drive)			
—	1.16 megabyte Floppy Disk	—	5 megabyte Hard Disk
Display Graphics Capability			
High resolution B/W	High resolution B/W	High resolution B/W or 4-color (color requires additional card)	High resolution B/W or 16-color
Software Available			
Word Processing Super Calc® — Communications — CP/M® library	Word Processing VisiCalc® 125 Business Graphics Communications — CP/M® library	Word Processing VisiCalc® — Communications — CP/M® 86 programs	Word Processing VisiCalc® III Business Graphics Communications Apple II software library CP/M® library (available Spring, 1982)

*"Fully configured" means system includes, at minimum, monitor, printer, 2-disk drives and RS-232 communicator. NOTE: Chart based on manufacturer's information available as of December, 1981.

'VE SEEN THEIR NERATION AT OUR THIRD.

SOS allocates system resources to make the most of dynamic memory, simplifies programming with standard device and file interfaces for all languages, and speeds software development by reducing program size and complexity.

OUR NEW PROFILE.

ProFile™ is Apple's new personal mass storage system—a quick, quiet 5MB hard disk ideal for software development or any mass storage application. Shown above twixt monitor and console, it comes with everything you need to get up and running, including interface card and driver software.

The III's standard built-in drive is a 140-K floppy that can be daisy-chained with three additional drives through a back panel connector. Which leaves you plenty of expansion slots for things like our new Universal Parallel Interface Card or our OEM Prototyping Card.

CHANGE KEYS.

The 128-character ASCII-encoded keyboard happens to be fully-programmable. So you can (with SOS) do neat things like remap it into DVORK. Or create armies of special function keys. Or teach your Apple to write Chinese.

Its own languages already include Business BASIC, UCSD Pascal,™ Assembly and, soon, a powerful new COBOL—and, in emulation mode, most languages available for Apple II.

Look up from the keyboard and you'll see our standard Monitor III green phosphor display (80 char. x 24 lines, u/l). It can show you some of the highest resolution graphics available—560 x 192 in 16 gray scales. Or choose a color monitor, and you could be seeing 16-color graphics in high 280 x 192 resolution.

HIT LIST.

We have a new edition of that monster hit, VisiCalc® with more modeling space than any other version. Plus AppleWriter III, a powerful new word processing package. Plus a new Business Graphics package. Plus a new Mail List Manager.

Plus a Pascal Utility Library that lets you take full advantage of all UCSD Pascal features.

Plus Access III—sophisticated smart terminal software to access mainframes with asynchronous communications up to 9600 bps.

Plus Apple II emulation to access that vast software library.

Plus, soon, a CP/M® card to access that other vast library.

Even with all this, the Apple III's potential remains essentially untapped.

So we're offering improved documentation, new programmer's aids, expert hotline counseling and an open channel to the industry leader in software publishing. Us.

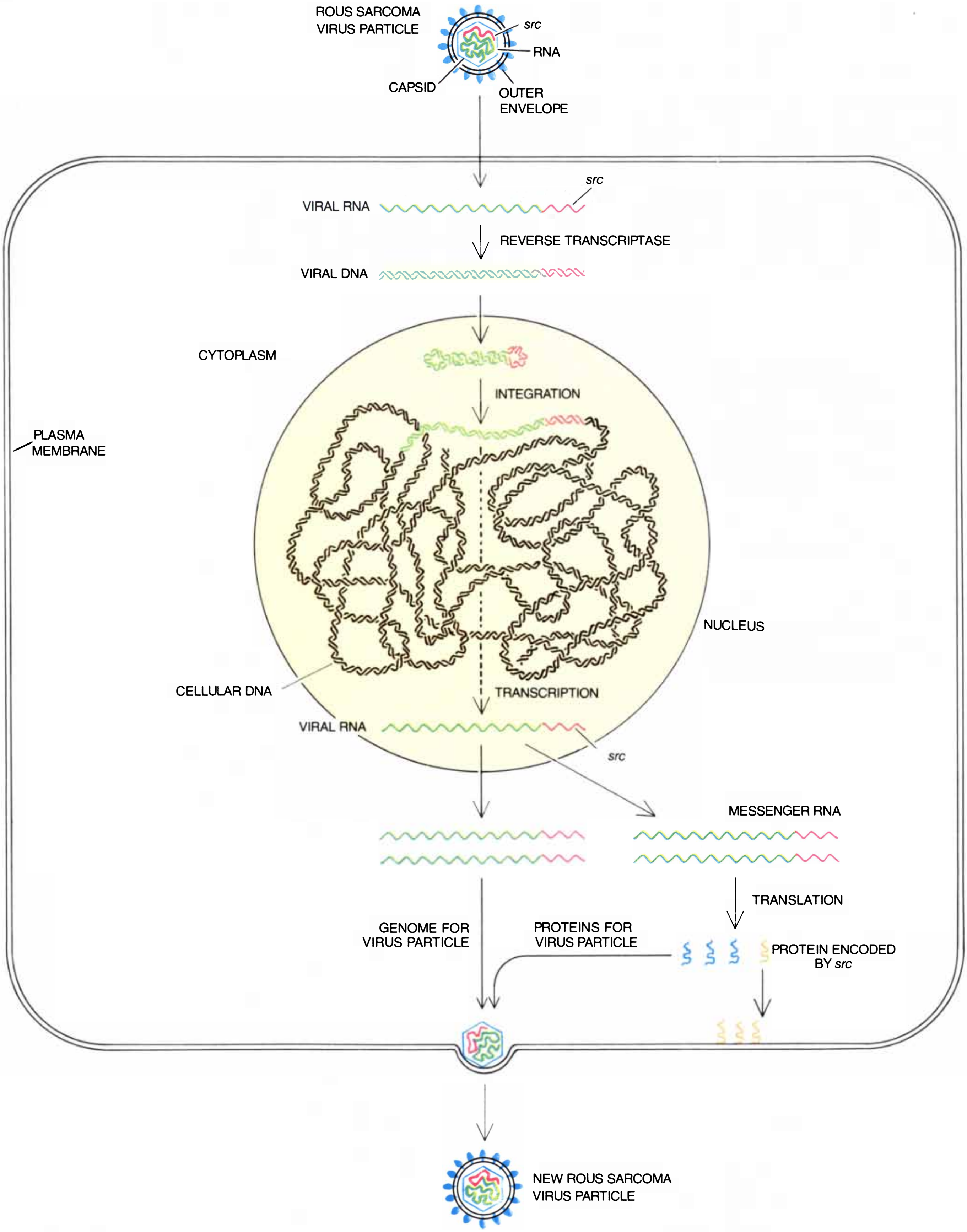
See your Apple dealer for more information and, perhaps, a little comparison shopping.

We're sure you'll find that between our third generation and their first, there's quite a gap.

The personal computer. apple®

For the authorized dealer nearest you, call (800) 538-9696. In California, call (800) 662-9238. Or write: Apple Computer Inc., 10260 Bandley Dr., Cupertino, CA 95014.

VisiCalc is a registered trademark of Personal Software, Inc. UCSD Pascal is a trademark of the Regents of the University of California. CP/M is a registered trademark of Digital Research, Inc. Apple is a registered trademark



ONCOGENE called *src,* which induces a cancer (a sarcoma) in chickens, is part of the small RNA genome of the Rous sarcoma virus, a "retrovirus." When a retrovirus infects a cell, the viral RNA is copied into double-strand DNA by the enzyme reverse transcriptase, which is supplied by the virus. The DNA forms a circle and then becomes integrated into the host cell's DNA. When the host DNA is transcribed into RNA by cellular enzymes (not shown), the viral DNA is also transcribed (*broken arrow*). Some of the viral RNA provides copies of the viral genome for inclusion in new virus particles and some of it is processed to make messenger RNA, which is translated by the cell's protein-synthesizing machinery into viral proteins. Several of these proteins are incorporated into new virus particles. The product of the *src* gene, however, is not a component of the virus particle. It is an enzyme called a protein kinase, which binds to the inner surface of the cell's plasma membrane and phosphorylates cellular proteins, thereby converting the cell to cancerous growth.

Oncogenes

They are genes that cause cancer. They were first found in viruses, but their evolutionary history implies that normal vertebrate cells have genes whose abnormal expression can lead to cancerous growth

by J. Michael Bishop

Can the cancer cell be understood? Since no one can yet explain how a normal cell controls its growth, it may seem foolhardy to think the abnormal rules governing the growth of a cancer cell can be deciphered. Yet the history of biology records many instances in which the study of abnormalities has illuminated normal life processes. Recent developments in cancer research have added a dramatic new example. For the first time investigators have perceived the dim outline of events that can induce cancerous growth. Enzymes that catalyze those events have been identified, and so have the genes specifying the structure of the enzymes.

These advances have come from the study of viruses that induce tumors. Recent years have seen an enthusiastic search for viruses that cause cancer in human beings. The search has been largely unsuccessful, leading many informed observers to doubt that viruses will ever prove to be a major cause of human cancer. Some viruses do induce tumors in other animals, however, and investigators have been studying these tumor viruses, attempting to define fundamental derangements of the cell that are responsible for cancerous growth. That quest has struck gold.

Although the genes implicated in the development of cancer were first observed in work with viruses, they are not native to the viruses. Indeed, it has turned out that the genes are not even peculiar to cancer cells. They are present and functioning in normal cells as well, and they may be as necessary for the life of the normal cell as they appear to be for the unrestrained growth of a cancer. A final common pathway by which all tumors arise may be part of the genetic dowry of every living cell.

Tumor Viruses

A virus is little more than a packet of genetic information encased in a protein coat. The information can be embodied in either DNA or RNA (whereas in the cells of higher organisms the genetic archive invariably consists of DNA). Both DNA and RNA are long strands of four of the chemical units called nucleotides. The sequence of the nucleotides constitutes a coded message, punctuated into the discrete units called genes. The instructions encoded in genes are carried out in various ways. Most commonly the sequence of nucleotides specifies the order in which amino acids are assembled to form a particular protein, typically an enzyme or a structural element. Viruses can have fewer than five genes and never have more than several hundred, whereas the cells of more complex organisms have a genome, or total genetic complement, of tens of thousands of genes. The reproduction of viruses mimics the processes by which cells grow and divide, but the simplicity of viruses makes them much easier than cells to study and understand.

In cells DNA is transcribed into a strand of messenger RNA and the RNA is translated into protein. An infecting virus insinuates its genetic information into the cellular machinery, so that the cell synthesizes viral proteins specified by viral genes. The proteins synthesize many copies of the viral genome, construct new virus particles and execute any other instructions of the viral genes. In some instances the instructions include a command that converts the host cell to a cancerous state.

The existence of tumor viruses was first suspected at the turn of the century. A critical discovery came in 1910, when Peyton Rous of the Rockefeller Institute for Medical Research showed that a cell-free filtrate from chicken tumors called sarcomas could induce new sarcomas in chickens. His reports were not well received; eventually Rous abandoned his work on tumor viruses because of his peers' disapproval. Decades later the reality of the virus first identified by Rous, and of other tumor viruses as well, was established beyond doubt by purification with physical techniques and visualization with the electron microscope. Tumor viruses became workaday agents in cancer research. In 1966, at the age of 85, Rous was awarded a Nobel prize.

Some tumor viruses are oncogenic (that is, they induce tumors) only in animals that are not their host in nature, whereas other tumor viruses are oncogenic in their natural host. Such differences are understood only in part, but for the investigator they are of no great concern. The ability to induce tumors at will with a rather simple and well-defined agent has been a great boon to cancer research, even if it is sometimes necessary to resort to an unnatural combination of virus and host.

Transformation

Many tumor viruses have a particularly valuable property: they elicit cancerous changes in cells in an artificial culture medium. This "transformation" of cultured cells makes it possible to examine the interaction of a tumor virus with a host cell under controlled conditions and to avoid the difficulties associated with experiments in animals. It is important to remember, however, that some tumor viruses do not transform cells in culture and yet are powerful oncogenic agents in animals.

The ability or inability of a virus to transform cultured cells is connected with its mechanism of oncogenesis. Two patterns have been recognized. Some viruses have a single gene that is solely responsible for their capacity to induce tumors, or in some cases a few such "oncogenes." The action of viral oncogenes is rapid, and it predominates over the activity of all other genes in the cell. Most viruses with oncogenes (and perhaps all of them) can transform cells in culture; the capacity for transformation is provisional evidence that a virus has an oncogene. Other viruses lack oncogenes and induce tumors by more subtle means. Such tumor viruses act slowly in animals, in many cases taking from six to 12 months to give rise to a tumor, in contrast to the few days or weeks required by an oncogene virus. And they do not transform cells in culture.

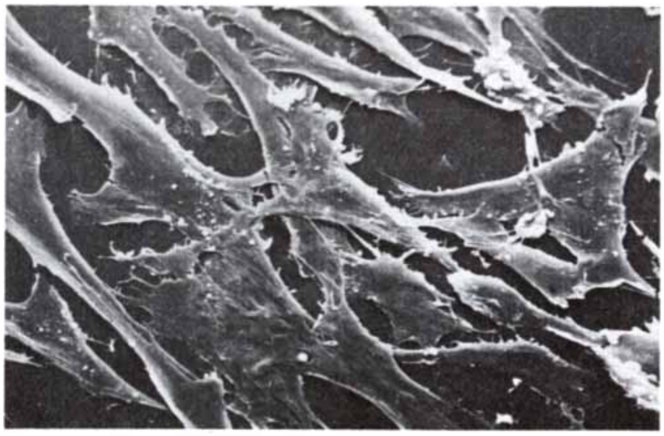
Both forms of oncogenesis are characterized by the persistence of the viral genome in the host cell for as long as the cell survives. In most instances viral DNA has been integrated, or chemically stitched, into the DNA of the host cell, but the genome of some tumor viruses appears to survive within the cell as a separate unit and to reproduce independently. At the moment it seems that the persistence of the viral genome is necessary for viral oncogenesis, either to maintain the influence of an oncogene over the cell or to sustain the less direct effects of viruses that induce tumors but do not carry an oncogene. The mysteries of viral oncogenesis have occasionally prompted the hypothesis of a "hit and run" mechanism in which a transient virus infection triggers a sequence of events eventuating in a tumor, with no trace of the virus necessarily persisting in the tumor cells. There is now very little evidence to support such models.

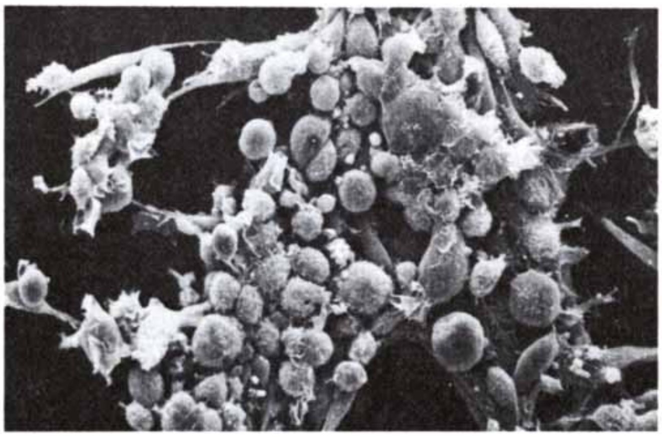
Retroviruses

The sarcoma virus discovered by Rous belongs to a family known as the retroviruses, which are the only tumor viruses with an RNA genome. Retroviruses have provided the most coherent view of oncogenesis now available. Three features of retroviruses account for their utility in the analysis of tumor development. First, they have been found in a large number of vertebrate species and they induce many types of tumors: experimental models for most major forms of human cancer. Second, it is relatively easy to identify and isolate retrovirus oncogenes and to find their products, and so they have provided the first glimpse of the chemical processes responsible for cancerous growth. Third, retrovirus oncogenes do not appear to be indigenous components of the viral genome; instead they seem to have been copied from genes of the vertebrate host in which the virus replicates. There is reason to suspect that the cellular genes from which the retrovirus oncogenes apparently arose are themselves involved in the production of tumors induced by agents other than viruses. Thus tumor virologists engaged in the arcane endeavor of tracking the evolutionary origin of oncogenes have been led to genetic mechanisms that may underlie many forms of cancer.

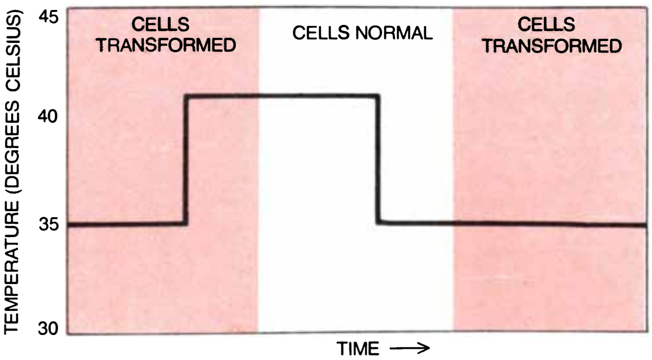
Retroviruses derive their name from a feature of their life cycle that makes them unique in biology: their RNA must be transcribed "backward" into DNA for them to propagate. This unusual process is accomplished by an enzyme called reverse transcriptase. The enzyme was discovered in the particles of viruses such as the Rous sarcoma virus in 1970 by David Baltimore of the Massachusetts Institute of Technology and by Satoshi Mizutani and Howard M. Temin of the University of Wisconsin. The discovery was important on several counts. It scuttled the widely held misconception that genetic information could flow only from DNA to RNA. It triggered a surge of research on retroviruses by clarifying the previously obscure mechanism of their replication. And it provided an essential reagent for the developing technology of genetic engineering with recombinant DNA.

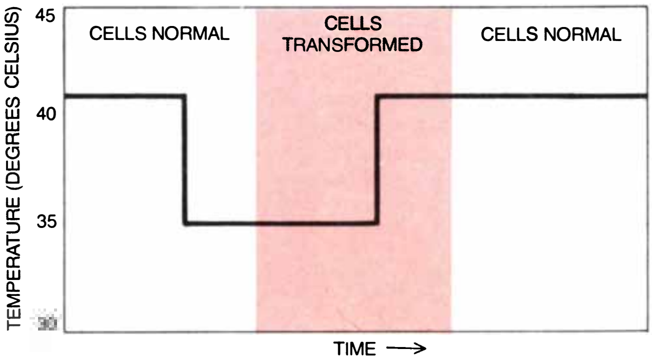
The life cycle of a retrovirus is a marvel of cooperation between parasite and host. The success of virus infection depends on the lavish hospitality offered by the cell, and yet the virus retains much authority to control events. During the early hours of infection the viral RNA genome is transcribed into DNA by reverse transcriptase. The viral DNA is then integrated into the cell's genome, with the result that viral genes are replicated along with cellular genes and are expressed by the machinery of the cell.





TRANSFORMATION of cultured cells by the Rous sarcoma virus is apparent in scanning electron micrographs made by G. Steven Martin of the University of California at Berkeley. Normal fibroblasts, or connective-tissue cells, adhere to the surface of a laboratory dish in which they are cultured and have a flat, extended configuration (*left*). On infection by the Rous sarcoma virus the cells become round instead of flat and cluster together in piles (*right*), presumably because cell proteins are phosphorylated by the enzyme encoded by *src*.





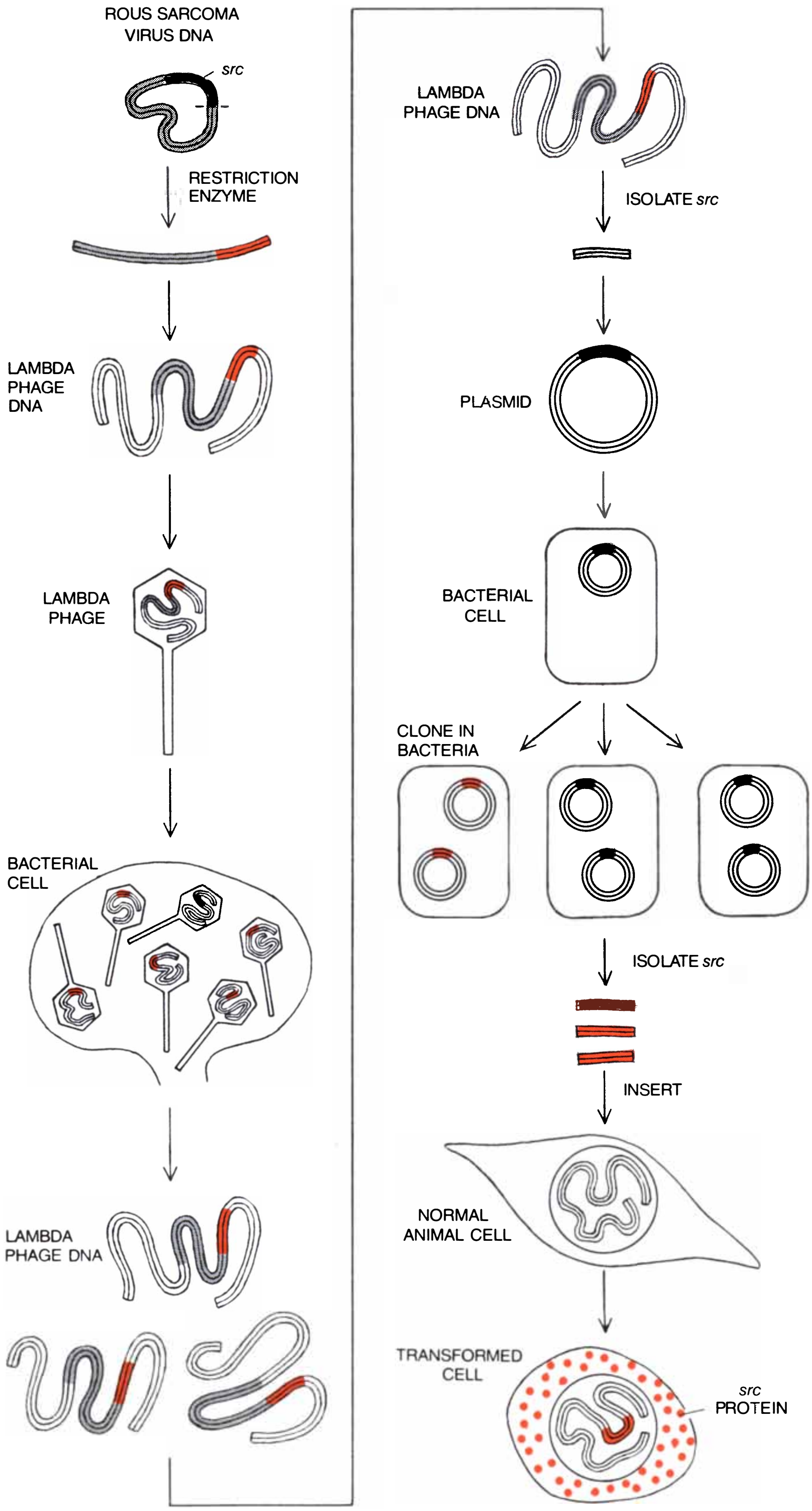
CONDITIONAL MUTATION of a gene results in the reversible inactivation of the gene or the enzyme it encodes. In the case of a temperature-sensitive mutant of *src*, infected cells are transformed when they are cultured at 35 degrees Celsius but revert to (or remain in) the normal state at 41 degrees. The discovery of such mutants implies that a viral gene must be responsible for transformation, that the action of the gene is probably mediated by a protein product and that the sustained action of the gene is necessary for transformation.

In many cases a retrovirus infection is innocuous to the cell. The virus acquires a new and potentially enduring home; new virus particles are manufactured and leave the cell, and yet the cell suffers no damage. The partnership can go awry, however, as a result of either of the two kinds of viral oncogenesis mentioned above. If the virus carries an oncogene, the activity of the gene can convert the cell to cancerous growth. If the virus lacks an oncogene, the integration of the viral DNA can interfere with a cellular gene at or near the point of insertion; in other words, the insertion can cause a mutation in the host cell's genome. Mutations at certain sites may engender cancerous growth. The induction of tumors by oncogenes and induction by the consequences of integration appear at first to be quite dissimilar events, but I shall show below that they are intimately related.

The *src* Gene

The oncogene of the Rous sarcoma virus was the first to yield to experimental analysis. An important early step was taken in 1970, when G. Steven Martin of the University of California at Berkeley identified temperature-sensitive "conditional" mutations that affect the ability of the virus to transform cells in culture. A conditional mutation is a powerful tool because it makes possible the reversible inactivation of a gene. When cultured cells infected with temperature-sensitive Rous sarcoma viruses are maintained at a "permissive" temperature, they are transformed. When the temperature is shifted to a higher, "restrictive" one, within hours the cells regain a normal appearance, only to be transformed once more when the temperature is again lowered. The interpretation is that at a restrictive temperature a mutated gene is inactivated. Transformation, then, is due to the action of a gene, which must be expressed continuously to maintain the cancerous state. (In most cases the elevated temperature probably does not act directly on the gene itself. Instead the mutation alters the structure of the protein product of the gene, with the result that the activity of the protein is impaired by the restrictive temperature.)

The gene first glimpsed by Martin is now called *src* (for sarcoma, the tumor it induces); it is the oncogene of the Rous sarcoma virus. The *src* gene was soon made more tangible by Peter H. Duesberg of Berkeley and by Charles Weissmann, Martin Billeter and John M. Coffin of the University of Zurich. They worked with strains of the Rous sarcoma virus that had been isolated by Peter K. Vogt of the University of Southern California. The strains are "deletion" mutants that have lost the oncogene and are therefore incapable of inducing tumors or transforming cells in culture.

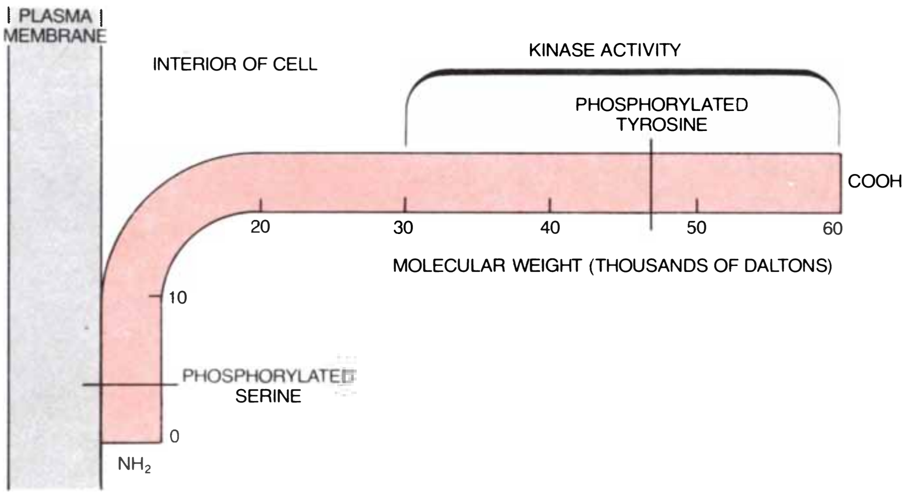


VIRAL ONCOGENE is purified, and its capacity to transform cells is tested, by methods of genetic engineering. The circular DNA of the Rous sarcoma virus is isolated from newly infected cells, cleaved with a restriction enzyme and inserted into the DNA of phage lambda, a bacterial virus. The growth of the phage in bacteria makes large quantities of the viral DNA, which is cleaved with a restriction enzyme to yield a fragment bearing only *src* and a bit of flanking DNA. The fragment is inserted into a plasmid (a small circle of bacterial DNA), which is introduced into bacteria for further amplification. Now many copies of the *src* fragment can be excised from the plasmids, purified and introduced into animal cells in culture. Fragment directs synthesis of a viral protein that induces transformation. Experiment was done by William J. DeLorbe and Paul A. Luciw to show that *src*, acting alone, gives rise to cancerous growth.

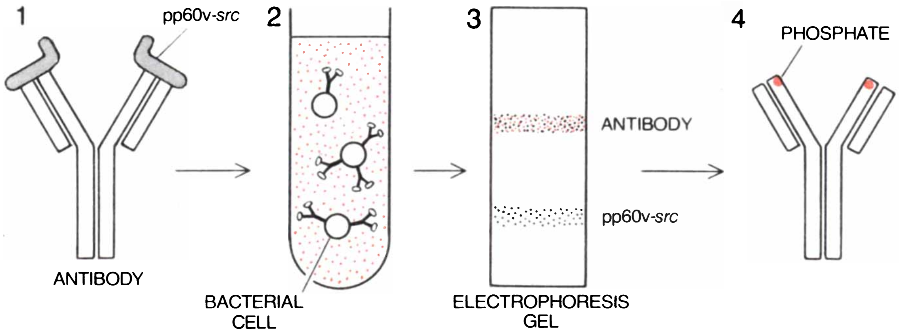
Duesberg and Weissmann and his colleagues fragmented the genomes of deletion mutants and of wild-type (oncogenic) viruses with the enzyme ribonuclease. By determining which fragment was missing in the mutants they were able to identify the oncogene as a segment of RNA near one end of the Rous sarcoma virus genome.

In the past few years the powerful new techniques of genetic engineering have been exploited to define oncogenes more precisely and to test their cancerous potential. DNA can now be cut into fragments at specific sites with the aid of a battery of enzymes called restriction endonucleases. Particular fragments can be grown in quantity in bacteria, then reisolated and inserted into cultured cells, where the genes carried by the DNA are expressed. In this way one can cut viral DNA into pieces that each carry a single gene and learn which of the pieces cause transformation. Analysis of the DNA of the Rous sarcoma virus has revealed a single gene capable of transforming cells; the gene encodes a single protein product. The implication is that one gene, by directing the synthesis of one protein, can bring about the changes characterizing a cancer cell. To know that protein and how it acts is to have in view the events that can generate a malignant tumor.

The protein encoded by *src* is known, owing largely to the work of Raymond L. Erikson of the University of Colorado School of Medicine and his colleagues. They began by identifying a protein that is synthesized in the test tube under the instructions of the wild-type Rous sarcoma virus genome but not under the instructions of the genome of a deletion mutant lacking *src*. Then they raised rabbit antibodies to a putative *src* protein by inducing tumors in rabbits with the Rous sarcoma virus. The antibodies combined specifically with the protein synthesized in the test tube and also with an identical protein in cells transformed by *src*. These findings persuasively identified a protein that is encoded by *src* and is responsible for the effects of the gene. The protein was designated pp60v-*src*; the "pp" signifies that it is a phosphoprotein (a protein to which phosphate groups are attached), the "60" refers to its molecular weight of 60,000 daltons and "v-*src*" indicates its genetic origin is the viral gene *src*.



ONCOGENE PRODUCT, a protein designated pp60v-*src*, is a chain of about 520 amino acid subunits. It is a protein kinase: an enzyme that adds phosphate groups to proteins. The protein is bound to the plasma membrane of cells by a domain near the amino (NH_2) end; the phosphorylating site is in the other half of molecule. Enzyme itself is phosphorylated at two sites.



PROTEIN-KINASE ACTIVITY of pp60v-*src* was discovered fortuitously. Rabbits in which tumors have been induced by the Rous sarcoma virus make antibodies to pp60v-*src*, and the antibodies form complexes with the protein (*1*). The complexes were immobilized on the surface of bacteria and a source of radioactively labeled phosphate (*colored dots*) was added (*2*). When pp60v-*src* and the antibody were separated by gel electrophoresis, the antibody band on the gel showed radioactivity (*3*): the immobilized pp60v-*src* had catalyzed the transfer of the phosphate to the amino acid tyrosine at a single site in each of the antibody's two heavy chains (*4*).

A Cancer Enzyme

How does the protein product of the *src* gene convert a cell to cancerous growth? That seemed to be a daunting question when the protein was isolated. Yet a first answer came quickly when it was discovered that pp60v-*src* is a protein kinase: an enzyme that attaches phosphate ions to the amino acid components of proteins in the reaction known as phosphorylation. The discovery was made by Erikson and his colleague Mark S. Collett and, independently, by Arthur Levinson, working with Harold E. Varmus and me in our laboratory at the University of California School of Medicine in San Francisco.

Soon thereafter Tony Hunter and Bartholomew M. Sefton of the Salk Institute for Biological Studies reported that pp60v-*src* attaches phosphate ions specifically to the amino acid tyrosine. That put pp60v-*src* outside the known classes of protein kinases, which phosphorylate the amino acids serine and threonine. The phosphorylation of tyrosine has turned out to be a common characteristic of oncogene-encoded enzymes; surprisingly, it also has a role in regulating the growth of normal cells.

Not many years ago phosphate appeared to most biologists to be a mundane material and its transfer to proteins a humble event. Now it is clear that the phosphorylation of proteins is one of the central means by which the activities of growing cells are governed. One enzyme, by phosphorylating a number of proteins, can vastly alter the functioning of a cell. In the case of pp60v-*src* two modes of action have been proposed. The enzyme could phosphorylate a single protein, precipitating a cascade of events that together give rise to the properties of a cancer cell; alternatively, the enzyme could phosphorylate numerous proteins, directly affecting the functions of each of them and perhaps precipitating secondary events or even cascades in turn. What little is known at the moment makes it seem likely that the second alternative correctly describes the action of pp60v-*src*.

Can the phosphorylation of tyrosine subunits in cellular proteins account for the ability of *src* to induce tumors? Hunter and his colleagues have shown that the amount of phosphorylated tyrosine in a cell increases approximately tenfold as a result of transformation by *src*. The increase is regarded as a manifestation of the activity of pp60v-*src*. The critical questions now are: What cellular proteins are phosphorylated by the enzyme and what are their func-

The networks use Scotch® Video Tape for the same reasons you should: true color and clarity.

All three major networks use Scotch® Video Tape. Because they *have* to get the best picture possible.

You can get true color and clarity at home, too. With Scotch Beta or VHS Videocassettes. Here's why:

We make video tape for home use just as carefully as the tape we make for the networks. It's not exactly the same. For one thing, network tape is wider. But we start with the same basic ferric oxide particles. It's made by the same processes, on the same machinery and to the same exacting quality control standards.

And because your home isn't a professional studio, our exclusive micro-fine Anachron™ particles are cobalt-encapsulated, giving you superb picture quality. And they are protected by a unique Static Barrier.™ This barrier virtually eliminates dust, a major cause of video dropouts—those white flashes in the picture.

We apply this kind of state-of-the-art technology to every video tape we make. That means you can count on true color and clarity in the extended play modes as well as at the fastest speed.

So take a tip from the experts. And start enjoying true color and clarity like the networks get in your own home.

Scotch® Videocassettes. The truth comes out.

S C O T C H® • V I D E O C A S S E T T E S

3M

We still don't think it's a sports car.

But if you insist.

No matter what the car experts say, the Honda Civic GL simply wasn't designed to be a sports car.

However, it does have the basic performance features of one. Such as 4-wheel independent suspension, rack and pinion steering and steel-belted radial tires.

The Civic GL is certainly fun to drive. With a peppy 1488cc engine to keep things lively. Front disc and rear drum brakes bring everything to a stop. They're power-assisted and self-adjusting too.

There's front-wheel drive for better traction and a smooth 5-speed transmission. There are also aerodynamic improvements. Which helped the Civic GL pass its mileage tests with flying colors.

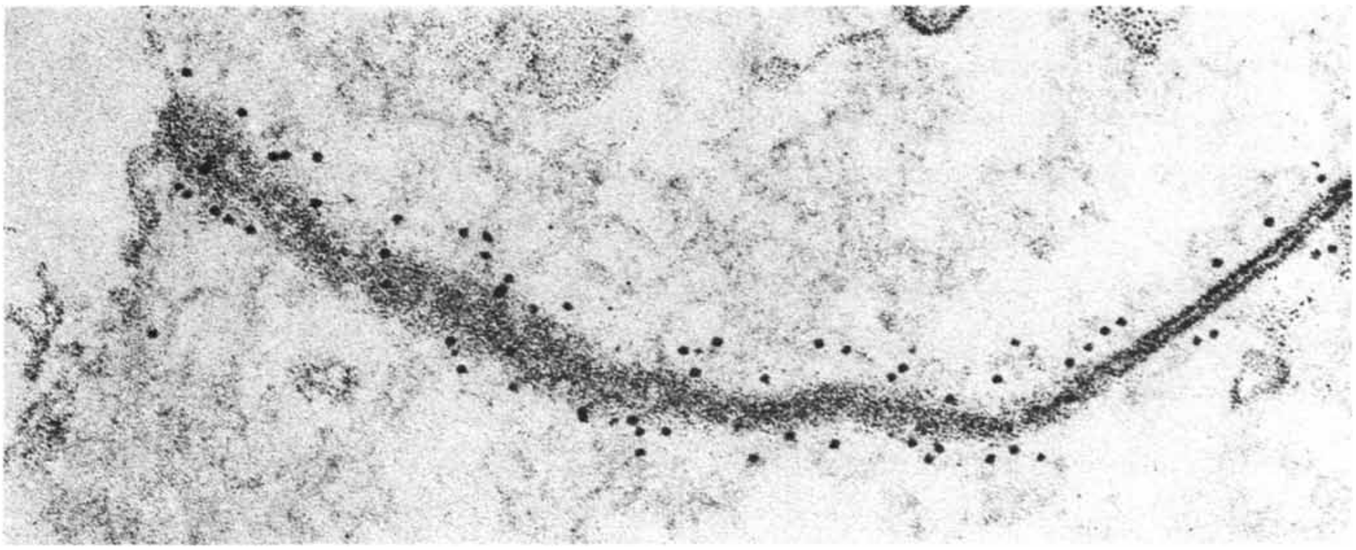
EPA estimated 35 mpg and 46 highway. Use 35 mpg for comparison. Actual highway mileage will probably be less. Your mileage may vary because of weather, speed or trip length. California figures will be only one mile per gallon lower.

The interior does nothing to deny a sporty image, either. Instruments include a tachometer and quartz digital clock. There are reclining front bucket seats and a 4-spoke sport steering wheel. And a remote control outside rearview mirror.

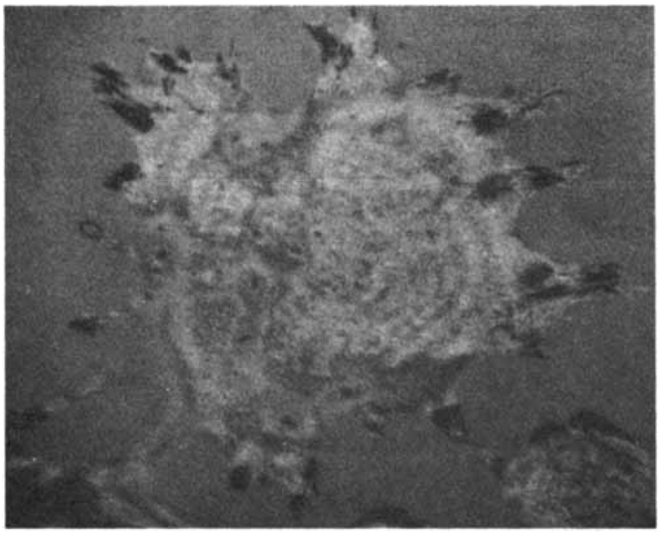
Is the Civic GL a sports car? There's only one way to settle this. And it isn't here.

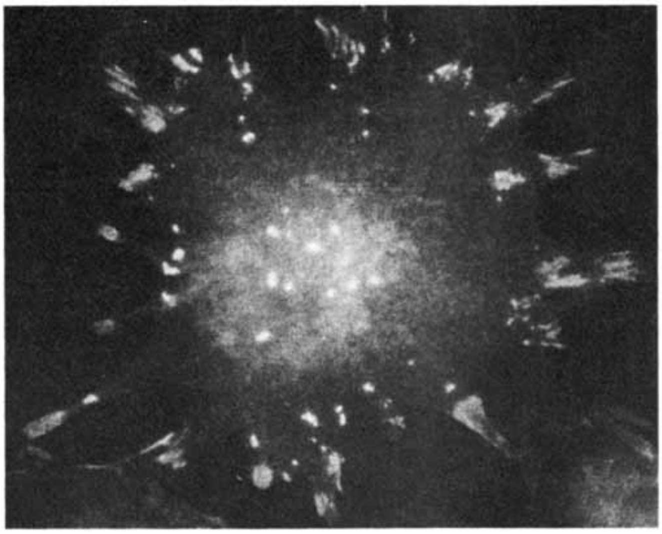
HONDA

We make it simple.



LOCATION of most of the *src* protein is demonstrated by an electron micrograph made by Mark C. Willingham and Ira H. Pastan of the National Cancer Institute. Cells infected with the Rous sarcoma virus were thinly sliced and treated with a rabbit antibody to pp60v-*src;* ferritin, a substance opaque to the electron beam, had been linked to the antibody indirectly (by way of a second antibody). The rabbit antibody became bound to pp60v-*src,* whose location along plasma membranes of two adjoining cells is revealed by dark ferritin dots.





ADHESION PLAQUES, which anchor cells to surfaces, are directly affected by pp60v-*src,* as is demonstrated by photomicrographs made by Larry R. Rohrschneider of the Fred Hutchinson Cancer Research Center in Seattle. In an interference-reflection micrograph (*left*) the adhesion plaques, which are in contact with the surface, are dark. When the same cell is treated with an antibody to pp60v-*src* that is indirectly labeled with a fluorescent dye, an ultraviolet micrograph (*right*) shows that most of the protein is in adhesion plaques.

tions? There are only a few clues, none of which can yet account for the unrestrained growth of the tumors induced by *src.* The pursuit of targets for pp60v-*src* is under way in many laboratories.

Site of Action

One approach is to find out where in the cell pp60v-*src* acts, in the hope of learning what proteins it affects and what those proteins do. Early studies indicated that the products of viral oncogenes might take up residence in the nucleus of the cell, where they could meddle directly with the apparatus responsible for replicating the cellular DNA and so drive the cell to unrestrained growth. Experiments by Hartmut Beug and Thomas Graf of the Max Planck Institute for Virus Research in Tübingen showed, however, that the effects of the *src* protein can be detected even in cells from which the nucleus has been removed. It came as no surprise, then, when several workers found that little if any of the pp60v-*src* in transformed cells is in the nucleus. Most of the protein is at the other extreme of the cell: it is bound to the plasma membrane, the thin film that encloses the cell and mediates its interactions with the outside world. Many cell biologists have argued that the control of cell growth may originate at the plasma membrane and its associated structures.

Inspection of the plasma membrane of cells transformed by *src* has provided the first correlation between the action of pp60v-*src* on a specific cellular protein and one of the typical changes in structure and function seen in cancer cells. By means of specialized techniques of photomicroscopy Larry R. Rohrschneider of the Fred Hutchinson Cancer Research Center in Seattle was able to demonstrate that pp60v-*src* is concentrated in adhesion plaques: regions of the membrane that adhere to solid surfaces. In cancer cells the adhesion plaques are dismantled; the resulting decrease in cell adhesion may contribute to the ease with which most cancer cells break away from their tissue of origin and metastasize to other sites.

Rohrschneider's findings suggested that pp60v-*src* might dismantle adhesion plaques by phosphorylating one of their component proteins, or perhaps several of those proteins. Pursuing that suggestion, Sefton and S. J. Singer of the University of California at San Diego showed that pp60v-*src* phosphorylates a tyrosine unit in vinculin, a protein that is a constituent of normal adhesion plaques and becomes dispersed throughout the cell following transformation by *src*. It seems reasonable to suggest that the phosphorylation of vinculin precipitates the dismantling of adhesion plaques, but the importance of such events in the unruly behavior of cancer cells has yet to be established.

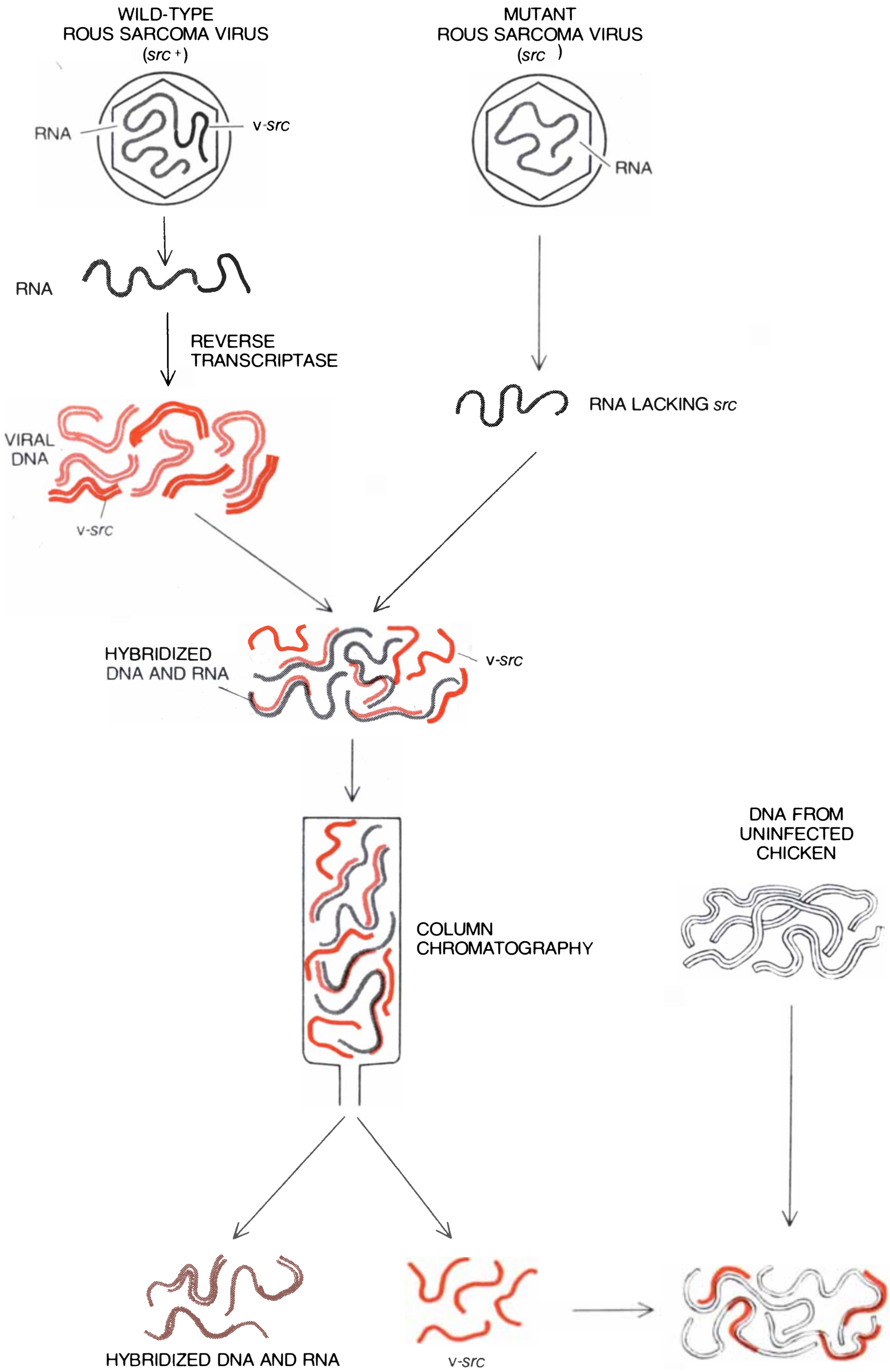
Once it was thought that the oncogenic effects of viruses might be ancillary manifestations of viral genes whose main function is to assist in the production of new virus particles. Now it is clear that the replication of retroviruses proceeds normally in the absence of oncogenes. How then can one explain the wide occurrence of oncogenes in retroviruses and their apparent conservation in the course of evolution? A decade of investigation has furnished a surprising answer. Retrovirus oncogenes are merely cellular genes in another guise, passengers acquired from the animals in which the viruses replicate. The discovery that cells too have oncogenes has implications extending far beyond tumor virology.

The Origin of Oncogenes

In 1972 Dominique Stehelin, Varmus and I set out to explore the "oncogene hypothesis" proposed by Robert J. Huebner and George J. Todaro of the National Cancer Institute. Seeking one mechanism to explain the induction of cancer by many different agents, Huebner and Todaro had suggested that retrovirus oncogenes are a part of the genetic baggage of all cells, perhaps acquired through viral infection early in evolution. The oncogenes would be innocuous as long as they remained quiescent. When stimulated into activity by a carcinogenic agent, however, they could convert cells to cancerous growth. We reasoned that if the hypothesis was correct, we might be able to find the *src* gene in the DNA of normal cells.

The DNA of vertebrates includes tens of thousands of genes. To search for *src* amid this vast array Stehelin fashioned a powerful tool: radioactive DNA copied solely from *src* by reverse transcriptase [*see illustration at left*]. The copied DNA served as a probe with which to search for cellular DNA with a nucleotide sequence similar to that of *src*. The search was carried out by molecular hybridization, in which chains of a nucleic acid (either DNA or RNA) hybridize, or form complexes, with nucleic acids to which they are related. We were exhilarated (and more than a little surprised) to learn that Stehelin's copy of *src* could hybridize with DNA from uninfected chickens and other birds. Deborah H. Spector joined us and went on to find DNA related to *src* in mammals, including human beings, and in fishes. We concluded that all vertebrates probably possess a gene related to *src*, and it therefore seemed the Huebner-Todaro oncogene hypothesis might be correct.

On closer inspection, however, the gene we had discovered in vertebrates proved not to be a retrovirus gene at all. It is a cellular gene, which is now called c-*src*. The most compelling evidence for this conclusion came from the finding that the protein-encoding information of c-*src* is divided into several separate domains, called exons, by intervening regions known as introns. A split config-



RADIOACTIVE PROBE with which to search for *src* in normal cells was prepared by Dominique Stehelin. RNA carrying the gene v-*src* was isolated from the Rous sarcoma virus (*top left*) and was copied into DNA by reverse transcriptase in the presence of radioactively labeled DNA precursors. The resulting radioactive DNA fragments (*color*) were denatured (complementary strands were separated) and mixed with RNA extracted (*top right*) from a mutant virus that lacks v-*src*. Single strands of DNA will "hybridize" with closely related strands of DNA or RNA. Most of the radioactive DNA hybridized with the deletion-mutant RNA, but the *src* fragments, finding no complementary RNA, could not. The DNA-RNA hybrids were then separated by column chromatography from the unhybridized *src* DNA to yield a radioactive *src* probe. When the probe was mixed (*bottom right*) with normal chicken DNA, it hybridized, revealing the presence of a cellular proto-oncogene that was designated c-*src*.

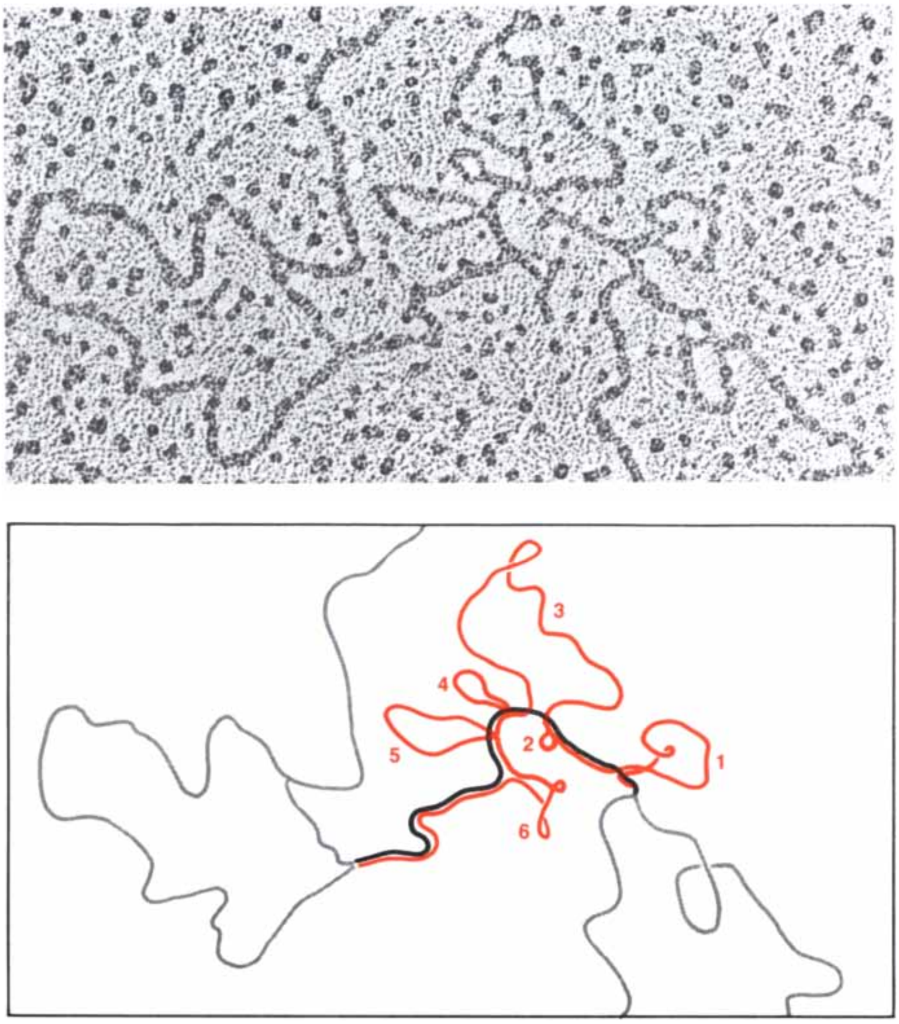
uration of this kind is typical of animal-cell genes but not of the genes of retroviruses. Apart from their introns, the versions of c-*src* found in fishes, birds and mammals are all closely related to the viral gene v-*src* and to one another. It appears the vertebrate *src* gene has survived long periods of evolution without major change, implying that it is important to the well-being of the species in which it persists.

The mystery presented by c-*src* deepened with the discovery that the gene not only is present in normal cells but also is active in them, that is, the gene is transcribed into messenger RNA and the RNA is translated into protein. Molecular hybridization with Stehelin's radioactive copy of v-*src* brought the RNA to view first, in both avian and mammalian cells. The protein was more elusive, mainly because it is synthesized in very small quantities in most cells. Success came when we and others probed for the cellular protein with antibodies prepared originally for the pursuit of the viral transforming protein, pp60v-*src*. The cellular protein isolated with the aid of these antibodies proved to be virtually indistinguishable from the viral protein, and it was therefore named pp60c-*src*. The two proteins are similar in size and chemical structure; both catalyze the phosphorylation of tyrosine and both are tightly bound to the plasma membrane of cells (transformed cells in the case of pp60v-*src*, normal cells in the case of pp60c-*src*). It is as if the two proteins were designed for the same purpose, even though one is a viral protein that causes cancer and the other is a protein of normal cells.

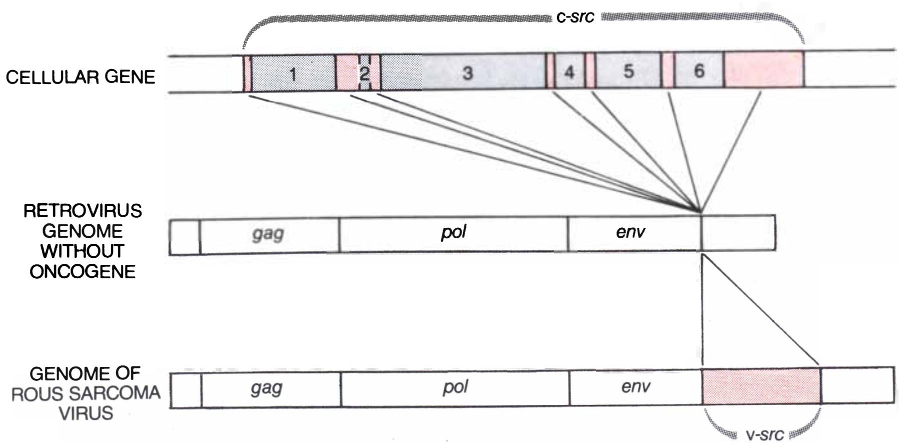
Cellular Oncogenes

The findings with respect to *src* were the first hint of a generalization whose extent and significance have yet to be established. Of 17 retrovirus oncogenes identified to date, 16 are known to have close relatives in the normal genomes of vertebrate species. Most of these cellular relatives of viral oncogenes obey the principles first deduced for c-*src*. They have the structural organization of cellular genes rather than of viral genes; they seem to have survived long periods of evolution, and they are active in normal cells. To account for these facts and for the remarkable similarity between retrovirus oncogenes and their normal cellular kin most virologists have settled on the idea that retrovirus oncogenes are copies of cellular genes. It appears the oncogenes were added to preexistent retrovirus genomes at some time in the not too distant past. How and why retroviruses have copied cellular genes is not known, but there is reason to think the copying continues today, and it may even be possible to recapitulate the process in the laboratory.

The vertebrate genes from which



CELLULAR AND VIRAL ONCOGENES are visualized in an electron micrograph made by Richard C. Parker. Viral DNA carrying the *src* gene and chicken DNA carrying the cellular version of the gene were isolated and the double-strand DNA's were denatured. Then the single strands were allowed to hybridize with any closely related strands. In this case, as is shown in the drawing, the viral gene (*black*) and the cellular one (*color*) have formed a heteroduplex. (Extraneous DNA required for cloning the genes is shown in gray.) The loops in the cellular strand are six introns: intervening sequences that interrupt the protein-encoding sequences (exons) in many cellular genes but not in retrovirus genes. Such electron micrographs helped to establish that the oncogenes in cells are native to the cell and were not introduced by viruses.



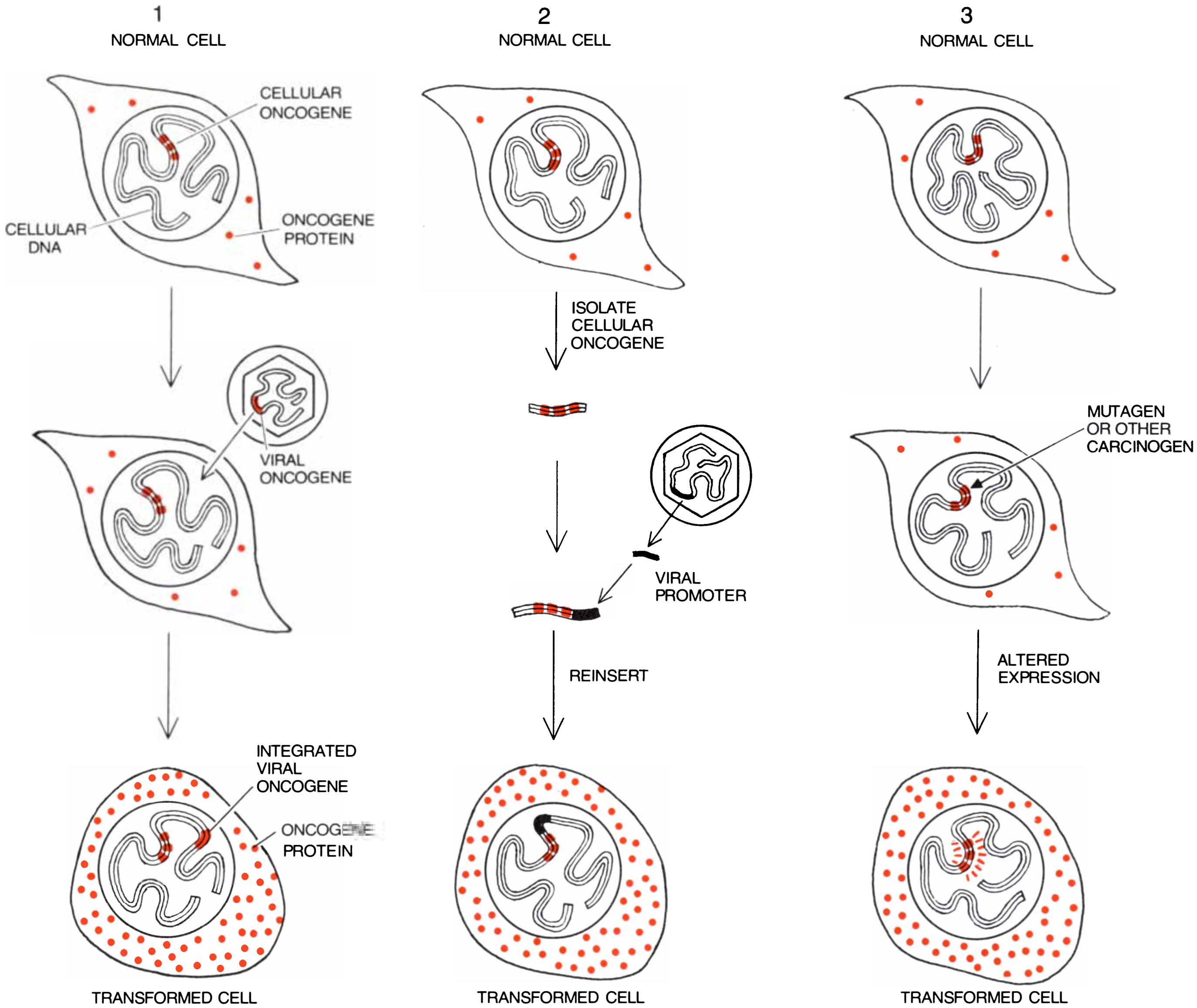
SPLIT CELLULAR GENE c-*src* (*top*) consists of exons (*color*) and introns (*gray*). The cellular gene was somehow picked up by a preexisting retrovirus; the introns were eliminated and the exons, spliced together, were inserted into the viral genome (*middle*) to complete the Rous sarcoma virus genome (*bottom*). In addition to *src* the genes are *gag*, which encodes the protein of the viral capsid, *pol* (encoding the enzyme reverse transcriptase) and *env* (encoding glycoprotein spikes of viral envelope). Other retrovirus oncogenes are thought to have similar origin.

retrovirus oncogenes apparently arose were at first called proto-oncogenes, to emphasize their evolutionary significance and to avoid implying that the cellular genes themselves have oncogenic potential. Now it is clear that they do have such potential. They are cellular oncogenes. The investigations that justify this designation began with this question: If retrovirus oncogenes are merely copies of genes found in normal cells, how can one account for the devastating effects of the viral genes on infected cells? Two explanations have been offered. The mutational hypothesis proposes that viral oncogenes differ from their cellular progenitors in subtle but important ways as a result of mutations introduced when the cellular genes were copied into the retrovirus genome. For example, the apparently similar enzymatic activities of pp60v-*src* and pp60c-*src* might actually have different targets in the cell and might therefore have very different effects on cellular behavior. The alternative dosage hypothesis suggests that retrovirus oncogenes act by brute force, overburdening cells with too much of what are essentially normal proteins carrying out normal functions. In this view the genesis of cancer by retrovirus oncogenes is related to the amount of the viral proteins rather than to any unique properties they have.

It is too early to know which of these views is correct, but initial indications favor the dosage hypothesis. First, the doses of retrovirus transforming proteins are unquestionably large. The signals directing the activity of retrovirus genes are quite powerful, with the result that the amount of protein produced from a viral oncogene is far larger than the amount usually produced from the corresponding cellular gene; it is clearly possible that the cell might be overwhelmed. More important evidence has come from attempts to test a central prediction of the dosage hypothesis: If retrovirus oncogenes and cellular oncogenes are indeed identical in function, it should be possible to find circumstances under which the cellular genes themselves can induce cancerous growth.

Cellular-Gene Oncogenesis

The first test of this prediction came from the remarkable experiments of Hidesaburo Hanafusa and his colleagues



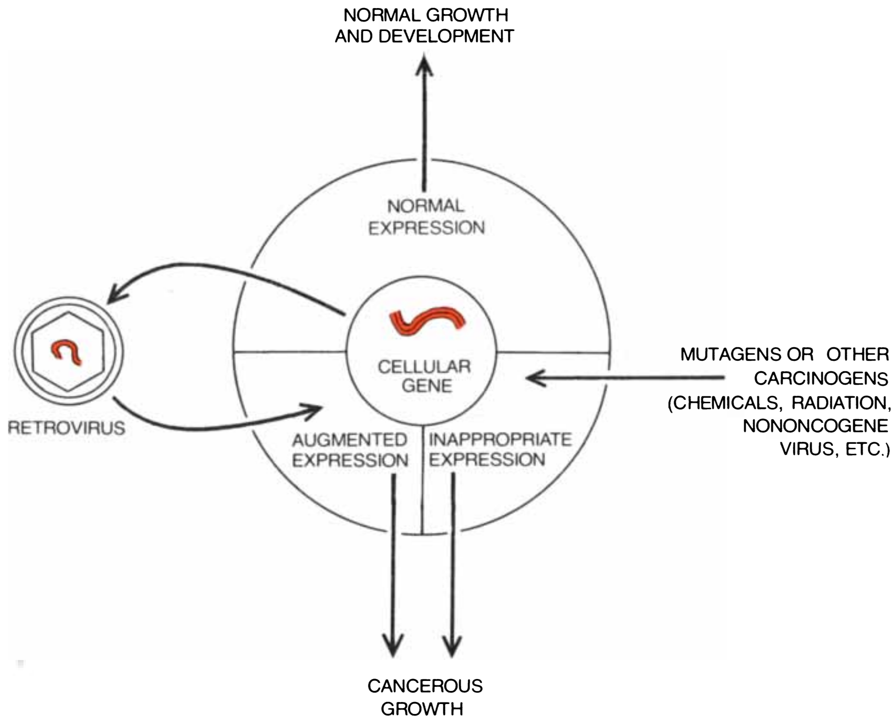
DOSAGE HYPOTHESIS holds that a cellular oncogene directs the synthesis of an amount of a normal protein product that is required for normal growth (*top row*) and that transformation to cancerous growth results from overproduction of the normal protein. In the case of infection by a tumor retrovirus (*1*) the overproduction is directed by the viral oncogene under viral control. The cellular oncogene itself has tumor-inducing potential, however, as has been demonstrated by a recombinant-DNA experiment (*2*). If a viral "promoter" is attached to the cellular oncogene and the activated form of the gene is reinserted into cells in culture, excessive quantities of the oncogene product may be synthesized, transforming the cell. A similar process may explain cancerous growth induced by a wide variety of agents other than retroviruses (*3*). If the cell's DNA is damaged by a mutagen or other carcinogen in some way that happens to increase the activity of the cellular oncogene, the oncogene's product may be made in excessive quantity, converting the cell to cancerous growth.

at Rockefeller University. Hanafusa found strains of the Rous sarcoma virus that had lost large portions of the *src* gene (but not all of it) and were therefore incapable of inducing the characteristic sarcoma in experimental animals. When Hanafusa injected the crippled viruses into chickens and then recovered the virus particles manufactured in the infected cells, he was astonished to discover that the v-*src* gene of the virus had been reconstituted. Apparently genetic material from the c-*src* gene was recombined with the viral genome while the virus was growing in the birds. The virus bearing the reconstituted gene was again fully capable of causing tumors, even though as much as three-fourths of its oncogene had just been acquired from a cellular gene. Hanafusa was able to repeat this extraordinary exercise at will, in quails as well as in chickens. His findings lent weight to the idea that the functions of c-*src* and v-*src* are the same, but many tumor virologists were unpersuaded in the absence of more direct evidence for the tumorigenic potential of the cellular genes.

Now such evidence is at hand. The research groups of George F. Vande Woude and Edward M. Scolnick of the National Cancer Institute exploited the techniques of genetic engineering to isolate three cellular oncogenes (one from mice and the other two from rats) and to show directly that these genes can induce cancerous growth in cultured cells. The feat was accomplished by attaching to the cellular genes a viral "promoter," a DNA-encoded signal that helps to regulate the expression of a nearby gene. In accordance with the dosage hypothesis, when the *src*-promoter complex was introduced into cells, some of the cells were transformed as if they had received a viral oncogene, whereas what they had received was a cellular gene under viral orders to work harder than usual. Moreover, cells transformed by the two rat cellular oncogenes could be shown to make very large quantities of the proteins encoded by the genes, again in accordance with the dosage hypothesis.

Why should an overabundance of a normal protein wreak such havoc? The question can be answered with assurance only when the role of cellular oncogenes in the orderly affairs of normal cells is understood. Perhaps cellular oncogenes are part of a delicately balanced network of controls that regulate the growth and development of normal cells. Excessive activity by one of these genes might tip the balance of regulation toward incessant growth.

There is evidence that the activities directed by viral and cellular oncogenes do help to control the growth of normal cells. Whereas at first the phosphorylation of tyrosine by pp60v-*src* seemed to be an anomalous process whose foreign nature might underlie the cancerous response to *src*, that view was reversed when Stanley Cohen of the Vanderbilt University School of Medicine found a role for tyrosine phosphorylation in the housekeeping of normal cells. Having discovered and purified a small "epidermal growth factor" whose binding to the surface of cells stimulates DNA synthesis and cell division, he considered how the signal for these events might be transmitted from the exterior of the cell to the interior. Cohen first showed that the binding of epidermal growth factor to cells brings about phosphorylation of proteins; prompted by the findings with pp60v-*src*, he ultimately found that the phosphorylation elicited by epidermal growth factor specifically affects tyrosine. Other workers have since shown that some proteins phosphorylated in response to epidermal growth factor can be phosphorylated by pp60v-*src*. A normal stimulant of cell division (epidermal growth factor) and an abnormal one (pp60v-*src*) thus appear to play on the same keyboard. The implication is that tyrosine phosphorylation by pp60c-*src* has a part in regulating the growth of normal cells.



CANCER-GENE CONCEPT, supported by oncogene data and other preliminary evidence, suggests a unifying explanation for various forms of carcinogenesis. The common central element is a group of cellular genes required for normal growth and development. Transplanted into a retrovirus genome (*left*), such a gene becomes an oncogene. Cancer can also result if the cellular gene is affected by any of a wide variety of mutagens or other carcinogens (*right*).

In Search of a Unified Theory

Retroviruses do not seem to be a major cause of human cancer. They may nonetheless have pointed the way to the central mechanisms by which the disease arises. It is generally thought that cancer begins with damage to DNA, although the exact nature of the damage is in dispute. How might the damage cause cancerous growth? Most recent efforts to answer the question in a way that might apply to all forms of cancer have invoked the existence of "cancer genes": components of the normal cellular genome whose activity is unleashed or augmented by carcinogens of various kinds and is then responsible for sustaining the undisciplined behavior of cancer cells. In this scheme cancer genes are viewed not as alien intruders but as normal, indeed essential, genes run amok; the damage done by a carcinogen turns friend into foe, perhaps by acting directly on the cancer gene or perhaps by crippling a second gene that normally polices the activity of the cancer gene.

Medical geneticists may have detected the effects of cancer genes years ago, when they first identified families whose members inherit a predisposition to some particular form of cancer. Now, it appears, tumor virologists may have come on cancer genes directly in the form of cellular oncogenes. In their viral form these genes are tumorigenic, and Vande Woude's and Scolnick's results imply that the cellular genes can also transform cells. It is therefore easy to imagine that cancer genes and the cellular oncogenes revealed by retroviruses are one and the same. The oncogene hypothesis has been restaged with the actors now cellular rather than viral genes; the dosage hypothesis serves to explain why the augmented activity of a normal cellular gene might cause cancer.

Evidence in support of these ideas comes from the study of chicken retroviruses that induce lymphoma, a lethal tumor of the immune system. The chicken lymphoma viruses have no oncogenes. Why then do they cause tumors? William S. Hayward and Benjamin G. Neel of Rockefeller University and Susan M. Astrin of the Institute for Cancer Research in Fox Chase, Pa., have discovered that in tumors induced by the chicken lymphoma viruses the viral DNA is almost always inserted into cellular DNA in the immediate vicinity of a single cellular oncogene (not c-*src* but a more recently recognized oncogene known as c-*myc*). As a seeming consequence of the insertion, the expression of the cellular oncogene is greatly amplified.

These findings fit the concept of cancer genes quite well. The insertion of lymphoma-virus DNA into the host genome is analogous to mutagenesis or other forms of damage introduced by carcinogens of many kinds. The insertion apparently stimulates the activity of a gene that is known to be oncogenic when it appears (as v-*myc*) in a different chicken retrovirus; the stimulated action of the cellular oncogene seems to be responsible, at least in part, for the genesis of tumors. Retroviruses without oncogenes induce a variety of tumors; by identifying the site where the viral genome is inserted into cellular DNA in some of those tumors, virologists may be able to discover cancer genes not yet identified by other means.

The unveiling of cancer genes (in the form of cellular oncogenes) by retroviruses was serendipitous. Must investigators be content with the pace at which retroviruses thus offer up new oncogenes from within the cell? Apparently not. Robert A. Weinberg of M.I.T. and Geoffrey Cooper of the Harvard Medical School have broadened the search for cancer genes beyond the purview of tumor virology. They have shown that gene-length bits of DNA isolated from some tumors (tumors that were not induced by viruses) can transmit the property of cancerous growth when they are introduced into previously normal cells in culture.

Weinberg and Cooper have evidently found a way of transferring active cancer genes from one cell to another. They have evidence that different cancer genes are active in different types of tumors, and so it seems likely that their approach should appreciably expand the repertory of cancer genes available for study. None of the cancer genes uncovered to date by Weinberg and Cooper is identical with any known oncogene. Yet it is clearly possible that there is only one large family of cellular oncogenes. If that is so, the study of retroviruses and the procedures developed by Weinberg and Cooper should eventually begin to draw common samples from that single pool.

A Final Common Pathway

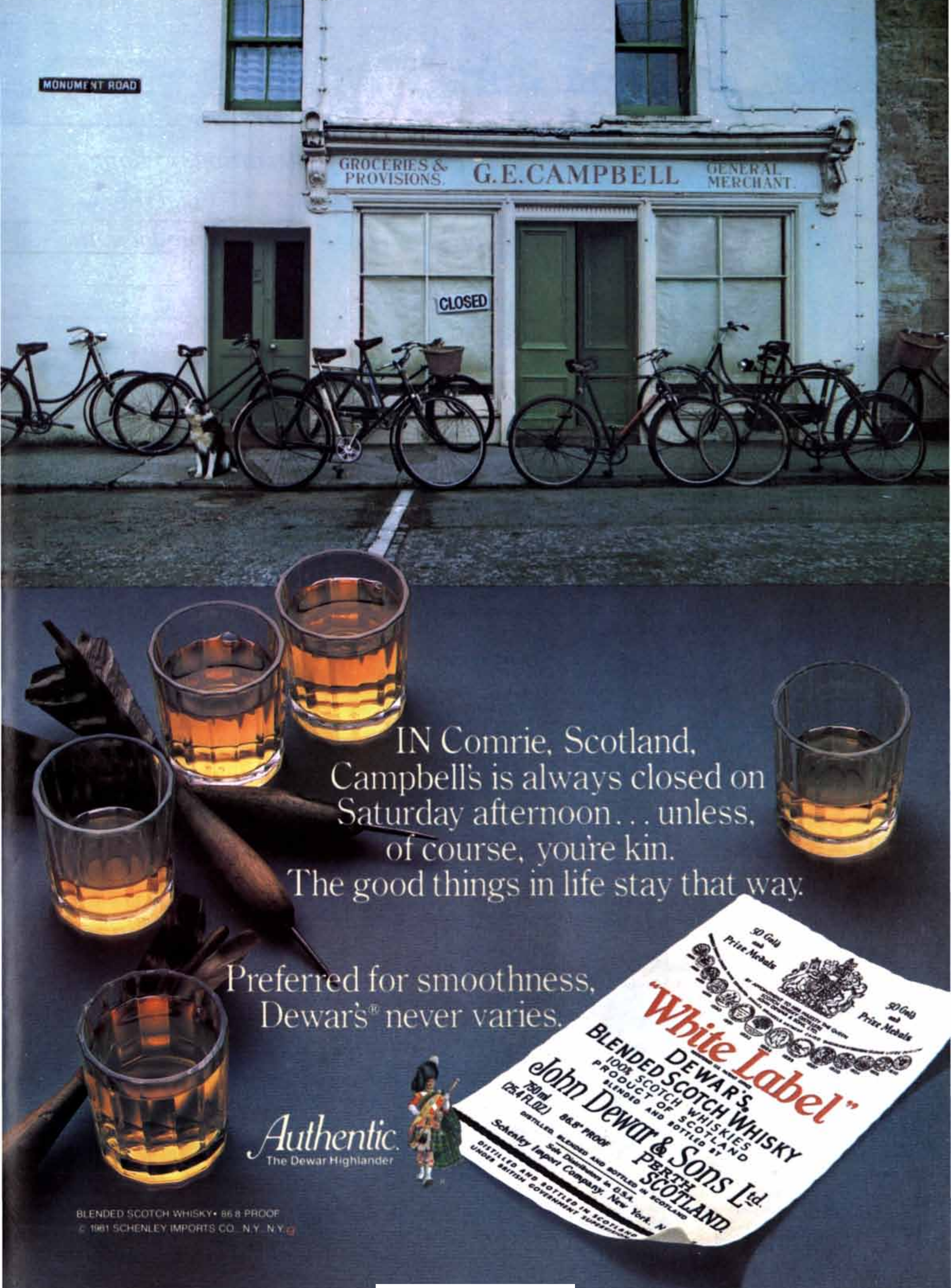
Normal cells may bear the seeds of their own destruction in the form of cancer genes. The activities of these genes may represent the final common pathway by which many carcinogens act. Cancer genes may be not unwanted guests but essential constituents of the cell's genetic apparatus, betraying the cell only when their structure or control is disturbed by carcinogens. At least some of these genes may have appeared in retroviruses, where they are exposed to easy identification, manipulation and characterization.

What has been learned from oncogenes represents the first peep behind the curtain that for so long has obscured the mechanisms of cancer. In one respect the first look is unnerving, because the chemical mechanisms that seem to drive the cancer cell astray are not different in kind from mechanisms at work in the normal cell. This suggests that the design of rational therapeutic strategies may remain almost as vexing as it is today. It will be of no use to invent means for impeding the activities responsible for cancerous growth if the same activities are also required for the survival of normal cells.

However the saga of oncogenes concludes, it presents some lessons for everyone concerned with cancer research. The study of viruses far removed from human concerns has brought to light powerful tools for the study of human disease. Tumor virology has survived its failure to find abundant viral agents of human cancer. The issue now is not whether viruses cause human tumors (as perhaps they may, on occasion) but rather how much can be learned from tumor virology about the mechanisms by which human tumors arise.

ONCOGENE	SPECIES OF ORIGIN	TYPE OF TUMOR	PROTO-ONCOGENE IN VERTEBRATE DNA	ONCOGENE PRODUCT		
				PROTEIN KINASE	PHOSPHORYLATES TYROSINE	LOCATED ON PLASMA MEMBRANE
v-*src*	CHICKEN	SARCOMA	YES	YES	YES	YES
v-*fps*	CHICKEN	SARCOMA	YES	YES	YES	YES
v-*yes*	CHICKEN	SARCOMA	YES	YES	YES	?
v-*ros*	CHICKEN	SARCOMA	YES	YES	YES	?
v-*myc*	CHICKEN	CARCINOMA, SARCOMA, LEUKEMIA	YES	?	?	?
v-*erb*	CHICKEN	LEUKEMIA, SARCOMA	YES	?	?	?
v-*myb*	CHICKEN	LEUKEMIA	YES	?	?	?
v-*rel*	TURKEY	LYMPHOMA	YES	?	?	?
v-*mos*	MOUSE	SARCOMA	YES	?	?	?
v-*bas*	MOUSE	SARCOMA	YES	?	?	?
v-*abl*	MOUSE	LEUKEMIA	YES	YES	YES	YES
v-*ras*	RAT	SARCOMA, LEUKEMIA	YES	YES	?	YES
v-*fes*	CAT	SARCOMA	YES	YES	YES	?
v-*fms*	CAT	SARCOMA	YES	YES	?	?
v-*sis*	MONKEY	SARCOMA	YES	?	?	?

ONCOGENES OF RETROVIRUSES are listed with their species of origin and the tumors they induce, which provide experimental models for most major human cancers. Each of these 15 viral oncogenes is apparently related to a proto-oncogene found in the DNA of vertebrate animals. The product of at least eight of the genes is a protein kinase, a phosphorylating enzyme. Six of the eight kinases have so far been shown to phosphorylate the amino acid tyrosine; at least four of the kinases bind to the plasma membrane of cancerous cell.



MONUMENT ROAD
GROCERIES & PROVISIONS.
G.E.CAMPBELL
GENERAL MERCHANT.
CLOSED
IN Comrie, Scotland,
Campbell's is always closed on
Saturday afternoon... unless,
of course, you're kin.
The good things in life stay that way.
Preferred for smoothness,
Dewar's® never varies.
Authentic.
The Dewar Highlander
"White Label"
DEWAR'S
BLENDED SCOTCH WHISKY
100% SCOTCH WHISKIES
PRODUCT OF SCOTLAND
BLENDED AND BOTTLED BY
John Dewar & Sons Ltd.
PERTH, SCOTLAND
750ml
(25.4 FL.OZ.)
86.8° PROOF
50 Gold and Prize Medals
BLENDED SCOTCH WHISKY • 86.8 PROOF
© 1981 SCHENLEY IMPORTS CO., N.Y., N.Y.

Laser Applications in Manufacturing

Laser tools have found a secure place for drilling, cutting, welding, heat-treating and alloying. Their unique capacities also suggest new approaches to the system represented by a product and its production

by Aldo V. La Rocca

One of the main characteristics of laser radiation is its intensity. "Laser" stands for light amplification by stimulated emission of radiation, and the amplification gives rise to an avalanche of photons (quanta of electromagnetic radiation) in a narrow beam. From the invention of the laser in the 1950's it was recognized that intense laser beams might be a good way to deposit large quantities of energy in materials for manufacturing purposes. That potentiality has now become a mature technology. Over the past decade high-power lasers have found a place in many manufacturing processes: the welding of automobile parts, electronic devices and medical instruments; the heat-treating of automobile and airplane parts to improve their surface properties; the cutting of sheet metal in the punch and die industry, and the drilling of small cooling holes (.007 to .05 inch) in airplane parts. In all these operations laser systems have made production lines more efficient and have reduced costs. Considering the number of applications now in the development stage, it seems certain that over the next decade laser technology will assume a role in many other manufacturing processes.

In manufacturing lasers serve basically as devices capable of applying an extremely high flux of energy to the surface of a workpiece. In this role they have significant advantages over such conventional heat sources as flames, torches, electric arcs and plasma jets. Among those advantages are a product of higher quality (in terms of better performance and a reduction in the number of parts that have to be reworked or scrapped); reduced outlays for materials, labor and processing; high productivity (with resulting reductions in floor space and depreciation costs); a better working environment, and the flexibility and versatility of the laser and the production system based on it.

It is becoming common to speak of two classes of high-power lasers, light and heavy. The classification depends mainly on the power. Light lasers operate in the range from a few tens of watts to a few hundred. They serve in such work as cutting and drilling ceramic substrates in the electronics industry, drilling rubies in the watchmaking industry and cutting not only metal but also cloth, plastics and wood in a variety of other industries. Many of the lasers are fairly small solid-state devices: ruby lasers (with a wavelength of .69 micrometer), neodymium-doped glass lasers and neodymium-doped yttrium aluminum garnet lasers (both with a wavelength of 1.06 micrometers in the infrared). These wavelengths couple well with most metals, making it possible to apply them in welding, drilling, cutting and heat-treating. Ruby and neodymium-glass lasers are usually operated in the pulsed mode, delivering a rapid series of pulses of intense energy to the surface of the workpiece. Neodymium-doped yttrium aluminum garnet lasers can be operated either in the continuous-wave mode (a steady beam) or in the pulsed mode, delivering as many as several thousand pulses per second.

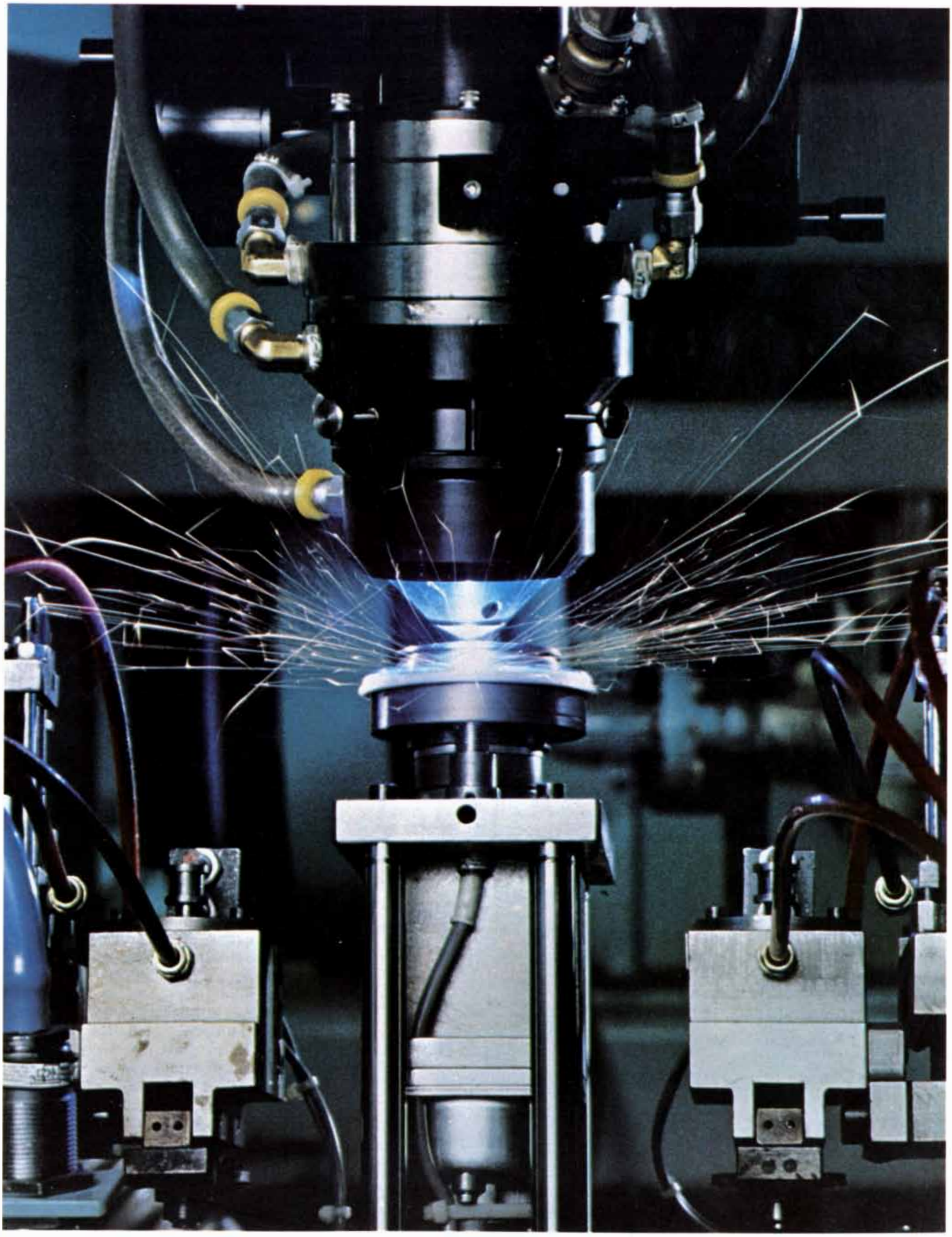
The light class of lasers also includes certain gas lasers (argon and carbon dioxide), which are operated mostly in the continuous-wave mode. The beam emitted by a gas laser has almost total collimation, meaning that it shows little of the divergence that characterizes the beam of, say, a flashlight. Hence the beam can be concentrated in a small spot (ranging in size from micrometers to a fraction of a millimeter), and it delivers power of great intensity. These characteristics are important, particularly in welding by the deep-penetration process, to which I shall be returning. Gas lasers also operate with notably good efficiency, which for the carbon dioxide laser can be as high as 15 percent from the power line to the power of the laser beam.

The heavy lasers range in power from a few kilowatts to a few tens of kilowatts. At present, however, only experimental or laboratory devices exceed 20 kilowatts. The heavy lasers in manufacturing serve in heavy-duty processing such as the welding of pipelines, the welding of automobile parts and the heat-treating of the surface of such parts as crankshafts and the cylinder walls of large diesel engines. The treatment hardens the surface, increasing the resistance of the part to wear. Most of the heavy-duty lasers are carbon dioxide lasers operating in the continuous mode.

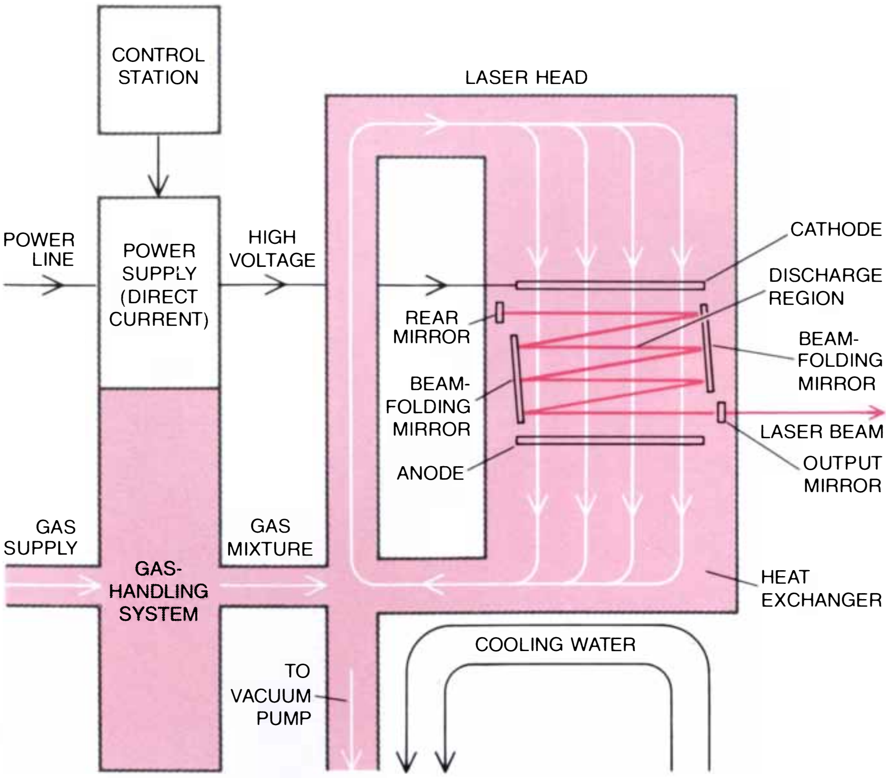
The high flux of electromagnetic energy applied to the surface of the workpiece by a laser is absorbed in an outer layer about 10 nanometers (.000001 millimeter) thick. In that thin layer a heat source of very high intensity is thereby established. An advantage of the laser is that the heat energy is maintained and made to work in the region where the work has to be done. For this reason the energy efficiency is high, ranging from 10 to 1,000 times higher than it is in conventional systems that heat proportionately larger volumes of the workpiece. Laser systems thus achieve notably fast processing times and unique processing properties.

Another important advantage of the laser is that it does not damage parts, since it delivers heat in less time than a conventional source because of the high power density of the beam. Therefore the heat has no time to flow into the part. Conventional sources heat far more of the workpiece than is necessary, giving rise to thermally induced distortion, cracks or stresses that can damage the part, making it necessary to rework or scrap it or impairing its performance. The economic implications are obvious for costly semifinished parts such as gears whose teeth need to be hardened, jet-engine turbine blades in which cooling holes must be drilled and engine blocks whose cylinder bores need to be hardened on the inside.

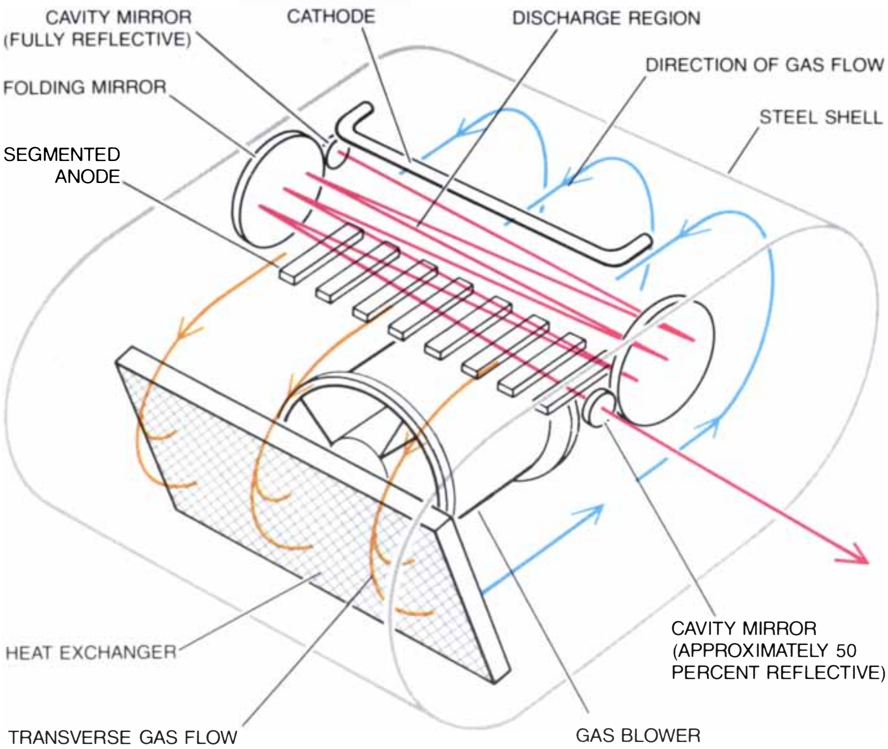
All these advantages result from the extremely high power density of the laser beam. Certain other advantages make the laser beam a highly flexible tool and explain why lasers can often be employed with good results even at the level of power obtainable from conven-



WELDING BY LASER of a synchronizing ring to a transmission gear for an automobile is done at the Mirafiori plant of the Fiat Auto Division in Turin, Italy. The laser beam, which is invisible, is delivered through the cone-shaped nozzle at the end of the welding head, just above the sparks. They are caused by the explosive ablation of surface impurities during a first (tacking) pass made by rotating the workpiece once under the beam at a speed of 75 revolutions per minute. Few sparks are seen during the immediately following second rotation (at 25 r.p.m.) in which the deep weld is made. A 2.5-kilowatt carbon dioxide laser made by Spectra-Physics generates the beam.



INDUSTRIAL LASER developing 1,200 watts of continuous infrared radiation is depicted schematically. It is a gas-transport laser made by Spectra-Physics. A mixture of carbon dioxide, nitrogen and helium is raised to a high energy level (a prerequisite of lasing) by an electric discharge between the cathode and the anode. The laser beam is amplified by seven passes between the mirrors of the optical system. During this stage the gas gets hot, and so it is moved next to the heat exchanger, where it is cooled. The gas is transported at high speed in a continuous loop in which it is successively excited, contributes to the laser beam and is cooled.



PERSPECTIVE VIEW of the laser head of the 1,200-watt gas-transport system shows the cycling of the gas and the folding of the laser beam. On emerging from the output mirror the laser beam is directed to the working head of a tool for welding, cutting, drilling or heat-treating.

tional sources. The beam has no mass and can be easily moved and controlled with short response times. It is easily accommodated in automatic processes. It acts from a distance, eliminating or reducing problems of mechanical interference. By the same token it does not generate mechanical responses, so that the workpiece does not vibrate and does not need to be clamped. Finally, laser technology makes for clean and fast processing that is compatible with work stations along production lines. As a result the technology has significant implications for the logistics and compactness of the production system.

These advantages become apparent in a discussion of the specific things lasers can accomplish in manufacturing. I shall discuss the applications in the order of their acceptance in industry, beginning with drilling and deburring and proceeding to cutting, welding and heat-treating.

In drilling and deburring lasers function as a means of removing material. The laser drilling process is expensive in terms of energy because the material must be melted and then vaporized. (Removing it as a liquid is difficult if the depth of the hole is more than twice the diameter.) For most metals vaporization calls for 10 times more energy than melting.

It should be added that the hot vapor, which is quickly superheated by the laser beam, is highly susceptible to excitation and ionization, the latter consuming at least 100 times more energy than melting does in most metals. The ionized vapor combines with the surrounding gases to form a plume of plasma: an ionized gas. The plasma plume absorbs, reflects and scatters the laser radiation, eventually preventing it from reaching the workpiece. The phenomenon is called blanketing. It is usually controlled by replacing the ambient gas with an inert gas such as argon or helium or a mixture of them. They are more transparent to the radiation and are less easily excited and ionized.

In drilling it has been found that by operating at higher power densities, which can ordinarily be achieved with pulsed lasers, good use can be made of the plasma and fluid-dynamic phenomena associated with the formation of a plume. Through them it is possible to establish processes akin to a laser-supported combustion and, with increasing power intensity, to a laser-supported detonation. The pressures and shock waves of the plasma leaving the hole at high speed (supersonic in a detonation) are quite effective in removing fluids and such solids as slag, oxides and nitrides. Therefore drilling is most often done with solid-state lasers, since with them the pulses can be more easily con-

Order Direct! Commodore VIC-20...The Friendly Computer.

"The first honest-to-goodness full color computer you can buy for only $299.95" – *William Shatner*

Complete Line of Hardware and Software Available to Expand your VIC-20

VIC-20 — Commodore's revolutionary personal computer features color, sound, graphics, programmable function keys, built-in BASIC, expandable memory, low-priced peripherals and more! Connects to any TV or monitor. *Includes RF Modulator,* switchbox, cables and self-teaching instruction book. / $299.95 (See coupon below)

Commodore Datassette — Provides handy economical storage of user-written or pre-recorded programs. / $75.00

VIC Graphic Printer — Economical dot matrix printer makes paper copies of BASIC programs, letters, business data. / $395.00

VIC-3K Memory Expander Cartridge / $39.95

VIC-8K Memory Expander Cartridge / $59.95

VIC-20 Super Expander — 3K RAM memory expansion, high resolution graphics plotting, color, and sound commands. / $69.95

Programmers Aid Cartridge — More than 20 new BASIC commands help new and experienced programmers. / $59.95

RECREATIONAL GAME CARTRIDGES:

VIC AVENGERS • SUPERSLOT • VIC SUPER ALIEN • SUPER LANDER / $29.95 each

COMPUTER PROGRAM TAPES:

(Requires Commodore Datassette)

Recreation Program Pack A — Car Chase; VIC-21; Blue Meanies from Outer Space; Biorhythm/Compatibility; Spacemath; Slither/Super Slither. / $59.95

Home Calculation Program Pack A — Personal Finance I - Home budget; Personal Finance II - Home budget; VIC Typewriter - Word processor for home use; Expense Calendar - income, expenses, appointments; Loan & Mortgage Calculator - Decision-making aid; Home Inventory - Home belongings list / $59.95

Programmable Character Set/Gamegraphics Editor — Lets the VIC user create up to 64 programmable characters and use them in BASIC programs. / $14.95

Introduction to BASIC Programming — A gentle but thorough introduction to BASIC programming. Excellent first book for any new computerist. / $24.95

VIC-20 Programmers Reference Guide — Master VIC-20 reference manual includes information on VIC BASIC, programming and much more. /$16.95

To order accessories simply list on separate sheet and clip to coupon.

Personal Computer Comparison Chart

Product Features	Commodore VIC-20	Atari 400	TI 99/4A	TRS-80 Color Computer
Price*	$299.95	$399.95	$525.00	$399.50
Total Memory Standard (ROM & RAM)	25K	26K	42K	12K
Memory (RAM) Expansion to...	32K	Not Available	Not Available	32K
Keyboard Style	Full-Size Typewriter Style	Flat Plastic Membrane	Half-size Typewriter	Calculator Style
Programmable Function Keys	4	0	0	0
Basic Language	Microsoft Basic	$59.95 Extra	TI Basic	Radio Shack Basic
Upper/Lower Case Characters	Yes	Yes	No	No
RS232 Interface	$49.95	$219.95	$225.00	$19.95
Number of Keys	66	57	40	53
Graphic Symbols on Keyboard	62	0	0	0
Displayable Characters	512	256	64	256

*Manufacturers suggested retail price September 1, 1981

A computer like this would have been science fiction a few years ago. Now it's a reality. It's the new VIC-20 by Commodore, a full fledged expandable color computer that costs little more than video games. And it's so easy to-use you can be writing your first program in 15 minutes!

Everybody loves video games and the Commodore VIC-20 has some of the best. But the VIC-20 can also help children with their homework. Mom can use it for home budgeting. Dad can even take the lightweight, portable VIC-20 to the office for financial and business application.

The Friendly Computer at a Friendly Price: At $299.95 the Commodore VIC-20 is the friendliest way we know to learn computing. It has a full computer keyboard even a small child can operate.

The VIC-20 also plays music, has exciting graphics, lets you create pictures. It even tells you when you've made a mistake and how to correct it. (That's very friendly!)

The VIC-20 can take your children from preschool through post-graduate studies.

Why get just another game that could end up in the closet? Get an honest-to-goodness computer for just $299.95.

Free with every VIC-20 computer
This 164 page guide tells you everything you need to know about your VIC-20 and how to operate it. Written for the beginner, you'll be programming on your VIC-20 in minutes!

Order now. We'll ship your new VIC-20 computer directly to you. 15 day free trial.

Contemporary Marketing, Inc.
790 Maple Lane, Bensenville, IL 60106

Faster Service for Credit Card Customers
Call Toll Free 800-648-5600 (In Nevada call: 800-992-5710)
24 hours a day

VIC-20 THE FRIENDLY COMPUTER

commodore

Yes, I want to start using the new VIC-20 personal computer right away.

Please send me:

______ **Commodore VIC-20 computer(s) at $299.95 each.**
(Item No. 2000). Add $4.95 per computer for shipping and insurance.

Illinois residents include 5¼% sales tax.

☐ Check/M.O. Enclosed ☐ Charge my credit card:
☐ MasterCard ☐ Visa ☐ American Express ☐ Diners Club

Card No. ______________ Exp. Date ______________
Name ______________
Address ______________
City ______________ State ______ Zip ______
Signature ______________

SCAF-060

Contemporary Marketing, Inc.
790 Maple Lane, Bensenville, IL 60106

©CMI 1982-263

trolled than they currently can be with gas lasers.

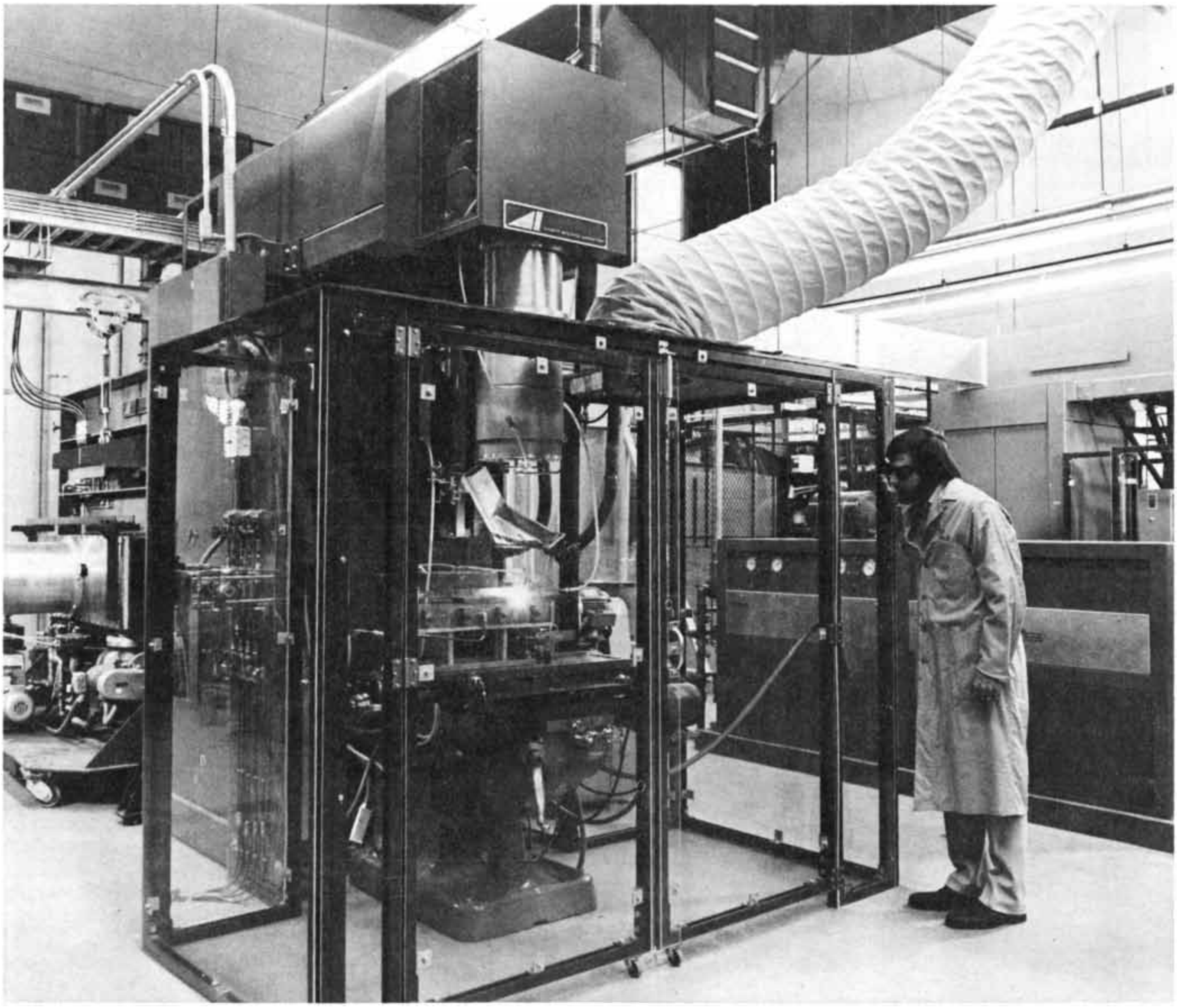
All things considered, drilling by laser remains a costly process. It is therefore industrially justified only for small holes in materials that are difficult to work by mechanical means because they either overload the tool, are too hard or abrasive or are so soft that holes cannot be made in them with high precision. With such materials, however, the advantages of lasers are so compelling that they have almost completely replaced conventional drilling processes.

The greatest success has been in drilling rubies in the watchmaking industry. Lasers are also widely used in drilling the diamonds employed as dies for drawing wire. In the electronics industry they serve for drilling and cutting the ceramic substrates for integrated circuits and for trimming such circuits and discrete electronic components. The aerospace industry uses lasers for drilling holes in parts made of exceptionally hard materials; the parts include turbine blades and the lining of combustion chambers. At the other extreme lasers have found application in making the holes in the nipples for babies' bottles, in rubber and polyethylene tubing and in plastic valves for aerosol containers.

The common factor in these tasks is that the holes are small. The cost of energy for drilling by laser is therefore low compared with other production costs. Moreover, even though the initial cost of a laser drilling system is high, the system pays out quickly because of its high productivity and because it eliminates the need for several other costly tools.

Laser drilling remains too expensive for the heavy metalworking industries because of the large amount of material that has to be removed. Laser tools do hold promise here, however, for the final drilling and the deburring of blind holes, that is, holes that do not go all the way through the material. Such a hole drilled mechanically usually has chips and burrs attached to its wall after the drill is withdrawn. The laser can remove them and do the final drilling if mechanical removal is difficult because the place is hard to get at or the tool might damage the surface finish. Master cylinders, crankshafts and fuel injectors for diesel engines are among the products that seem to be likely candidates.

Cutting with a laser tool calls for less



MACHINE SETUP for welding and cutting and for the heat-treating of metal surfaces incorporates a first-generation 15-kilowatt laser made by Avco Everett. The work station at the right rear is for surface treating; the one in the foreground is for welding and cutting. The operation in progress is the welding of half-inch stainless steel at 25 inches per minute with a beam power of 7.5 kilowatts. The laser delivers a continuous infrared beam that is directed downward onto the workpiece through a telescope. A standard milling-machine carriage moves the workpiece. The power of the beam can be regulated according to the demands of the task from 1.5 to 15 kilowatts.

energy than drilling per unit of volume of material removed because all that is needed is to bring the material to the melting state. The liquid is then removed by a jet of covering gas. The gas can be oxygen or nitrogen if oxidizing and nitriding effects (which can cause the surface of the material to become brittle or undesirably hard, diminishing its resistance to corrosion) are of no concern. Oxygen in particular increases the cutting rate several times because of its exothermic reaction with most metals at high temperatures. High-carbon steels and titanium alloys can be cut in this way because the products of the reaction are solid and are easily removed by the gas jet.

A laser beam put to work in cutting can easily be made fully automatic by computerized numerical control. Such a beam can follow highly complex patterns, starting from the edge or the inside of a sheet of material as is necessary. Cuts of this kind are difficult and sometimes impossible with mechanical tools such as diamond grinding wheels, which are often the only tools capable of attacking a hard material. For this reason lasers seem likely to extend the useful range of such materials.

Running on a fully automatic program, a laser cutting tool can greatly improve precision, repeatability, flexibility and productivity. The combined effect of these improvements on the entire production schedule can make a manufacturing operation more productive by a factor that has been estimated to range from eight to 20. Laser systems bring still other advantages: the absence of vibration, noise and dust, the reduction of fumes and the elimination of the need to buy, stock and maintain a variety of often expensive cutting heads.

Together these reasons explain why laser tools can offer attractive cost benefits in spite of their high initial cost, which is estimated as being from two to five times that of conventional cutting systems. A short depreciation time (a year or less) can be achieved with laser tools, which explains the rapid extension of such tools to a wide variety of cutting tasks. It has been estimated that more than 100 laser cutting systems are installed per year throughout the world, operating in the power range from 400 to 2,000 watts.

The greatest success has been in industries working with precoated sheet metal, organic fibrous materials (wood, leather, fiberboard and textiles), plastic and fiber-reinforced plastic materials, wire-reinforced hose rubber, ceramic materials, quartz, glass, composite materials and exotic aerospace materials. Even for materials that are easy to cut lasers offer unexpected advantages. An example is the cutting of fabrics for making clothes.

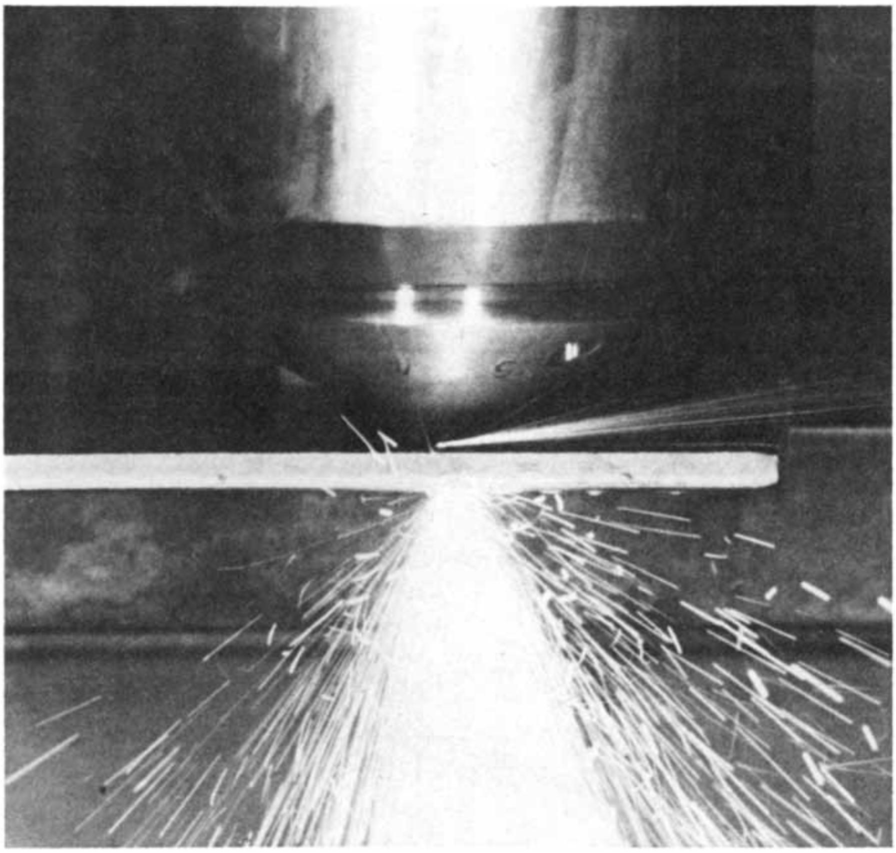
With such fabrics it is often necessary to cut complex patterns. It is also desirable to make a precise cut with well-sealed edges, so that the seaming and sewing operations usually needed to prevent fraying can be eliminated. Fully automatic laser systems operated by computerized numerical control have been notably successful for these purposes. The fabric is cut one layer at a time, eliminating the tendency of piled layers to stick together in the area of the cut. The high speed of the process makes it possible to cut the fabric for more than 40 suits of clothing per hour with a laser of relatively low power (from 200 to 400 watts).

One such system, built by the Culham Laboratory in Britain, has a 400-watt laser whose beam is guided by movable mirrors. The beam cuts a moving web of fabric two meters wide at a feed rate of up to 80 meters per minute. The computer memory holds all the necessary information on styles and sizes and on variations that might be handled as special orders. About 20 similar systems are in production in the U.S. and Britain.

In welding laser tools can operate in two ways: by conduction and by deep penetration. In conduction welding the heat source is held at the surface of the workpiece. The surface must be kept below the vaporizing temperature in order to avoid the loss of material, which would weaken the weld. The welds obtained have an aspect ratio of 1.5 : 1, which is to say that the depth of the weld can be only about 1.5 times the width. Conduction welding therefore serves mainly for joining thin sheets or plates. It is not very efficient (from 8 to 15 percent) and can leave fairly large areas that have been affected by heat.

These drawbacks are overcome with deep-penetration welding, which is the more important of the two processes. The deep penetration is achieved by moving the heat source (the region of maximum temperature) below the surface of the material. The efficiency of the process increases sharply to values higher than 80 percent, notwithstanding the fact that energy-expensive phenomena such as vaporization, excitation, ionization and plume blanketing now come into the picture.

An explanation is found in beneficial effects arising from the interplay of the many complex phenomena that develop when the heat source brought deep into the workpiece creates a core of super-



CUTTING BY LASER is done at the Mechanical Technology Research Institute in Italy. A plate of stainless steel five millimeters thick is fed from left to right at a speed of 1.2 meters per minute under the stationary cutting head. The invisible infrared laser beam passes through a concentric nozzle that also delivers oxygen, which in reacting with the metal heated by the laser induces a combustion that speeds up the cutting. The sparks are caused by the expulsion of liquid or solid particles from the cut. This cutting operation is done at 2.5 kilowatts with a beam generated by an Avco Everett laser that is programmable from 1.5 to 15 kilowatts.

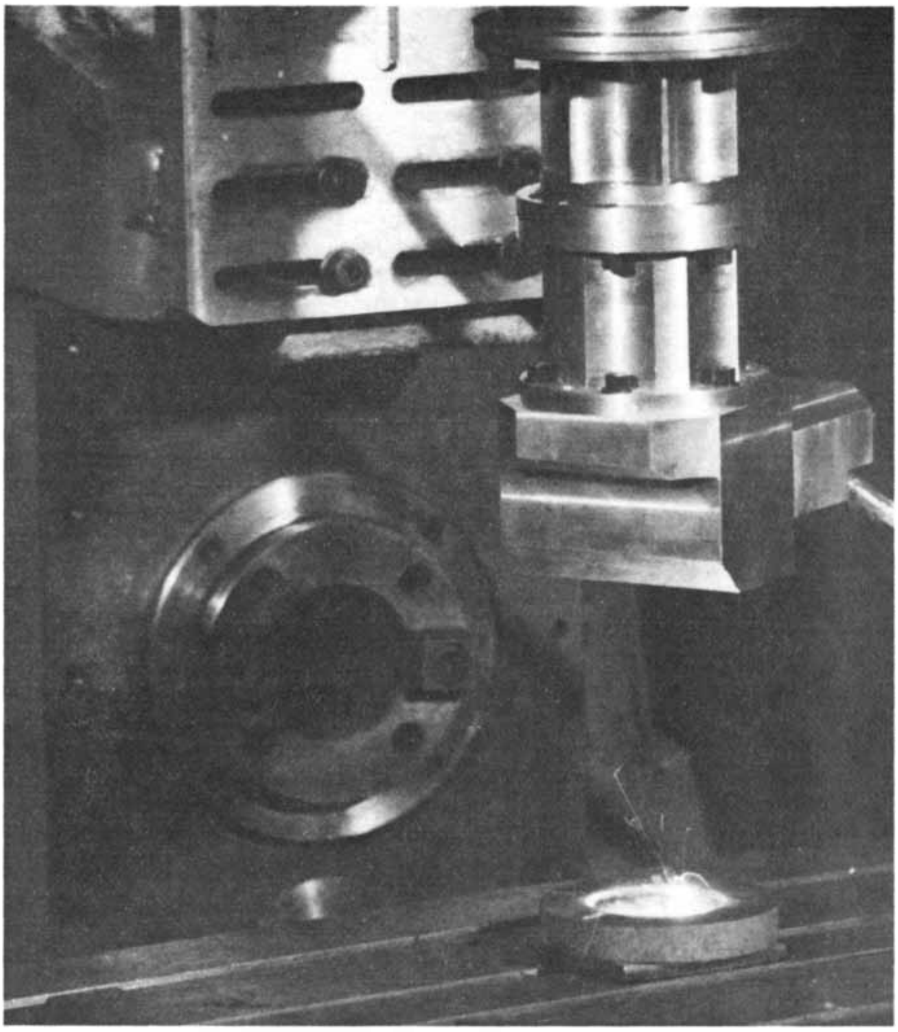
heated vapor surrounded by superheated molten material. The core occupies a small zone that in a full-penetration weld extends throughout the thickness of the material. This zone is called the keyhole. With a well-focused beam of high intensity a keyhole can be established in milliseconds. It can be maintained when it is made to move as the weld is formed.

The molten material is lifted above the outer surface by the pressure of the superheated vapor. There it is held by a combination of gravity, viscosity and surface tension that aids the formation of surface beads, which enhance the mechanical properties of the weld. It is also significant that because of the steep temperature and pressure gradients the fluid material in the keyhole is subjected to violent stirring motions. They induce heat transfer by the transport both of material and of the latent heat of phase changes (such as the change from the gas phase to the liquid phase). The keyhole is thus made to function as a heat pipe, but one that is far more effective than any other kind of heat pipe because it benefits from the latent heats of excitation and ionization, which are orders of magnitude greater than those of the phase transformations that serve in ordinary heat pipes.

The tremendous capacity for transporting heat established in the keyhole means that the material around the keyhole acts as an insulator and inhibits the unwanted flow of heat to the rest of the workpiece. This phenomenon explains the high efficiency of the deep-penetration laser weld, which is evidenced by the high ratios of depth to width (ranging from 5 : 1 to 12 : 1) and the limited extent of the heat-affected area around the weld.

From the metallographic standpoint a deep-penetration weld made by a laser beam has a good crystal structure with few impurities. As a result such mechanical properties as hardness, tensile strength and impact strength are the same as those of the base material and in some instances better. The refining and purifying effects of laser welding have been demonstrated in shipbuilding steel, arctic-pipeline steel, nickel steels and the new HSLA (high-strength, low-alloy) steel.

At first the welding of aluminum and its alloys by laser met with some difficulties because the melt had a low surface tension and low viscosity. The result with sheets more than three millimeters thick was that the melt did not stay in place, so that the weld zone was undesirably thin and porous. A developmental program sponsored by the National Aeronautics and Space Administration overcame these problems for plates of titanium alloy and aluminum by heating the plates before welding. Plates as



VALVE SEAT is alloyed by laser at the Fiat Research Center in Orbassano, Italy. The intense heat a laser beam delivers in a concentrated area makes it possible to alloy the surface of a part after it has been formed. This valve seat is made of cast iron; it is being alloyed with a powder of elements chosen as a substitute for Stellite, an alloy containing almost 50 percent cobalt, a costly strategic material. The new alloy is being tested to see how it compares with Stellite in wear resistance in the environment of hot and chemically reactive exhaust gases in the region where the valve head will repeatedly make contact. In this heat-treating process the laser beam quickly melts and stirs the alloying elements, causing them to diffuse rapidly in the iron.

CLOSE VIEW of the valve seat shows the area where the laser beam, which is moved automatically around the surface being alloyed, is delivering its intense heat. The materials being alloyed with the cast iron are in the form of a powder. Because a laser beam does not make mechanical contact with a workpiece the valve seat does not have to be clamped in position.

much as half an inch thick have now been welded successfully.

Laser welding offers many practical advantages. No welding rods, fluxes or protective materials are needed. Dissimilar materials can be welded. Warping, internal stresses and cracks are minimized, which is particularly important when the materials to be welded are dissimilar or brittle. No physical contact need be made with the parts being welded; it is sufficient that they can be reached by a beam. The beam can be guided by mirrors, which means it can be made to impinge on the inside of a hollow workpiece. Welds can therefore be made in areas that are difficult to reach, and no mechanical stress need be applied to the parts. For example, with mirrors laser systems have proved capable of making deep-penetration welds inside pipelines.

The flexibility of the laser beam and the ease of handling, positioning and controlling it greatly facilitate automatic operation. Hence systems of extremely high productivity can be built to handle complex structures and shapes. In these applications laser technology reduces or eliminates the need for highly skilled people, simplifying the heretofore magic art of making difficult welds.

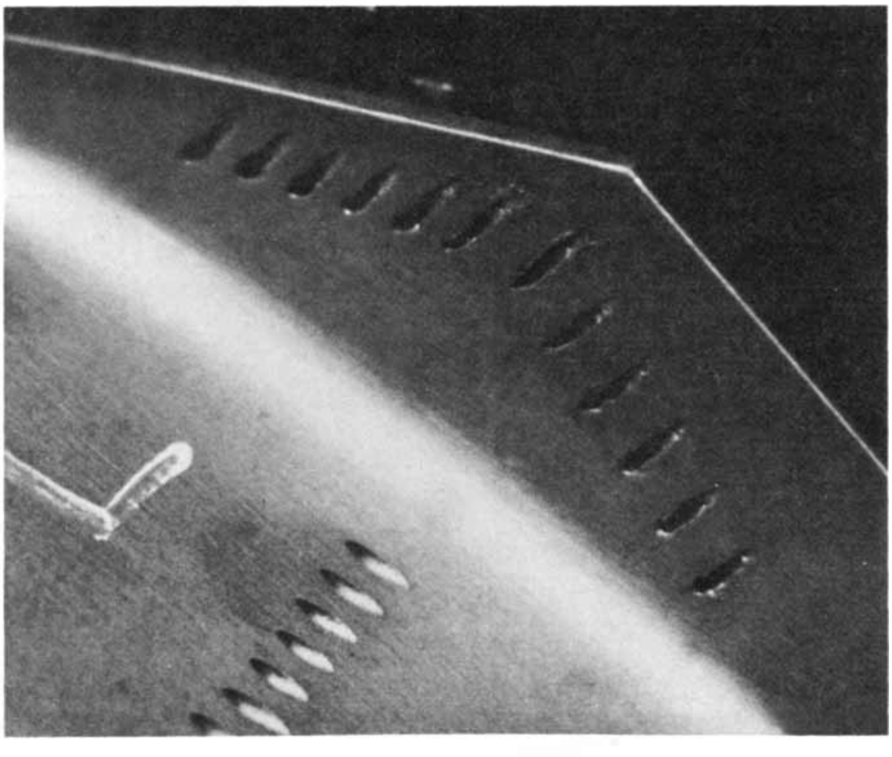
On the other hand, laser welding demands a more accurate positioning of the workpieces. If the gap between two pieces to be welded is more than about 25 percent of the thickness of the pieces and no filler material is added, material may drop out, resulting in welds that are thin or porous. The necessity of precise positioning has not, however, proved to be a serious problem.

Laser welding is found in the electronics and electrical-engineering industries for such tasks as joining dissimilar materials, precoated materials and electric-motor laminates and fixing electrodes to batteries and wires to terminals. As always, the laser welding is superior to conventional welding because it makes a weld with less overall input of heat, with the result that the heat-affected area is a small one. With dissimilar materials the metallurgical refining achieved by laser welding prevents the formation of brittle joints. With precoated materials a laser welding tool will remove the coating from the area of the weld simultaneously with making the weld, whereas with conventional methods the steps must be separate. In welding electrodes to batteries and wires to terminals the small heat-affected area means that the heat of welding does not damage other parts of the battery or the circuit.

An example of what laser welding can accomplish is provided by an automatic system manufactured by the Hamilton Standard Division of the United Technologies Corporation to weld battery electrodes. It is based on a three-kilowatt carbon dioxide laser made by United Technologies. The system is said to have demonstrated at three different plants an increase in productivity of from seven to 10 times over the productivity of conventional systems, offering at the same time greatly improved quality control.

Heavier welding, involving parts that are more than four millimeters thick, is in the transition stage between development and application. A number of companies are investigating lasers for welding in shipbuilding, pipe fabrication and the construction of bridges and prefabricated-steel buildings.

In the aerospace industry a number of welding operations have been conducted with electron beams. Laser beams are now substituting for electron beams because the laser beam does not have to operate in a vacuum and the electron beam usually does. The applications include the internal welding of reinforcing ribs in rocket shells, the internal welding of enclosures for instruments, the welding of aluminum and titanium vessels for holding gas at high pressure and the welding of blades and gears for turbine engines. The automotive industries are in the first phases of applying laser systems in the precision welding of such high-volume parts as clutch and differential housings, gears, axles, exhaust-gas sensors and brake components.



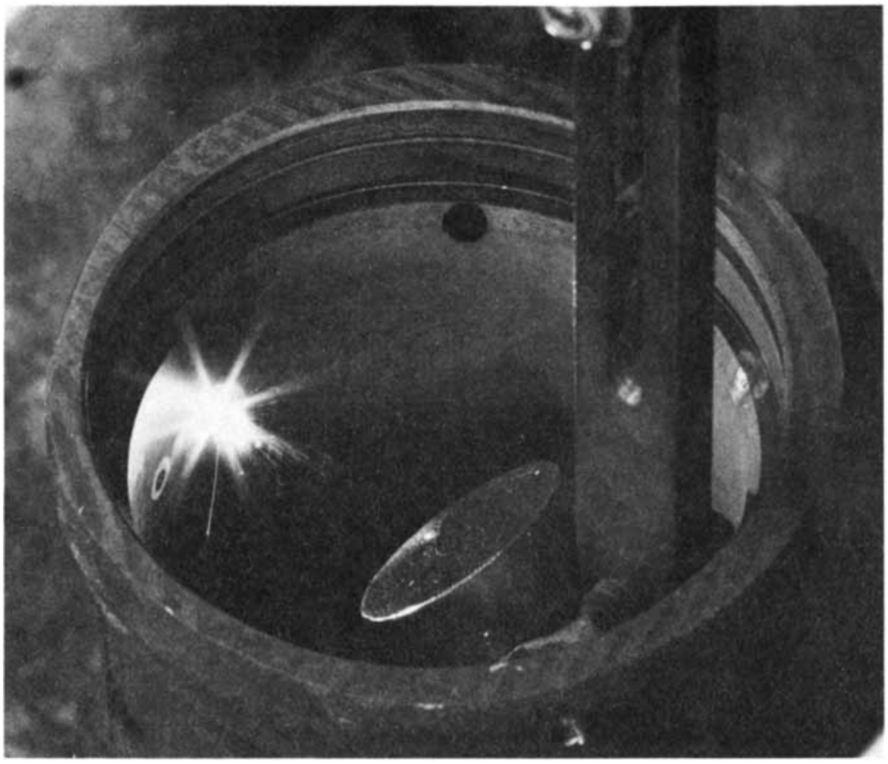
DRILLING BY LASER achieves a complex pattern of tiny cooling holes in a vane for a jet engine. For the part to function properly the diameter and depth of the holes and the angle of entry of the laser beam must be controlled precisely. This drilling was done at the Pratt & Whitney Aircraft Division of the United Technologies Corporation by a pulsed ruby laser that delivers 45 kilowatts of power to the workpiece in less than one millisecond per pulse. The industrial laser that does the drilling of the cooling holes was manufactured by United Technologies.

Applying the heat generated by a laser to treat the surface of a part is a process of great importance in several mass-production industries. The purpose of the treatment, whatever the source of the heat, is to give the surface of the part special properties needed for long-term operation under demanding conditions such as heavy mechanical loads, high temperatures and chemically aggressive environments. Two of the main heat-treatment processes are cladding and surface alloying.

Cladding adds to the base material of a machine part a layer of another material with the desired properties. An example is cladding steel with a cobalt alloy. To such a process laser heat treatment brings higher productivity and clean and controllable operation through fully automatic systems. Cladding can now be done on factory production lines rather than in outside metallurgical shops.

Surface alloying is a process that has been made industrially possible by the laser. The material added can be in the form of powders, rods, sheets or rings of the alloy or powders of the alloying elements themselves, which can now be alloyed on the spot. The implications for the design and production of parts are striking. Operating at high beam intensity, the laser quickly generates the high temperatures and temperature gradients required for alloying. Through them the added material is melted, vigorously stirred and caused to diffuse into the underlying base material. Because of the powerful driving forces extremely fast rates of diffusion are attained. Because cooling is fast the alloyed layer has a fine microstructure. Moreover, the distribution of the alloying elements in depth can be closely adjusted by the appropriate selection of the conditions under which the laser's heat is applied.

Surface alloying by laser is valuable because it endows parts with unusual properties and yields parts of notably high quality. It also results in significant economic gains, which stem from major reductions in the amount of costly materials consumed because with a laser such materials are added only to the small areas where high performance is required. Further gains result because the bulk material is easier to work and because only a small amount of the hard alloy material has to be removed in finishing operations. (Often the tool in such an operation is a costly diamond grinding wheel, since nothing else will attack the hard material.)



SURFACE HARDENING of the inside of a power-steering gear housing is done by laser at the Saginaw Steering Gear Division of the General Motors Corporation. The housing is made of cast ferritic malleable iron. The beam from a 1,000-watt carbon dioxide laser is directed downward through a focusing lens onto a gold-plated mirror, which is moved up and down inside the housing to harden the surface in several equally spaced strips. The laser beam is invisible; the bright spot in this photograph is from the area of the bore that the beam is heating.

A third major form of heat-treating is phase transformation, in which a surface can be hardened (with no addition of material) by causing the base material to change from one solid phase to another that is harder. To avoid making the part brittle it is important to keep the hardening from going too deep. Usually a layer a few millimeters thick is sufficient to provide the necessary hardness. The same thickness seems sufficient to achieve a state of compressive stress that improves the resistance of the part to fatigue.

In surface hardening the laser beam generates a temperature profile that is chosen to keep the surface below the melting point while causing the phase transformation down to the desired depth. Since the heat is applied only in a surface layer, it is quickly quenched by the cold bulk material. This fast self-quenching gives the treated layer a fine grain structure. In extreme cases, as in the process known as laser glazing, the result is a surface layer that is totally amorphous or glassy.

Hardening by phase transformation is the only heat-treating now being done with lasers on an industrial scale. A major operation has been installed by the General Motors Corporation at its Saginaw Steering Gear Division. There some 20 laser systems heat-treat the guiding surfaces on the housing of power-steering units at a rate that can reach 30,000 units per day. General Motors also has at its Electromotive Diesel Division several five-kilowatt laser systems that heat-treat the inside walls of the cast-iron-alloy cylinder sleeves for the large diesel engines made at the plant.

Although the achievements I have described are impressive, they represent only a beginning of the contributions laser systems can be expected to make to industry. What one sees now are first-generation laser systems at the lower power levels (from 200 to 2,000 watts). Lasers of higher power are becoming available, and increasing numbers of them are being installed on production lines. It seems likely that when reliable systems become available in the range from 10 to 20 kilowatts, they will affect a sector of much greater interest for the economies of the industrialized nations, namely the mass-production industries. In order to achieve this transition effectively the nations and industries concerned must make preparations, not only by developing improved laser systems but also by reviewing the entire spectrum of industrial processes to ensure that the new technology's full potential for productivity and versatility is exploited.

The novelty of the new tool, arising from disciplines and technologies foreign to the present generation of production engineers, will initially make its introduction difficult on conventional production lines. Therefore what is needed is a pragmatic approach, starting with relatively simple applications of assured success. Most likely they will be single-purpose applications, relying on the combined effects of several favorable factors rather than pushing as far as possible the productive capability of the laser system.

At the same time it will be necessary to prepare for a second phase of multipurpose, versatile systems that do exploit fully the productivity made possible by lasers. Although the improvements already obtained are regarded as being substantial, they represent a small fraction of the potential. They appear to be satisfactory only because today's production systems are incapable of accepting higher rates of productivity, particularly in the metalworking steps, where lasers have the most to offer.

The situation will change radically when the product and the plant are designed as a unified system that fully utilizes the advantages offered by laser technology. The design will have several well-balanced, highly productive multipurpose working lines that feed into a single second-generation laser system. The laser system will be centrally located and flexible and will perform in a superior manner all the metalworking steps for all the feeder lines. Then the laser's extremely high productivity, which sometimes seems useless in today's manufacturing plants, will be properly exploited and the revolutionary impact of laser technology will be fully felt.

"A phone that automatically redials a busy number? GEE!"

(No, GTE!)

GTE's unique Flip-Phone® II is like something out of a spy movie.

First of all, the entire unit is just one, small, compact piece that flips open when picked up. (Thus, its name.)

It has the extraordinary capacity to automatically redial the last number you dialed. All you do is simply press a button. (Extremely handy for redialing things like busy numbers, people who aren't in, or just someone you want to call back.)

It also has a one-button feature that lets you talk to someone who's in the room with you without the party on the other end of the line hearing (though you can still hear them).

The Flip-Phone® II telephone is really one of the most advanced as far as home phones go. What's more, it's fun.

GTE

Superclusters and Voids in the Distribution of Galaxies

Red-shift surveys of selected regions of the sky have established the existence of at least three enormous superclusters of galaxies. The surveys also reveal that huge volumes of space are quite empty

by Stephen A. Gregory and Laird A. Thompson

Astronomers and cosmologists are much preoccupied these days with explaining the emergence and distribution of aggregates of matter in the universe. How soon after the big bang, the presumed explosion of the primordial atom some 10 to 20 billion years ago, did matter begin to coalesce into the stars and galaxies we see today? Assuming that matter was more or less evenly dispersed before coalescence began, is the universe on the grand scale uniformly populated today by stellar aggregates of one kind or another? Recent observations by several groups of astronomers are helping to answer these questions. Large-scale surveys have verified the existence of superclusters of galaxies: organized structures composed of multiple clusters of galaxies. Each cluster, in turn, may consist of hundreds or thousands of individual galaxies. Although the existence of superclusters has long been conjectured, their confirmation has been accompanied by at least one major surprise: equally large regions of space contain no galaxies at all.

Superclusters are so vast that individual member galaxies moving at random velocities cannot have crossed more than a fraction of a supercluster's diameter in the billions of years since the galaxies came into being. Evidently superclusters offer an insight into evolutionary history that is simply not obtainable with smaller systems. At scales smaller than those of superclusters the original distribution of matter is smeared out by evolutionary "mixing." Astronomers hope that an understanding of the largest structures in the universe will clarify the processes that give rise to structures of all dimensions, ranging downward from galaxies to stars and planets.

It is impossible to determine who first conceived the idea that clusters of galaxies might be members of still larger aggregates, namely superclusters. As one reads old technical papers on extragalactic astronomy one is struck by the similarities between the speculations of 50 years ago and the better-understood concepts of today. What our immediate predecessors lacked were the observational tools that have finally provided the evidence to substantiate some of the early speculations. Although observations in the X-ray, ultraviolet, infrared and radio regions of the electromagnetic spectrum have opened exciting new windows on the universe, it is fair to say that the most important information for cosmology has been collected by telescopes that gather visible and near-visible light.

Even before the invention of the telescope observers could see in the night sky not only planets and stars but also nebulous patches of light. As large telescopes evolved in the 19th century some of the nebulas were resolved into individual stars. Some astronomers contended that such nebulas were huge independent systems of stars outside our own system. In any event the positions of nebulas listed in John Herschel's *General Catalogue* of 1864 and in J. L. E. Dreyer's *New General Catalogue* of 1888 clearly show clustering.

As a consequence astronomers who believed some of the nebulas were external systems began to speculate about their clustering tendency. In 1908 the Swedish astronomer C. V. L. Charlier outlined a cosmology characterized by a hierarchy of clustering. He identified a number of nebular clusters, two of the largest being in the constellations Virgo and Coma Berenices. In 1922 the British astronomer J. H. Reynolds observed that a band of nebulas seemed to stretch from Ursa Major through Coma Berenices to Virgo, a distance of some 40 degrees across the northern sky. Although Reynolds believed these nebulas were within our own stellar system, he may be regarded as the first to identify what is now recognized as the Local Supercluster, of which our galaxy is a member.

By the mid-1920's Edwin P. Hubble of the Mount Wilson Observatory proved that many nebulas are indeed external galaxies. By 1929 he had announced the profound discovery, made in collaboration with Milton L. Humason, that the more distant the galaxy is (as judged by its relative faintness), the more its light is shifted toward the red end of the spectrum. Such red shifts indicate that the galaxies are streaming away from ours (and from one another) as part of an expansion of the cosmos. Today the red shift/distance relation is called Hubble's law. It is the basis of modern observational cosmology.

The value of the red shift, z, is obtained by subtracting the rest wavelength, or non-red-shifted wavelength, of a galaxy's stellar spectral lines from the observed wavelength and dividing the difference by the rest wavelength. The largest red-shift value found by Humason (in the late 1940's) was .2, equivalent to a recession velocity of 60,000 kilometers per second, or 20 percent of the speed of light. There is no general agreement on the value for the Hubble relation, but a plausible one equates a recession velocity of 75 kilometers per second with a distance of one million parsecs, or 3.26 million light-years. Humason's galaxy with a red shift of .2 is therefore about 2.6 billion light-years away. The most remote objects known, the quasars, with red shifts as large as 3.5, are apparently receding at more than 90 percent the speed of light. They are presumed to be about 15 billion light-years away.

In the 1930's both Hubble and Harlow Shapley of the Harvard College Observatory drew attention to the fact that in the northern galactic hemisphere (which includes Virgo and Coma Berenices) bright galaxies are more numerous than they are in the southern hemisphere. Hubble also photographed many galaxies so faint that he believed he had found a probable end to the phenomenon of clustering. "For the first

CORE OF THE CLUSTER OF GALAXIES in the constellation Coma Berenices, a small part of the vast supercluster that includes not only the Coma cluster but also another rich cluster, A1367, is depicted in a color image reconstructed from black-and-white photographs. At least 300 moderately bright elliptical galaxies and galaxies of the type designated S0 can be counted in the original photograph, each a giant collection of tens of billions of stars at a distance of some 300 million light-years. The only prominent object in the picture that is not a galaxy is a bright blue star, slightly above and to the right of the center, that is a nearby member of our own galaxy. The two giant elliptical galaxies near the center are dominated by the ruddy light from much older stars. The picture was prepared at the Kitt Peak National Observatory with the aid of a computer-controlled television monitor in which one can carefully balance colors appropriate to black-and-white images recorded separately in the red, green and blue (or ultraviolet) regions of the electromagnetic spectrum. The original photographs were made by one of the authors (Thompson) with the 1.2-meter Schmidt telescope on Palomar Mountain.

SPIRAL GALAXY NGC 4535 is in the constellation Virgo at a distance of some 60 million light-years. It is a member of the Local Supercluster of galaxies centered on the Virgo cluster. Our own galaxy lies in a sparse cluster called the Local Group, which is now thought to be an outlying member of the Local Supercluster. The red-orange hue of the nucleus of NGC 4535 indicates an abundance of old, cool stars. Blue knots in the spiral arms mark regions where hot young stars have recently formed. NGC 4535 is believed to resemble our galaxy in shape, size and luminosity. The picture was reconstructed from images made with the Palomar Schmidt telescope by Thompson.

time," he said, "the region now observable with existing telescopes may possibly be a fair sample of the universe as a whole."

Another of Hubble's contributions, which is pertinent to our discussion, is his classification scheme for the forms of galaxies. He divided the galaxies into two main classes, ellipticals and spirals, with several subcategories. Elliptical galaxies range from the spherical to the lenticular and are generally devoid of structural features. Spiral galaxies, such as our own, are flattened disks in which spiral arms are usually visible (unless the disk is seen from the edge). Later a third group, the S0 galaxies, were found to have intermediate properties.

By 1950 astronomers could agree on the general characteristics of clusters of galaxies. Several very rich clusters were known, the largest being the cluster in Coma, with more than 1,000 individual members. Clusters had been found to consist predominantly of elliptical and S0 galaxies. No more than half of all galaxies appeared to lie within clusters; the rest, classified as "field" objects, were thought to be isolated galaxies, mostly spirals, that lie well outside clusters. A few astronomers suggested that the Virgo region might consist of more than a single cluster, but Charlier's model of a hierarchy of successively larger clusters seemed doomed by Hubble's counts of faint galaxies.

Gerard de Vaucouleurs of the University of Texas at Austin, who has been studying the brighter galaxies in the northern galactic hemisphere since the early 1950's, was the first to define and describe the Local Supercluster. His work shows that it is centered on the Virgo cluster, about 60 million light-years away, and has perhaps 50 outlying clusters called groups, along with individual galaxies scattered among them. Our own galaxy lies in one of the sparse clusters, in what astronomers call the Local Group, so that it is almost certainly an outrider of the great supercluster.

A second major development in the past 30 years is the growing awareness that the Local Supercluster is not unique. Between 1950 and 1954 the entire northern sky was mapped with the wide-angle 1.2-meter Schmidt telescope on Palomar Mountain. Soon thereafter George O. Abell of the University of California at Los Angeles compiled a catalogue of 2,712 rich clusters of galaxies. Abell pointed out that many of the clusters seemed to be members of superclusters with an average of five or six clusters each. His proposal was disputed, however, on the basis of another catalogue of clusters compiled from the same survey by Fritz Zwicky and his colleagues at the California Institute of Technology. The Zwicky catalogue suggested that the clusters do not form clusters of a higher order. The disagreement can be resolved by recognizing that the clusters as defined by Zwicky are generally larger than Abell's clusters and include several centers in which galaxies are concentrated.

At about the same time, but on the basis of a different sky survey (one completed at the Lick Observatory), Jerzy Neyman, Elizabeth L. Scott and C. D. Shane of the University of California at Berkeley reported finding huge "clouds of galaxies," their term for superclusters. They also suggested tentatively from their study that every galaxy in the universe belongs to a cluster; there might be no isolated galaxies. More recently a thorough and formal analysis of all available galaxy catalogues has been carried out by P. J. E. Peebles and his colleagues at Princeton University. They have quantified the entire spectrum of galaxy clustering and conclude, among other things, that clusters tend to lie close to one another.

A third major development in the analysis of the clustering phenomenon since 1950 has been the introduction of large-scale red-shift surveys. The first step in such a survey is to measure the red shifts of all galaxies brighter than some faint limit in a selected area of the sky. By applying Hubble's law to the observed red shifts the distance of each galaxy can be estimated. This approach has certain clear advantages over the statistical analysis of existing catalogues of galaxies, which provide only the two angular coordinates of galaxy positions, namely the two that can be measured on the plane of the sky. On the basis of the existing catalogue data the third dimension, distance, can only be inferred approximately from the faintness of the galaxy. In the red-shift-survey method the distance is derived explicitly from Hubble's law. The disadvantage of the method is that whereas thousands of galaxy positions can be derived from a single photograph, spectral red shifts can be obtained only one at a time. The two methods are nonetheless complementary. Catalogue studies can analyze large numbers of galaxies in large volumes of the universe; red-shift studies supply three-dimensional detail in much smaller sampling volumes.

Large-scale red-shift surveys have been made possible by major advances in telescope instrumentation. Although Hubble and Humason had access to the largest telescopes (the 100-inch reflector on Mount Wilson and later the 200-inch on Palomar), the available photographic emulsions were slow compared with current ones. Modern spectrographic cameras usually include an electronic device that intensifies the image by a factor of at least 20 before it enters the detector. For certain observations it is sometimes possible to use digital detectors so sensitive that they are capable of counting individual photons. As a result of these advances and others astronomers can now record as much spectrographic information in half an hour as Hubble and his contemporaries could record in an entire night.

It is of historical interest that the first red-shift survey was presented at a conference held in 1960 on the application of image-intensifying tubes to astronomy. With one of the new devices Nicholas U. Mayall, working with the 120-inch reflector at the Lick Observatory, had recorded the spectra of 40 of the 82 brightest galaxies lying within four degrees of the center of the cluster of galaxies in Coma Berenices. In 1972 Herbert J. Rood and Thornton L. Page of Wesleyan University completed and extended Mayall's initial survey. Additional red shifts were recorded by Eric C. Kintner of Wesleyan, who then analyzed the enlarged sample in collaboration with Rood, Page and Ivan R. King of the University of California at Berkeley. Their paper represents the first modern, detailed study of the red shifts in a single cluster of galaxies. They reported that the cluster consists predominantly of elliptical and S0 galaxies, some moving with speeds of more than 1,000 kilometers per second, and that they could place no limit on the cluster's size.

Some four years later William G. Tifft of the University of Arizona and one of us (Gregory) extended the Coma Berenices survey to both fainter and wider angular limits. We found that the Coma cluster itself ends three degrees from the center but that a number of galaxies form an armlike projection pointing westward toward the nearby cluster A1367 and perhaps linking up with it. (A1367 stands for cluster No. 1367 in Abell's catalogue. The Coma cluster itself is A1656.) Our analysis emphasized that red-shift surveys yield not only a detailed picture of distant clusters of galaxies but also important information about galaxies that may lie in the foreground. Because the galaxies in the foreground seem to be found in the sparse collections called groups (or clouds if they are even sparser) a single red-shift survey is able to identify collections of many different sizes, from the biggest down to the smallest. Indeed, the sparse foreground samples may have as much to tell about how clusters form as the more dramatic rich clusters. Our analysis also drew attention to the paucity of field galaxies.

In a rapidly moving area of investigation the same observations or similar ones are often made independently by different workers. It happened that Rood and Guido L. Chincarini of the University of Oklahoma, who had been studying galaxies to the west of the Coma cluster, found that the Coma cluster's westward arm was still detectable at a distance of more than 14 degrees. They also suggested that a bridge of galaxies might join the Coma cluster and A1367. At this stage the two of us (Gregory and Thompson) initiated a survey that systematically extended the Coma survey westward all the way to A1367. Our more complete analysis confirmed the existence of a bridgelike connection between the two clusters. Hence a study that had originated with Mayall's observation of 40 galaxies in an area of 16 square degrees ultimately covered 238 galaxies in 260 square degrees and in the process established the existence of a true supercluster.

The Coma cluster lies near the extended pole of our own galaxy, nearly 90 degrees away from the veil of dust and gas that limits visibility in the central plane of the galaxy. In our study of the Coma-A1367 supercluster we decided to secure spectra of all galaxies brighter than magnitude 15, about a million times fainter than Vega, the brightest star in the northern sky. When the galaxies in our sample are simply mapped in two dimensions, as they appear in the sky, one can see two main concentrations: the Coma cluster itself in the northeastern corner and A1367 in the southwestern corner [*see illustration on page 112*]. Otherwise one has a strong visual impression that the map is made up of many unattached galaxies distributed more or less randomly between the two centers.

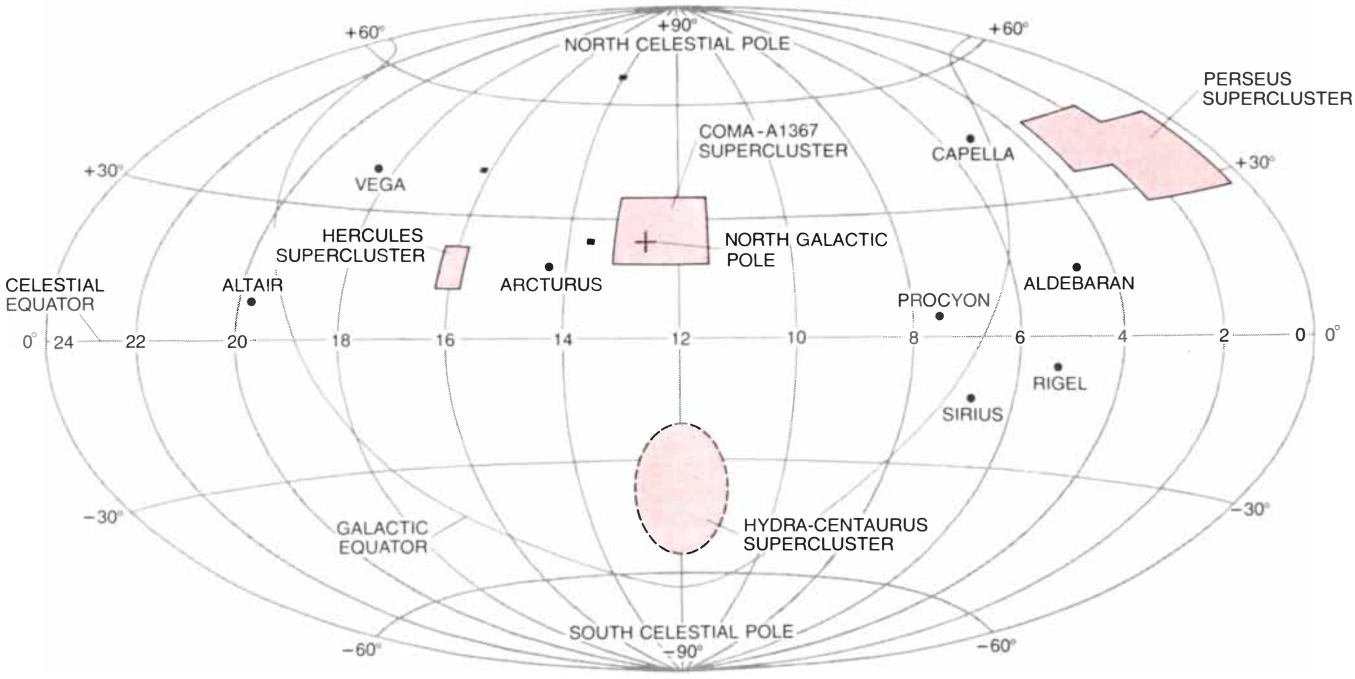
If the results of the red-shift survey are now plotted to show how the same galaxies are distributed in the third dimension, that is, according to distance, a quite different picture emerges. For this purpose it is sufficient to use two positional coordinates: radial distance (derived from red shifts) and east-west angles in the sky [*see illustration on page 113*]. In this picture the galaxies are seen to be much less smoothly distributed.

There are a few small clumps close to our galaxy, represented by the vertex of the wedge. The most impressive feature, however, is a densely populated region stretching all the way across the map at a distance of 315 million light-years. This is the feature that qualifies as a supercluster, since it incorporates the two rich clusters, Coma and A1367, and several less populous clusters that together form a continuous megagalactic structure stretching more than 70 million light-years from end to end.

Somewhat surprisingly the map also shows a few "voids," regions seemingly empty of galaxies. At the time we completed the study we felt confident that the voids were real, but we had doubts about their universality. Conceivably, we thought, they were peculiar to this one region of the sky.

Since superclusters themselves may be as diverse in their structure and content as individual galaxies one would like to have examples other than the Coma-A1367 supercluster before generalizing about supercluster properties. At least three more large systems are currently under study, and all of them exhibit intriguingly different characteristics. The region around the Hercules cluster has recently been investigated by one of us (Thompson) in collaboration with Chincarini, Rood, Tifft and Massimo Tarenghi, working with the two-meter telescopes at the Steward Observatory of the University of Arizona and at the Kitt Peak National Observatory. Once again the observational evidence shows the presence of a complex supercluster occupying a broad band at a distance of between 400 and 600 million light-years. Unlike the Coma-A1367 supercluster, the Hercules system is not dominated by one or two clusters. Nevertheless, it is similar to the Coma-A1367 supercluster in having a vast empty region in the foreground. Perhaps the most surprising aspect of the Hercules system, however, is that its densest clusters are dominated by spiral galaxies rather than elliptical ones. This peculiarity alone makes the Hercules system remarkable.

The third supercluster under study is in the region of the sky that coincides with the constellations Perseus and Pisces. It appears to be a long filament that runs more than 40 degrees across the sky, from the well-known Perseus cluster of galaxies to a small group of galaxies near the elliptical galaxy NGC 383. Our new observations, made in collaboration with Tifft, indicate that the depth of the apparent filament is no greater than its width in the sky. Consequently we can presume not



LOCATION OF FOUR SUPERCLUSTERS is shown on a homolographic projection of the celestial sphere. The Hydra-Centaurus cluster in the southern hemisphere has not yet been intensively studied. The first supercluster to be recognized was the Coma-A1367 system. The authors estimate that it holds well over a million galaxies, 280 of whose red shifts have now been measured. It occupies a volume of at least 10^{23} cubic light-years. The three black squares identify regions where a recent survey shows a near absence of galaxies with recession velocities of between 12,000 and 18,000 kilometers per second, corresponding to the interval between 520 and 780 million light-years. If the region between the three sampled sites is equally lacking in galaxies, it would be a void of some 30×10^{24} cubic light-years.

only that the system is a true filament but also that the individual galaxies within the filament have random motions of rather low velocity. There is also evidence that many of the individual galaxies in the Perseus-Pisces system have planes of rotation that are either parallel to the axis of the supercluster filament or perpendicular to it. This observation, if it is confirmed, may tell something about the way galaxies and superclusters are formed.

The three red-shift surveys cover only about 2 percent of the entire sky. Groups at several observatories are trying to obtain a more comprehensive view of the superclustering phenomenon. For example, Jaan Einasto, Mikhel Jôeveer, Enn Saar and S. Tago in the Estonian S.S.R., who independently discovered the Perseus supercluster and the existence of voids, have analyzed the largest existing catalogue of galactic red shifts. Although their catalogue lacks the details of our recent red-shift surveys, it has enabled them to verify on a larger scale the same features found by the more detailed survey methods.

Similarly, Chincarini and Rood have analyzed the distribution of giant spiral galaxies first observed by Vera C. Rubin, W. Kent Ford, Jr., and their colleagues in the Department of Terrestrial Magnetism of the Carnegie Institution of Washington. The Rubin-Ford survey covers the entire sky but has little detail in any one region. It has nonetheless enabled Chincarini and Rood to verify the existence of the three superclusters we have described and to add a previously unrecognized structure in the southern hemisphere: the Hydra-Centaurus supercluster. The work of Chincarini and Rood and of Einasto, Jôeveer, Saar and Tago strongly suggests that superclusters extend far beyond the regions covered in our red-shift surveys. By their reckoning the Coma-A1367 and Perseus superclusters may occupy an area of the sky 10 times larger than the area we have cautiously proposed.

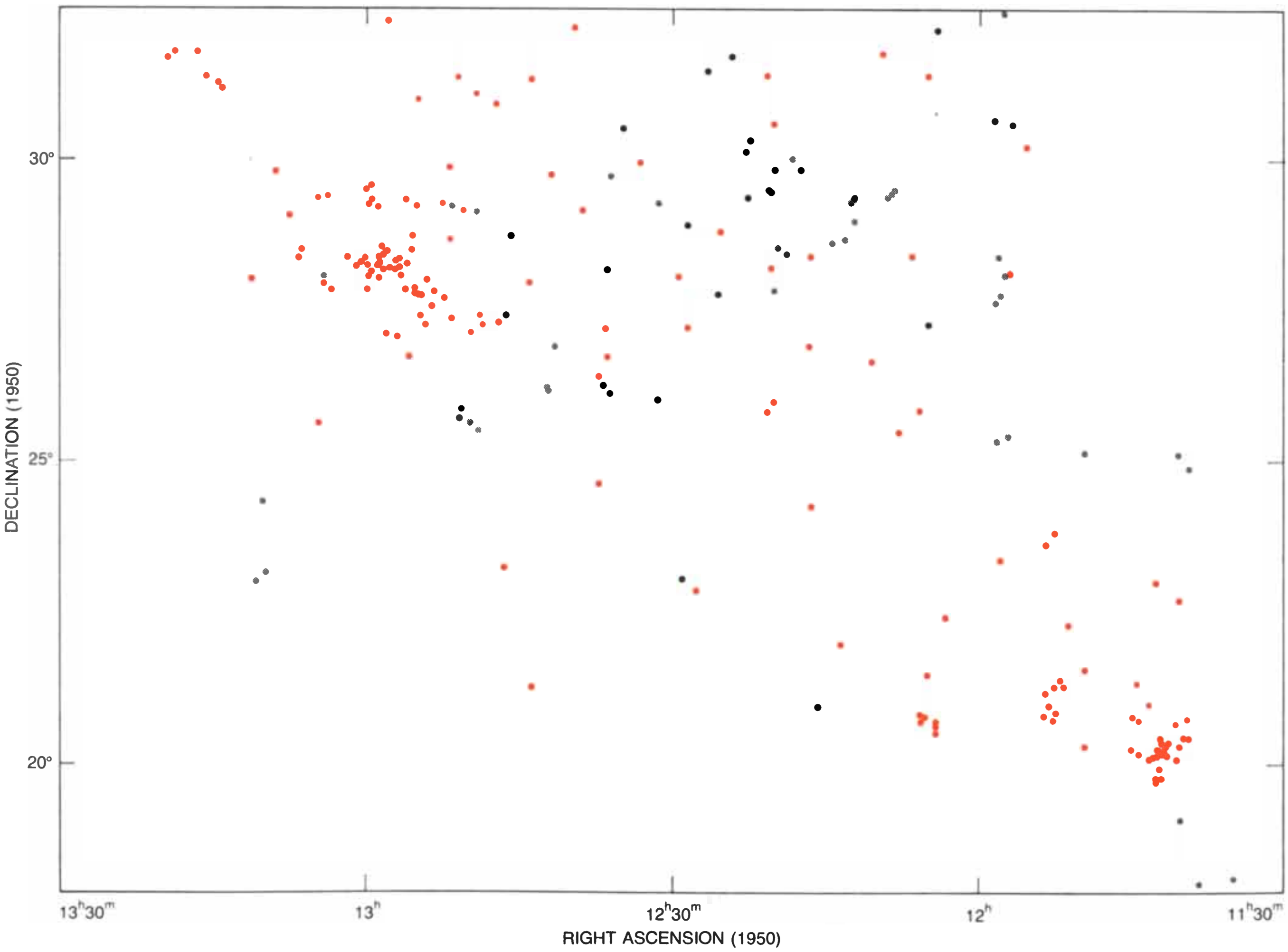
This hypothesis receives added support from a survey by Robert P. Kirshner of the University of Michigan, Augustus Oemler, Jr., of Yale University, Paul L. Schechter of Kitt Peak and Stephen A. Shectman of the Mount Wilson and Las Campañas observatories. Their survey covers three small regions of the northern galactic hemisphere. In each region they find galaxies with red shifts similar to those observed in the Coma-A1367 supercluster. They also believe they have detected an immense void whose volume may approach 30×10^{24} cubic light-years. In their small search areas, centered near the north galactic pole, there seems to be an almost total absence of galaxies with red shifts of between 12,000 and 18,000 kilometers per second. In some four square degrees of sky, where they would have expected to find about 25 galaxies with red shifts in that range, they found only one such galaxy. The size of the void, which would be by far the largest known, is calculated on the assumption that the region between the sampling points is equally empty of galaxies. The void lies at a distance of between 520 and 780 million light-years.

On the basis of the present surveys we have plotted the three-dimensional distribution of galaxies in the three well-defined superclusters: Coma-A1367, Hercules and Perseus [*see illustration on page 114*]. In this representation our own galaxy is at the center. The three surveys provide wedge-shaped windows looking out into the vastness of the cosmos. The tendency of galaxies to clump is seen to be pervasive. The existence of voids, which we were initially hesitant to credit, can no longer be doubted. The universe might have arranged itself so that the space between clusters was filled not by voids but by many small groups of galaxies. Instead the voids are evidently an integral part of the process of clustering and superclustering.

The study of superclusters is not reserved to optical astronomy; radio and X-ray astronomy are also making fundamental contributions. Radio astronomers are able to detect the presence of intergalactic gas by showing that some radio sources in clusters and superclusters have been distorted by what is presumably a gas of low density but high temperature. If gas fills entire superclusters in the same way as it fills the denser cluster regions, its contribution to the total mass of the supercluster could be tremendous.

X-ray astronomers have already detected extremely hot gas in the direction of distant superclusters. It is unclear, however, whether the emission comes exclusively from the cores of bright clusters or whether some of it originates in the regions between cluster cores. Jack O. Burns, Jr., of the University of New Mexico and one of us (Gregory) are combining red-shift data from Kitt Peak, radio maps of distorted sources from the Very Large Array radio telescope at Socorro, N.M., and X-ray images from the Einstein X-ray satellite to examine a complete sample of clusters.

Other astronomers are applying the methods of radio astronomy to carry out their own red-shift surveys. The red shifts are determined by observing the displacement in the 21-centimeter radio-emission line of neutral (un-ionized) hydrogen in interstellar space. One survey, conducted by J. Richard Fisher and R. Brent Tully of the University of Hawaii at Manoa, has mapped the galaxies of the Local Supercluster out as far as the nearest large void. The most sensitive radio telescope available for such studies is the 303-meter antenna at Arecibo in Puerto Rico, where work is being done on all three of the superclusters we have described. The astronomers working on these projects include Chincarini, Thomas M. Bania, Riccardo Giovanelli, Martha P. Haynes, Trinh Xuan Thuan and one of us (Thompson). The observa-



DISTRIBUTION OF GALAXIES in the region of the sky embracing the Coma and A1367 clusters was mapped by the authors in their extensive red-shift survey. The most distant galaxies on the map are shown in color, the nearest are shown in black and those at intermediate distances are shown in gray. The dense concentration of galaxies at a right ascension of 13 hours and a declination of 28 degrees represents the core of the Coma cluster. The smaller aggregation at a right ascension of 11 hours 40 minutes and a declination of 20 degrees represents the A1367 cluster. Because the galaxies elsewhere in the sky seem to be fairly scattered astronomers were misled into thinking most galaxies are sprinkled at random throughout space. A different view of the same region of the sky appears on the opposite page.

tions are valuable because they provide not only additional galaxy red shifts but also data on neutral hydrogen at various locations within the superclusters. Although these studies have not advanced far enough to yield new conclusions on conditions within superclusters, they hold great promise for the future.

It has become abundantly clear from the red-shift surveys that the present-day distribution of galaxies is highly inhomogeneous out to a distance of several hundred million light-years. It seems probable that the inhomogeneity extends out to billions of light-years and characterizes the entire universe. We must add the caveat, however, that the universe may contain much matter that is nonluminous. The possible existence and volume of such matter is currently the subject of wide speculation.

If the universe is inhomogeneous today, there is evidence that at very early epochs it was homogeneous. The evidence lies in the fact that the soft background radiation that bathes the earth at microwave-radio wavelengths is remarkably uniform with respect to direction. The prevailing view is that the background radiation represents the expanded and cooled remnant of the hot early universe. Overall the microwave background radiation is smooth to one part in 3,000. Recently, however, some inhomogeneities of small amplitude but extending over large angles of the sky have been detected.

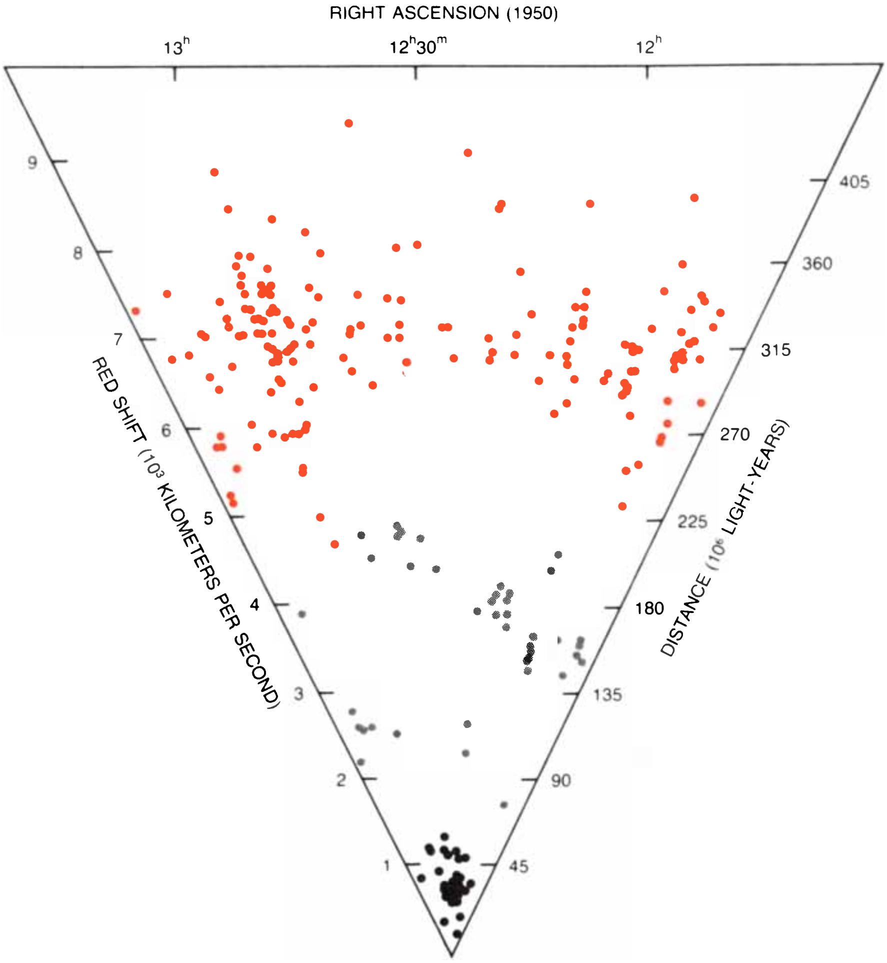
Can the path from homogeneity to the rich assortment of present-day structures be traced? We believe the lack of isolated galaxies and the presence of large voids may provide direct evidence for establishing the relative times of formation of galaxies, clusters of galaxies and clusters of clusters. There are two competing hypotheses. The more conventional model assumes that individual galaxies arose out of a nearly homogeneous primordial soup. The main trouble with this model is explaining how the universe proceeded from its smooth state to the state in which matter was gathered into galaxies. The model assumes that once galaxies formed, small irregularities in their distribution would slowly be amplified by the operation of long-range gravitational forces. The end result of such amplification would be the superclusters seen today.

A competing theoretical explanation was suggested in 1972 by two Russian astronomers, Yakov Zel'dovich and Rashid Sunyaev. In their model the gas of the early universe did not condense into stars and galaxies immediately. Instead slight but very-large-scale irregularities in the general distribution of the gas grew larger in response to gravitational attraction and became increasingly irregular. Eventually the gas became dense enough to collect into vast sheets of material (dubbed "pancakes"), which then fragmented into galaxies. According to this hypothesis, clusters and superclusters form first as concentrations of gas, and only then do galaxies appear.

Do either of these models find support in the observations we have made of superclusters? Since the Zel'dovich-Sunyaev model requires all galaxies to have formed in clusters or superclusters, field galaxies, or random stragglers, should be rare. If the conventional model is correct and galaxies can arise almost anywhere at random, only later to be shepherded by gravity into groups or clusters, stragglers should be rather common. Actually the only populations of isolated galaxies we have discovered in our red-shift surveys are galaxies scattered within the boundaries of superclusters. Moreover, the voids are genuinely empty.

We suggest that the isolated galaxies scattered within superclusters were once members of small groups that were subsequently disrupted by collisions within the dense superclusters. It seems realistic to suppose at one time all galaxies were members of groups or clusters. In sum, the observed distribution of galaxies within superclusters and the existence of huge voids between superclusters are entirely consistent with the Zel'dovich-Sunyaev model. Advocates of the alternative model are searching for support in computer simulations they hope will show how small-scale irregularities can grow into large ones by random processes. How the debate will turn out is not clear.

In describing the filamentlike Perseus-Pisces supercluster we suggested the possibility that the rotation axes of some galaxies were correlated not only with the rotation axes of other galaxies but



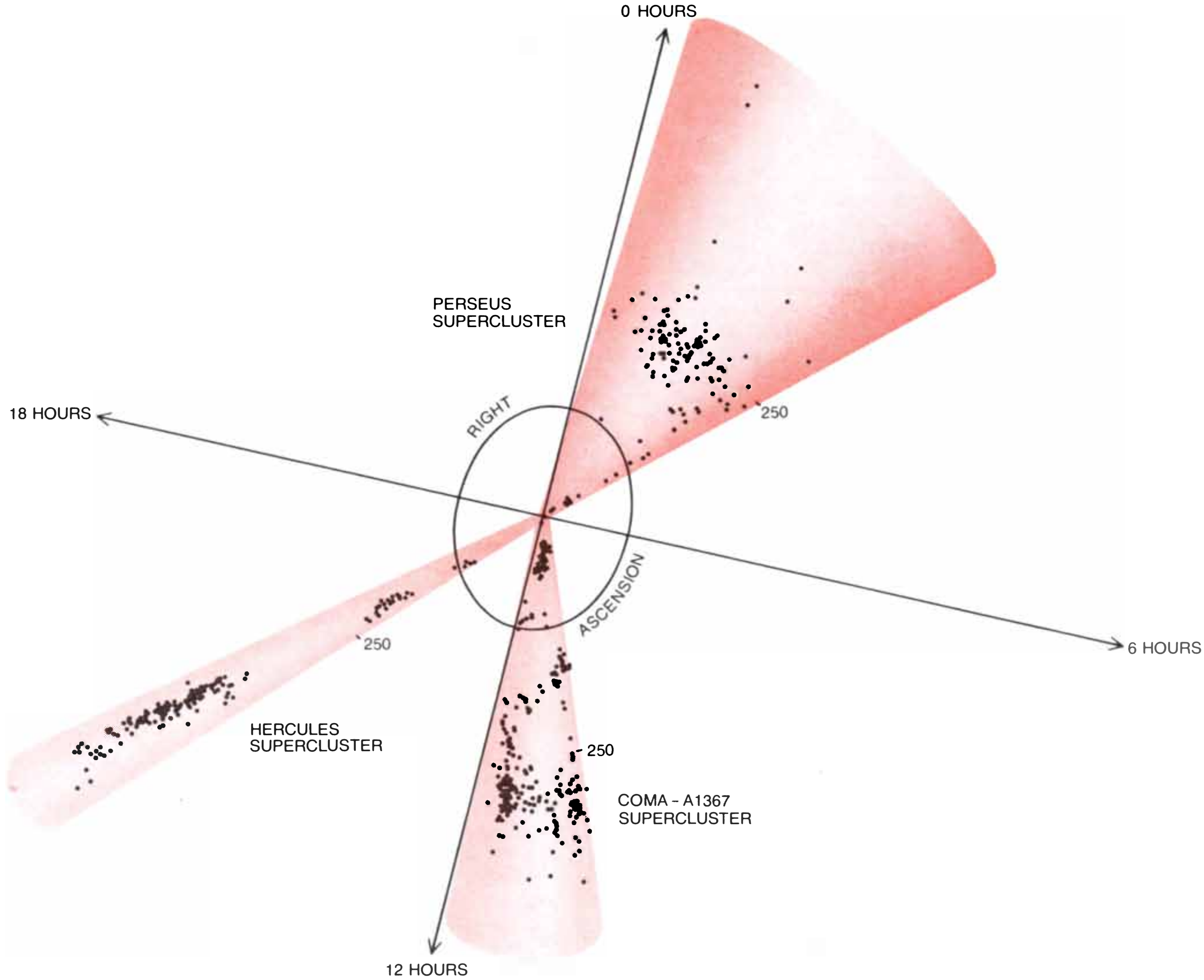
"PLAN VIEW" OF THE COMA-A1367 SUPERCLUSTER can be drawn on the basis of the authors' red-shift survey. The galaxies are the same ones in the illustration on the opposite page but are distributed according to their velocities of recession as deduced from their red shifts. The Hubble constant, which relates recession velocity to distance, provides the scale at the right side of the triangle. The conversion factor adopted here is that a recession velocity of 1,000 kilometers per second corresponds to a distance of 43.5 million light-years (or 75 kilometers per second per million parsecs). Our own galaxy would be at the lower apex of the triangle at a distance of zero light-years. It can now be seen that the Coma cluster (which is at a right ascension of 13 hours with a recession velocity of about 7,000 kilometers per second and the A1367 cluster (which is at 11 hours 40 minutes and the same recession velocity) are merely the richly populated ends of a continuous band of galaxies that stretches across the sky at a distance from our galaxy of about 315 million light-years. This band of galaxies is the Coma-A1367 supercluster. Note also the huge voids where the survey found no galaxies at all.

also perhaps with the gross structure of the supercluster filament. The idea has received support from a recent study by Mark T. Adams, Stephen E. Strom and Karen M. Strom of Kitt Peak, who find similar rotational correlations in the combined data from several flattened clusters. If such correlations are confirmed, supporters of the conventional model of galaxy formation would probably face insurmountable obstacles in producing an explanation. The random statistical processes in the conventional model are not conducive to generating organized rotational motion over any large scale. The Zel'dovich-Sunyaev model, on the other hand, would readily explain such correlations.

What are the prospects of resolving such issues in the near future? One of the most promising avenues of inquiry is continuing improvement in the measurement of the primordial microwave background radiation. The very slight irregularities observed in this radiation are evidence for the existence of structure in the universe from the earliest epochs. The upper limits for the observed smoothness are close to the limits needed for a test of the two competing models of galaxy formation. Perhaps the next 10 years will show that neither model is satisfactory, in which case astrophysicists will have to rethink matters altogether.

Our final comments concern the very concept of a supercluster of galaxies. First, is supercluster the appropriate term? To many of our colleagues the term should be limited to aggregates that are bound together by gravity. It is not clear from the observations that this condition is satisfied. Our own view is that the term supercluster properly describes a present-day aggregate of galaxies that is well separated from any similar structure. No dynamical relations among supercluster members are necessarily implied.

A second comment concerns the universality of superclusters. We have now established that every nearby richly populated cluster in the Abell catalogue is part of a supercluster. We speculate, therefore, that a necessary condition for the formation of a rich cluster is the presence of companion clusters. Finally, we want to leave the reader with a proper sense of the dimensions of superclusters. The Coma-A1367 supercluster is more than 300 million light-years from our galaxy, yet even when it is viewed from such a tremendous distance, it still stretches at least 20 degrees across the sky, through the constellations Coma Berenices and Leo. Chincarini and Rood argue, moreover, that the same supercluster could be as much as 10 times bigger. For astronomers and cosmologists organized structures of such vastness leave plenty of room for study.



THREE DIRECTIONS IN THE SKY have now been intensively studied in galaxy red-shift surveys. In this projection the solar system is at the apex of three cones that embrace the galaxies that have been surveyed in each of the three regions where superclusters of galaxies have now been identified. The scale marker adjacent to each cone gives the distance from the solar system in millions of light-years. The scale allows for foreshortening in the projection. Thus far only about 2 percent of the entire sky has been surveyed in this detail.

The Infrared "Vision" of Snakes

Snakes of two families can detect and localize sources of infrared radiation. Infrared and visible-light information are integrated in the brain to yield a unique wide-spectrum picture of the world

by Eric A. Newman and Peter H. Hartline

Rattlesnakes can strike accurately even at night when darkness hides their prey. The strike is guided by heat: infrared radiation emitted by the warm-blooded target. The radiation is sensed by an extraordinary system that enables rattlesnakes to "see" in a region of the electromagnetic spectrum where animals generate their own radiant energy.

The infrared "eyes" of the rattlesnake are the pit organs, a pair of deep cavities in the head that open on the side of the head below and in front of the eyes. The pit organs are richly supplied with heat-sensitive nerve fibers connected to the brain. All snakes of the subfamily Crotalinae, the pit vipers, have pit organs and are sensitive to infrared radiation. In North America these snakes include the cottonmouth (water moccasin) and the copperhead as well as the rattlesnake. Pythons and many other members of the family Boidae (boid snakes), distant relatives of the pit vipers, also have heat-sensitive pits. In contrast to the pit vipers, they have many pits on the scales bordering their mouth, as many as 13 pairs.

Infrared sensitivity in pit vipers and boid snakes has apparently evolved from the somatic sensory system. In man and other mammals this sensory system is concerned with the sensations of, among other things, touch, pressure, temperature and pain. In mammals sensory messages are conducted from the facial area to the brain by the trigeminal nerve. In infrared-sensitive snakes a substantial part of the trigeminal nerve is devoted to the pit organs and infrared sensitivity. In the course of evolution the somatic sensory system of pit vipers and boids has developed into a sense much like vision. Pit vipers and boid snakes have also evolved specialized nuclei, or clusters of nerve-cell bodies, in the brain to receive and process the sensory information coming from the pit organs. The nuclei relay information to the optic tectum of the midbrain, a structure known for its role in vision and in the spatial representation of sensory information.

Like other animals, human beings learn to associate several sensory modalities with real objects of many kinds. For example, both the sight of a bird and the sound of its song help to localize and identify it. In an analogous way the optic tectum of infrared-sensitive snakes combines infrared signals from the pit organs with visual signals from the eyes. The integration of visual and infrared information in the tectum gives pit vipers and boid snakes a unique view of the world, a view that compares and contrasts visual and infrared images.

Although the infrared pit organs were described long ago by anatomists, their function remained obscure until the 1930's, when G. Kingsley Noble and A. Schmidt of the American Museum of Natural History in New York showed that the pits detect heat. They found that rattlesnakes could differentiate between a warm light bulb (covered with an opaque cloth) and a cold one. Snakes oriented and struck toward warm bulbs as long as their pits remained uncovered. They ignored all bulbs, warm as well as cold, if the pits were blocked.

In the 1950's Theodore H. Bullock and his colleagues at the University of California at Los Angeles demonstrated that the pit organ functions by sensing the infrared radiation given off by warm objects. These investigators monitored the nerve impulses in the axons (long nerve-cell fibers) of the trigeminal nerve that are connected to the pit. They found that impulses are triggered equally well by a warm object whether it is presented to a rattlesnake in the light or in a totally dark room. The axons do not respond, even in a lighted room, if the pit is exposed to the same object once it has cooled. Furthermore, the axon's activity in response to a warm object vanishes if a heat-absorbing glass filter is inserted between the pit and the object, even though the glass lets through nearly all the visible light. In the complementary experiment, where a filter was inserted that let through no visible radiation but passed most of the infrared, the response was only slightly reduced. Thus it was conclusively demonstrated that the pit organ is sensitive to radiation at infrared wavelengths.

Just how good is the rattlesnake's infrared sense? How accurately can a rattlesnake localize a warm object? How sensitive are the infrared detectors? Answers to these questions have come from a combination of behavioral and neurophysiological experiments.

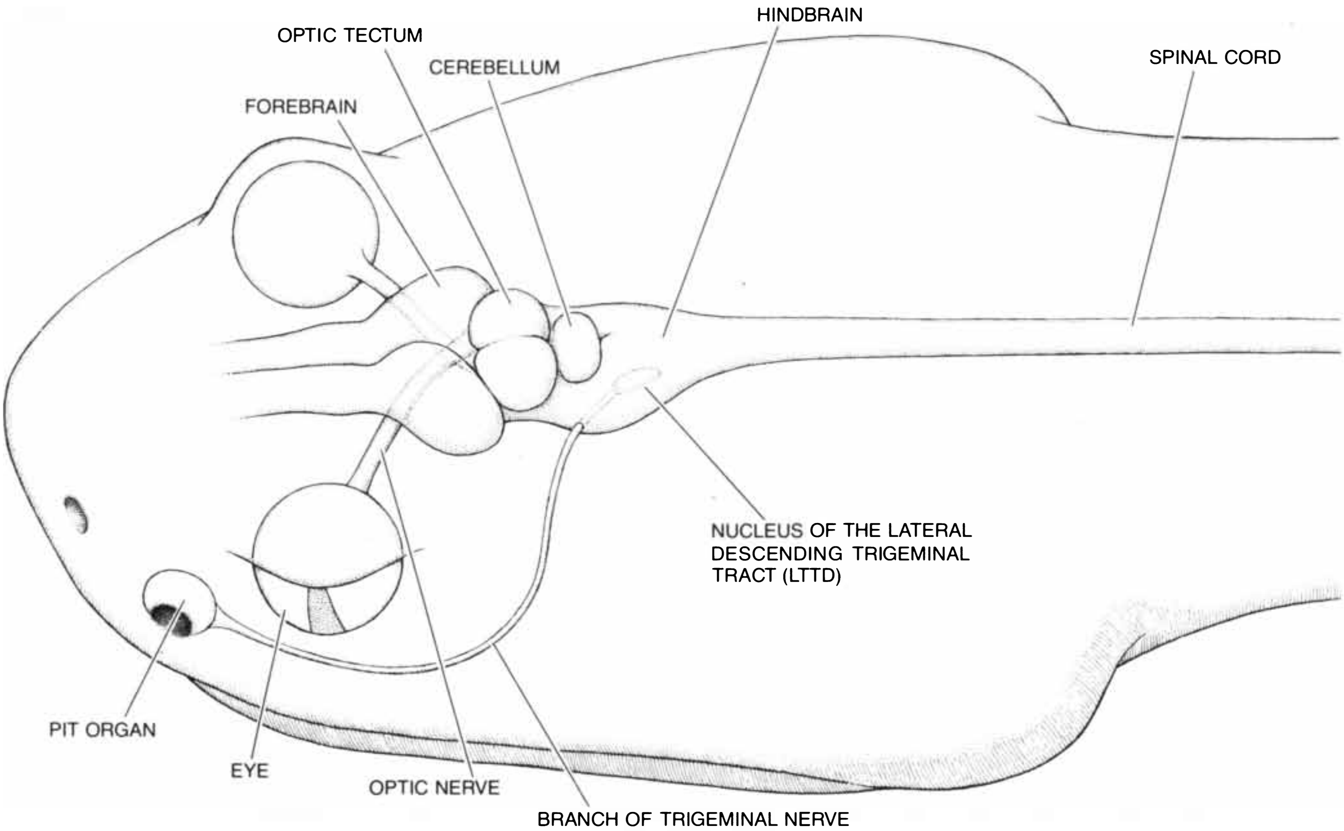
In order to determine how accurately a rattlesnake can localize a warm object Laurence R. Stanford and Michael S. Loop, working with one of us (Hartline) at the University of Illinois at Urbana-Champaign, measured the striking accuracy of blindfolded snakes. A rattlesnake whose eyes (but not its pits) had been occluded with blinders was put on a raised pedestal in the middle of a circular enclosure. An infrared source (the heated tip of a soldering iron) was then presented to the snake just beyond its striking range. An experimenter introduced the soldering iron at various angles from zero to 60 degrees to the left or right of where the snake was facing. A video camera was ready to record the expected strike; the accuracy of the strike would be measured from the videotape.

The snakes refused to cooperate; they almost never struck toward the soldering iron. Apparently the stimulus was not sufficiently meaningful to them. The somewhat diabolical solution was to give the snake a mild electric shock just before each presentation of the soldering iron. The snakes now struck often enough so that we could measure the accuracy of their response. Even with both eyes securely covered the strikes were accurate to within about five degrees of dead center, regardless of whether the target originally lay straight ahead or off to the side. That is very impressive, and for a mouse it is deadly.

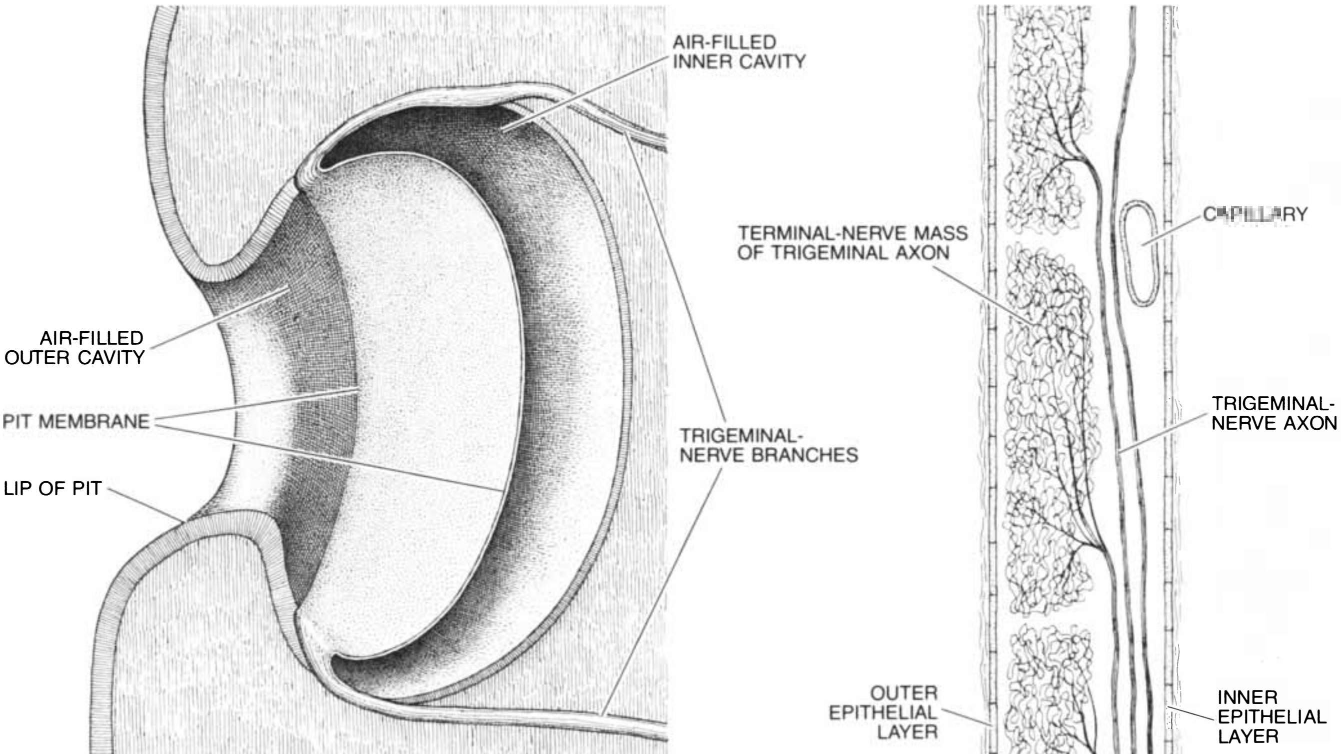
How can an investigator measure the sensitivity of the infrared sense? Bullock's method was to record the frequency of action potentials, or nerve impulses, in the trigeminal-nerve axons leading away from the pit organ. This neurophysiological approach can give an indication of the ultimate sensitivity

SOUTHERN PACIFIC RATTLESNAKE, a native of the western U.S., can locate weak sources of infrared radiation by means of its pit organs, a pair of cavities in front of and under its eyes. In rattlesnakes and other members of the pit-viper family the radiation is detected by an extremely thin heat-sensitive membrane that stretches across the back of each pit. The field of view of each pit extends from about 25 degrees beyond the animal's midline to 105 degrees at the side and from 45 degrees below the horizontal to 60 degrees above it.

RETICULATED PYTHON, a native of southern Asia, has 13 pairs of pits in the scales above and below its mouth. This photograph shows five pits on the left side of the snout and one on the right. Six other pits border the lower lip, below and behind the eye. Other boid snakes (snakes of the family Boidae), including the tree boas, have similar arrays of pits. The back surface of each pit is highly sensitive to heat energy. Each pit has a slightly different field of view. Still other boid snakes, including the boa constrictors, do not have discrete pits but have infrared-sensitive scales on their face. Both of the photographs on this page were made by one of the authors (Newman).



STRUCTURES IN THE BRAIN and the nerve pathways associated with the infrared sensory system of pit vipers have only recently been worked out. A principal puzzle, now solved, was how the information gathered by the pit organ and carried by impulses in the trigeminal nerve finds its way to the optic tectum, a midbrain structure that plays a prominent role in the processing of visual information.



RATTLESNAKE PIT ORGAN is formed out of a cavity in the soft tissues and bone of the face (*left*). A heat-sensitive membrane, 30 square millimeters in area but only 15 micrometers thick, is stretched across the cavity. The membrane intercepts infrared radiation that passes through the aperture of the pit. Inside the pit membrane (*right*) the bushy nerve endings of the trigeminal-nerve fibers terminate within a few micrometers of the membrane's outer surface. A change in the membrane temperature of only .003 degree Celsius is sufficient to raise the firing rate of the trigeminal-nerve fibers. The biophysical mechanism that accounts for the response to heat is not known.

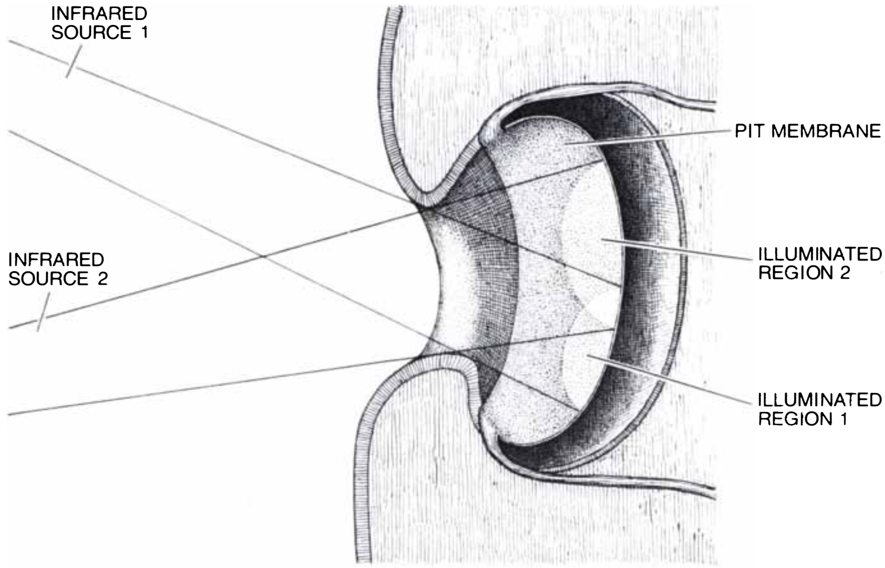
of the pit receptor. Bullock found that in the absence of a stimulus individual axons carry an irregular low-frequency barrage of action potentials. When the pit is suddenly illuminated with moderately strong infrared radiation, the frequency of the firing of the action potentials increases dramatically. The firing rate rises within 100 milliseconds of the onset of the stimulus, remains high for a brief time and then slowly decays back to nearly the background level. When the illumination is turned off, the firing rate temporarily drops below the background level, then slowly recovers to its prestimulus rate.

Bullock and his colleagues found that individual axons are remarkably sensitive to infrared stimuli. A small warm object such as a human hand or a live rat, introduced as far as half a meter from the pit, triggers a noticeable increase in the firing rate of some of the trigeminal-nerve fibers. Such a stimulus corresponds to an increase of the energy incident on the pit organ of approximately .1 milliwatt per square centimeter. For the purpose of comparison, on a clear winter day in the northern U.S. the sun's rays deliver about 50 milliwatts of infrared per square centimeter.

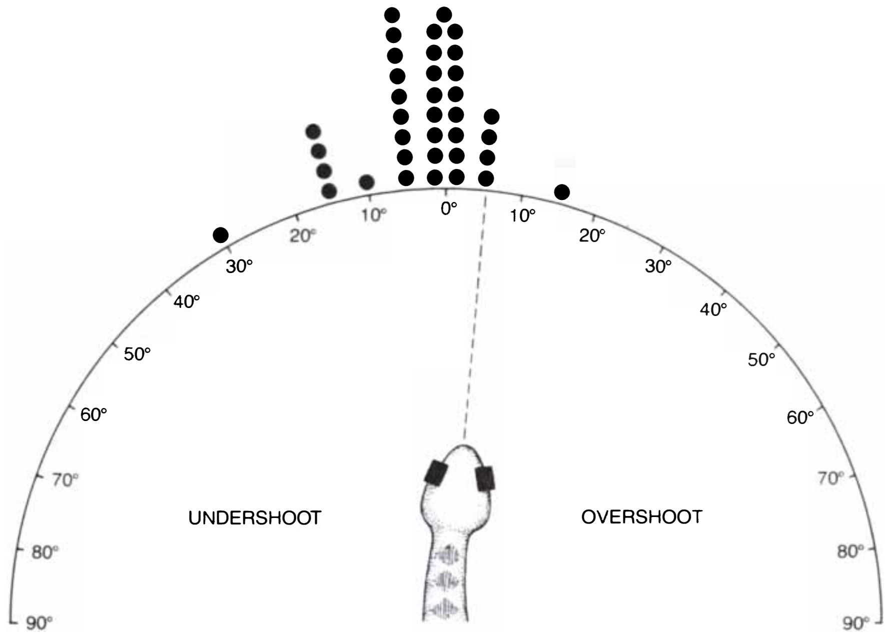
Bullock and F. P. J. Diecke also made direct measurements of the temperature sensitivity of the trigeminal fiber endings in the rattlesnake pit organ. They controlled the internal temperature of the pit by directing a stream of warm water into the pit cavity. A change in the temperature of the flowing water as small as .003 degree Celsius produced a noticeable change in the firing rate of the more sensitive nerve fibers!

The pit organ's remarkable sensitivity to radiant heat can be traced to its remarkable structure. Each pit organ of a rattlesnake or boid snake consists of a cavity from one millimeter to five millimeters in diameter in the soft tissue and bone of the head. In rattlesnakes a thin membrane about 30 square millimeters in area is suspended across the air-filled space of the cavity. Only 15 micrometers thick, or about a fourth the thickness of this page, the membrane is the key component of the sense organ. Distributed across the membrane surface are some 7,000 thermosensitive endings of trigeminal sensory axons. Each of the endings can excite its nerve fiber if it is warmed. Similar heat-sensitive nerve endings cover the bottom of each pit in boid snakes. In boids the sensitive epithelium is not suspended in air as it is in rattlesnakes. The axon endings, however, come to within 30 micrometers of the surface.

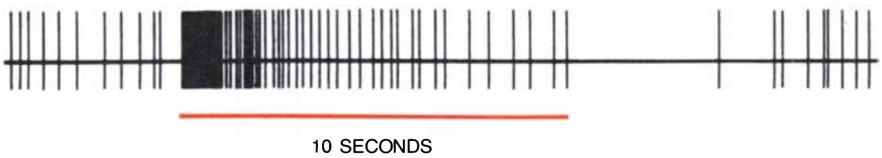
The extreme heat sensitivity of the pit organs is a direct result of the proximity of the thermoreceptors to the outward-facing surface of the sensory epithelium. Infrared radiation falling on the surface immediately raises the tempera-



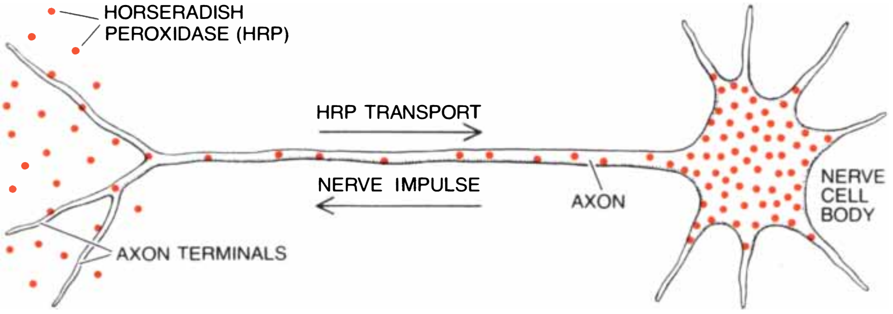
OPTICS OF THE PIT ORGAN resemble a pinhole camera. The diameter of the pit opening in a rattlesnake is about two millimeters, or roughly a third the diameter of the heat-sensitive membrane. Hence the radiation from an infrared source will illuminate only the area of membrane that falls within a cone defined by the location of the source and the pit aperture. Accuracy of the system is limited only by the snake's ability to localize edges of illuminated region.



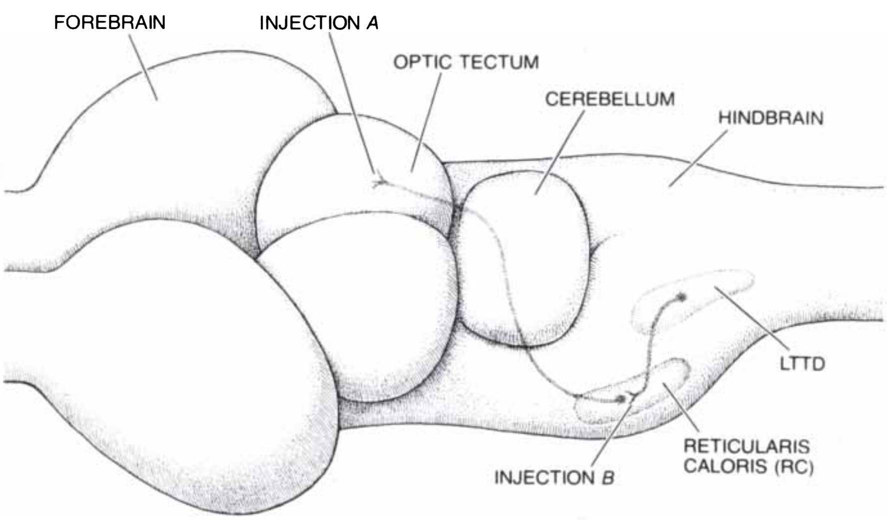
SPATIAL ACCURACY of the rattlesnake's infrared system can be measured by presenting a warm object at various angles to the left or the right of a snake whose eyes are covered with blinders. Each filled circle represents a single strike of the snake and shows the angular error between the strike and the target. The stimuli were presented at up to 60 degrees to the left or the right of the animal. The average error of the strikes is less than five angular degrees.



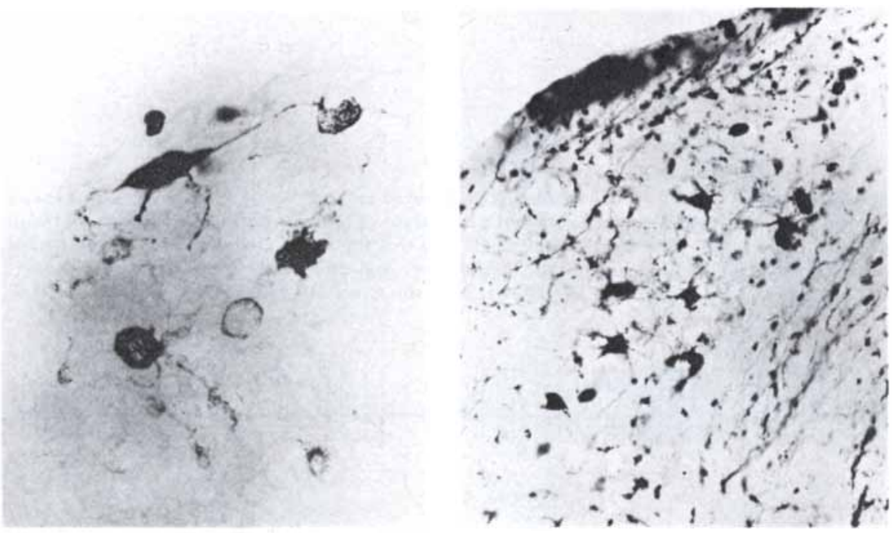
TRIGEMINAL-NERVE FIBERS signal changes in the temperature of the rattlesnake's pit membrane. The colored bar indicates a 10-second test period during which a snake's pit organ was exposed to infrared radiation. The traces on the line above the bar show the impulses recorded from a single trigeminal-nerve fiber. The firing rate is highest when the radiation is turned on but drops to nearly the background level. When the radiation is turned off, there is a short silent period. The properties of infrared trigeminal-nerve fibers were originally discovered by Theodore H. Bullock and his colleagues at the University of California at Los Angeles.



ANATOMICAL PATHWAY OF THE INFRARED SYSTEM was traced with the aid of an enzyme extracted from horseradish, horseradish peroxidase (HRP), which is taken up by the terminals of neurons and transported back to their cell bodies. HRP molecules are colored dots.



HORSERADISH PEROXIDASE INJECTED INTO BRAIN of a rattlesnake was traced by staining thin slices of tissue to reveal the presence of the enzyme. HRP injected into the optic tectum (*A*) was transported to a previously unknown nucleus, or collection of cell bodies, called the nucleus reticularis caloris (RC). When HRP was injected into the RC nucleus (*B*), stained cell bodies were found in the LTTD nucleus, which was known to receive a direct input from the pit organ. LTTD stands for a Latin expression that translates as "nucleus of the lateral descending trigeminal tract." The HRP experiments solved the puzzle of how the infrared information originating in the rattlesnake's pit organ ultimately reaches the optic tectum.



MICROGRAPHS OF BRAIN SLICES show how cell bodies labeled with HRP stand out as dark, irregular profiles after staining. The micrograph at the left is a section of the RC nucleus after horseradish peroxidase was injected into the optic tectum. The micrograph at the right is a section of the LTTD nucleus after the enzyme was injected into the RC nucleus. The experiments were done by the authors and Edward R. Gruberg of the Massachusetts Institute of Technology. Earl Kicliter of the University of Puerto Rico collaborated on first experiment.

ture of the sensors. The same amount of radiation falling on the bare skin of a mammal would raise the temperature of the first few micrometers of the skin by about the same amount. The thermoreceptors of mammals, however, lie much deeper. The entire thickness of the tissue between the surface and the receptors, about 300 micrometers, must be heated in order to warm them, and it takes 20 times as much heat energy to warm a layer of tissue 300 micrometers thick as it does to warm a layer only 15 micrometers thick. The sensitivity of mammalian thermoreceptors is further reduced by conductive heat loss to even deeper layers. Rattlesnakes avoid this problem with yet another specialization: since the heat-sensing membrane is surrounded by air on both sides, it does not lose heat to deeper tissues. Energy absorbed from infrared radiation almost exclusively heats the receptors.

Although the temperature sensitivity of rattlesnake trigeminal-nerve axons is impressive, it is not extraordinary when it is compared with the thermosensitive nerve cells of other animals. The skin of the human forehead, for example, has about the same threshold sensitivity to temperature. The great difference between rattlesnakes and other animals in sensitivity to incident radiation is primarily due to the anatomical specializations of the sensory organ and not to physiological differences in the sensory receptors.

The directional capabilities of the infrared sense can also be attributed to anatomical specializations of the pit. The pit organs of rattlesnakes have openings that are less than half the diameter of the sensory surface. Thus a small warm object half a meter or so in front of the pit will illuminate no more than a fourth of the entire thermosensitive membrane. The location of the illuminated patch on the membrane surface should therefore give an adequate clue to the location of a warm object. In short, the pit organ functions much like a crude pinhole camera.

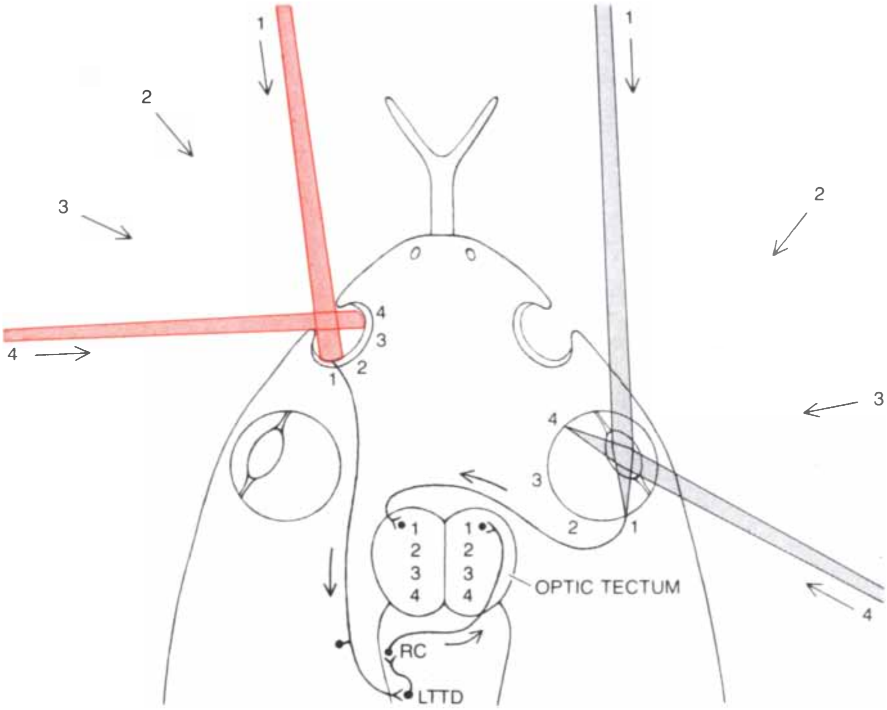
The accuracy with which the snake can localize a warm object must depend on how accurately it can determine the position of the illuminated patch on the membrane. Simple trigonometric calculations show that in order to localize a small source to within five degrees of angle the snake must be able to measure the position of the warm spot on its receptive membrane to within about 175 micrometers. With a mosaic of receptors, each one about 60 micrometers across, such accuracy is quite reasonable. Diffraction of the infrared radiation (whose dominant wavelengths are in the band between six and 12 micrometers) may blur the edges of the warm spot somewhat, but not enough to reduce the pit's accuracy significantly.

Specializations evolved by the infrared-sensitive snakes are not limited to the pit organ. Pit vipers and boid snakes have also developed unique brain structures to process the novel information gathered by the pits. Early neurophysiological experiments on the rattlesnake brain demonstrated that the activity of many neurons (nerve cells) in the optic tectum of the midbrain (which was then thought to be primarily concerned with vision) was controlled by infrared stimuli. Evidently the infrared system, in parallel with its evolution from a facial-skin sense to a visionlike remote sense, has achieved a prominent representation in the tectum. This discovery, made independently by one of us (Hartline), who was then working at the University of California at San Diego, and by S.-I. Terashima and R. C. Goris of the Tokyo Medical and Dental University, presented an interesting anatomical problem. In other reptiles and in mammals the trigeminal nerve terminates in the trigeminal region of the hindbrain. How does infrared information originating in the trigeminal nerve find its way to the optic tectum of the midbrain in the rattlesnake?

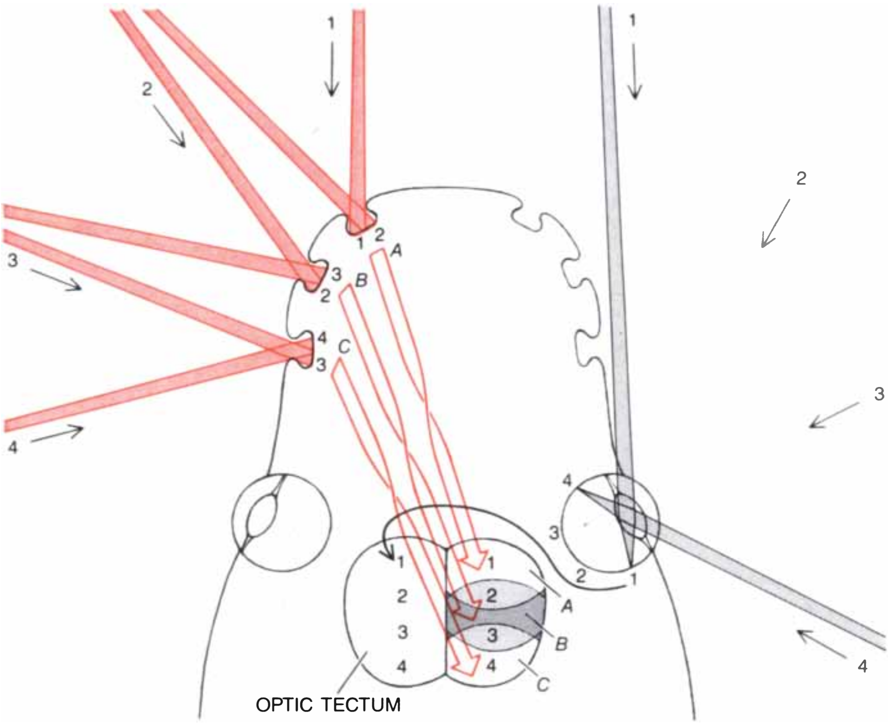
Two groups of investigators began to work on the problem. Dolores M. Schroeder and Loop, working at the University of Virginia Medical School, employed cobalt ions to trace the connections of the rattlesnake trigeminal nerve. They placed the cut end of a branch of the nerve serving the pit in a cobalt chloride solution and passed an electric current between the solution and the brain. The current drove cobalt ions up the axons of the nerve to the axon terminals. When Schroeder and Loop treated sections of brain with a cobalt stain, they found that the trigeminal-nerve fibers, instead of ending in the normal trigeminal hindbrain area, ended in a new nucleus, now known as the LTTD (an abbreviation for the new structure, originally named in Latin "nucleus of the lateral descending trigeminal tract").

G. J. Molenaar of the University of Leiden independently identified what is almost certainly the same nucleus in the python, a boid snake. The LTTD nucleus appears to be peculiar to snakes that have a specialized infrared capability; it has been found in no other animal. It seems to be exclusively devoted to the infrared sensory system.

Our curiosity was piqued. If the nerve impulses carried by the trigeminal axons are delivered to the LTTD nucleus, what is the pathway that relays the infrared sensory information to the optic tectum? One approach to tracing connections within the brain is to use the peroxidase enzyme extracted from horseradish. The large horseradish-peroxidase molecule is taken up specifically by the axon terminals of individual neurons and is transported to other parts



INFRARED SPACE AND VISUAL SPACE are represented in similar orientations on the surface of a rattlesnake's optic tectum. For clarity the pathways of the two sensory systems are depicted for opposite sides of the head. The front area of the tectum (*1*) receives input from the back areas of the pit membrane and from the retina, which survey roughly the same regions of space. The back area of the optic tectum (*4*) receives input from the front areas of the pit and the retina, which both "look" toward the side and behind the animal. Although the representations of the infrared and visual fields of view on the surface of the optic tectum have some systematic differences, they are similar enough so that each region of the optic tectum receives information through both sensory modalities from the same general region of space.



CONNECTIVITY IN THE PYTHON INFRARED SYSTEM is more complex than it is in the rattlesnake. The python pit organs look out on different but overlapping areas of infrared space. (Only three pits of the multipit array are shown.) The field of view of each pit is displayed over the pit's projection region on the tectal surface. Anterior pits (*A*) are represented in the front of the tectum, the more posterior pits (*C*) toward the rear of the tectum. The front area of each pit's infrared-sensitive epithelium is connected to the back of its projection region. The complex pattern of connections yields a more or less continuous representation of infrared space on the python's tectal surface that corresponds fairly well with the visual representation.

of the same cells. A staining procedure makes the horseradish peroxidase clearly visible in slices of brain and reveals the locations of the cell bodies of the neurons that took up the peroxidase.

Working with Edward R. Gruberg of the Massachusetts Institute of Technology and Earl Kicliter of the University of Puerto Rico, we injected horseradish peroxidase deep into the rattlesnake optic tectum where infrared sensory input is found. If cells in the LTTD projected directly to the optic tectum, we reasoned, their axon terminals in the tectum would pick up the peroxidase and transport it back to the LTTD. Somewhat to our surprise we found no stained neurons in the LTTD. Instead we found a group of prominently stained cells in a different region of the hindbrain. We postulated that this group of cells, which we named the nucleus reticularis caloris (RC), served as a relay station between the LTTD and the tectum.

Now we had to determine whether or not our new nucleus was connected directly to the LTTD. We investigated this question with Gruberg by injecting horseradish peroxidase into the RC nucleus. This time we did find stained cells in the LTTD. Stanford and Schroeder looked into the same question with a technique based on the degeneration of nerve cells. They selectively destroyed the cells of the rattlesnake LTTD nucleus and used a staining procedure that revealed the degenerated axons of these cells. They were able to trace the axons from the LTTD to the RC. Therefore both experimental approaches yielded the same answer: cells of the LTTD nucleus project directly to the RC. Although neither of these experiments could clarify the functional roles the LTTD nucleus and the RC nucleus play in the infrared sensory system, they at least explained how infrared information reaches the optic tectum. In the rattlesnake the LTTD and RC nuclei have evolved to process infrared sensory information and relay it to the tectum.

It is clear from the effectiveness of the infrared system in guiding the snake's strike that the system is able to localize objects in space with high accuracy. In this role it functions remarkably like the visual system of other animals. Our neurophysiological studies and those of our colleagues have revealed many similarities between the mechanisms that enable the visual and infrared systems to localize objects.

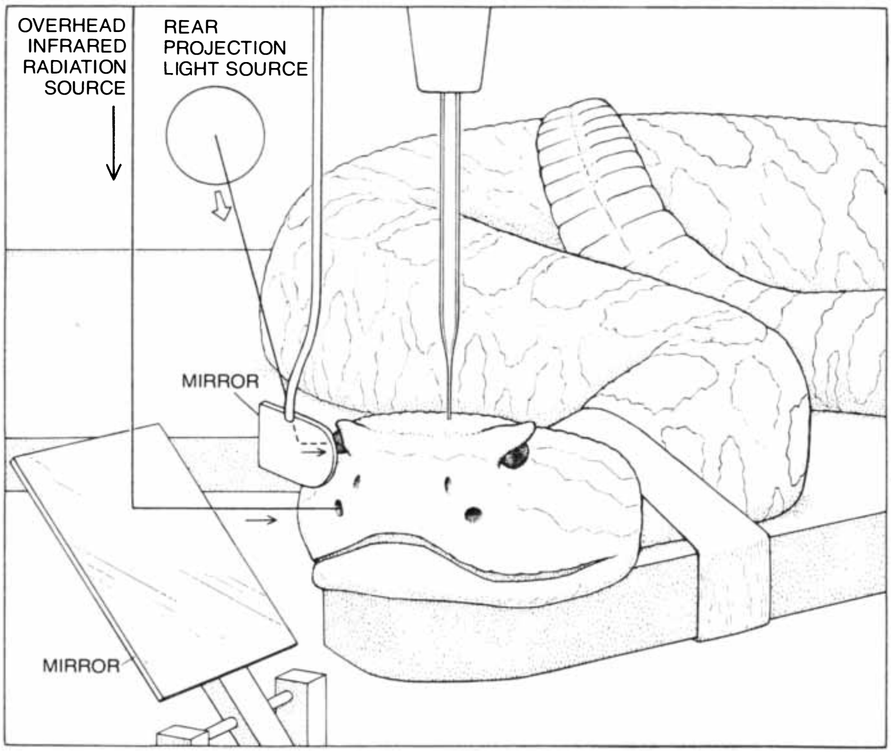
Bullock, and later J. A. DeSalvo, working with one of us (Hartline), described how individual trigeminal-nerve axons respond to warm stimuli anywhere within a field of view described by a cone originating at the pit. Such cones of view are usually between 45 and 60 degrees across and are called, by analogy with their visual counterparts in the retina, excitatory receptive fields. Most trigeminal fibers therefore behave as if their receptive fields were limited simply by the shadows cast by the lip of the pit. The response of any one trigeminal neuron would not convey the location of a stimulus accurately, but a population of many neurons, whose activity could be analyzed by the brain, might do so. In such analysis an important role is probably played by the optic tectum.

The optic tectum, also known in mammals as the superior colliculus, is one of the major visual centers of the brain. It is perhaps the main center for the processing of information about the location of objects in space. In monkeys the tectum helps to initiate and control "looking," or direction of gaze. In frogs and toads it probably initiates and controls the capture of flies and other prey.

In the optic tectum visual information is organized in an orderly way. Connections between the retina and the tectum (which are linked by the optic nerve) are made in such a way that a map of visual space is formed on the tectal surface. Just as a road map represents the spatial organization of roads on the earth's surface, so the visual map on the tectal surface represents space as it is viewed by the eye. For example, the front part of the tectum represents the part of visual space in front of the animal; the back part of the tectum represents the space toward the side and the rear. The orderly representation of visual space on the tectal surface is a constant feature of vertebrates, from fishes through primates.

Vision is not the only sense that can tell the brain where objects are. Sound also yields valuable location cues. So does touch. Both of these senses are also represented in the optic tectum of mammals. Furthermore, both the auditory system and the somatic sensory system are represented in the tectum in an orderly spatial way that coincides roughly with the visual system's tectal map. The front part of the tectum represents the front region of auditory space; sounds coming from in front of an animal preferentially excite this tectal region. Similarly, the front part of the tectum responds preferentially to touch stimulation of the front part of an animal such as the snout.

Is the infrared system organized in a similar spatial manner in the rattlesnake optic tectum? Leonard J. Kass, Loop and one of us (Hartline) investigated this question at the University of Illinois with the aid of neurophysiological techniques. We recorded the electrical responses of tectal cells from a grid



RATTLESNAKE'S INFRARED AND VISUAL SENSES were stimulated with special equipment in the authors' laboratory at the Eye Research Institute of the Retina Foundation in Boston, Mass. The microelectrode above the snake's head recorded the activity of single tectal cells. Infrared radiation from a source directly above the snake was reflected off the large tiltable mirror and into the right pit organ. Visual images from a rear-projection screen were directed into the right eye by a smaller mirror. The visual stimulus was a bright circular spot.

of positions on the tectal surface and determined where in space an infrared stimulus elicited a maximal neuronal response.

After analyzing the data from several snakes we were gratified to discover not only that the infrared sense is organized spatially in the tectum but also that the infrared tectal map follows the same rules as the visual map. The front part of the tectum responds to infrared stimuli coming from in front of the snake; the back part, to stimuli coming from the side. In short, the visual and the infrared maps are, at least to a first approximation, in register. As we shall see, the similar spatial organizations of the visual and the infrared systems offer advantages in the processing of signals from the two sensory modalities.

The orientation of the map of infrared space on the tectum represents a major departure from the general vertebrate pattern of connections in the somatic sensory system, even though the infrared system presumably evolved from a somatic sense. In most other animals the front part of the tectum represents the front of the body surface. In the rattlesnake, however, the front part of the tectum represents the part of the pit membrane farthest from the animal's snout rather than the part nearest it. This inverted connectivity preserves in the tectum the normal representation of space but not of the body surface.

Boid snakes present an even more complex example of connectivity between the skin and the tectum. The reticulated python, for example, has 13 pits on each side of its head. Each pit is in essence an inverting optical system, which calls for an inverted tectal map, like the map in the rattlesnake. The pits toward the front of the python's head look farther forward, however, than the pits farther back along its lip. One would expect the front pits to be connected to the front part of the tectum, conforming to the usual noninverting vertebrate scheme. How, then, can the array of pits be connected to the tectum to form a reasonable representation of infrared space there? E. C. Haseltine of Indiana University tackled this intriguing problem in collaboration with one of us (Hartline).

What we found is the most complex somatic connectivity scheme observed in any animal to date. First we located the region of the tectum that receives connections from the entire sensory surface of a single pit, that is, the pit's projection region. When we mapped the pit's connections within that region, we found that the back part of each pit's sensory surface mapped to the front of its projection region in the tectum. The map of each pit is therefore inverted compared with the somatic maps of typical vertebrates. When we compared the projection regions of the different pits, however, we found that the usual vertebrate mapping scheme holds true: the front pits project to the front of the tectum and the back pits project to the back. There is a final bit of complexity: adjacent pits, whose fields of infrared view overlap substantially, have substantially overlapping projection regions in the tectum.

The end result is a single, coherent representation of infrared space on the python tectal surface. Most important, the visual and the infrared maps in the python tectum correspond fairly well to each other. Evidently an intricate rearrangement of the connections between the somatic sensory system and the tectum has evolved to preserve the spatial registration between the two sensory modalities.

An intriguing possibility is suggested by the correspondence between the visual and the infrared representations of the world in the rattlesnake and python optic tectum. Perhaps the tectum compares or combines signals from the two senses. We knew that some tectal neurons in mammals are responsive to both visual and auditory stimuli or to both visual and tactile stimuli. We conjectured that similar "multimodal" neurons may exist in the rattlesnake tectum.

We were not disappointed. The first bimodal neurons in the tectum of a snake were identified by one of us (Hartline), Kass and Loop. The infrared stimuli were supplied by a hand or a warm soldering iron, the visual stimuli by a cardboard bar (which was thermally neutral). We found that some tectal neurons respond to warm objects whether the room is lighted or totally dark. The same neurons respond to a cardboard bar, but only if the room lights are on. We called these cells "or" neurons, since they can be excited by stimulation of either of the two modalities.

We also found a small number of tectal neurons that respond to warm objects, but only if the room is lighted. These neurons do not respond to a thermally neutral cardboard bar even when the room is brightly lighted. Nor do they respond to a warm object in a darkened room. They were designated "and" neurons, since both visual and infrared stimuli seem to be essential for their response. Neurons such as these had never been described for any combination of sensory modalities in other animals.

The two of us (Newman and Hartline) realized that we might find other kinds of interactions of the visual and the infrared modalities if we tested specifically for them. We built a computer-controlled stimulator that could present infrared and visual stimuli whose intensity, location and time courses were independent of one another. Like some illusions of magic, this one was achieved with a few well-placed mirrors [*see illustration on page 122*].

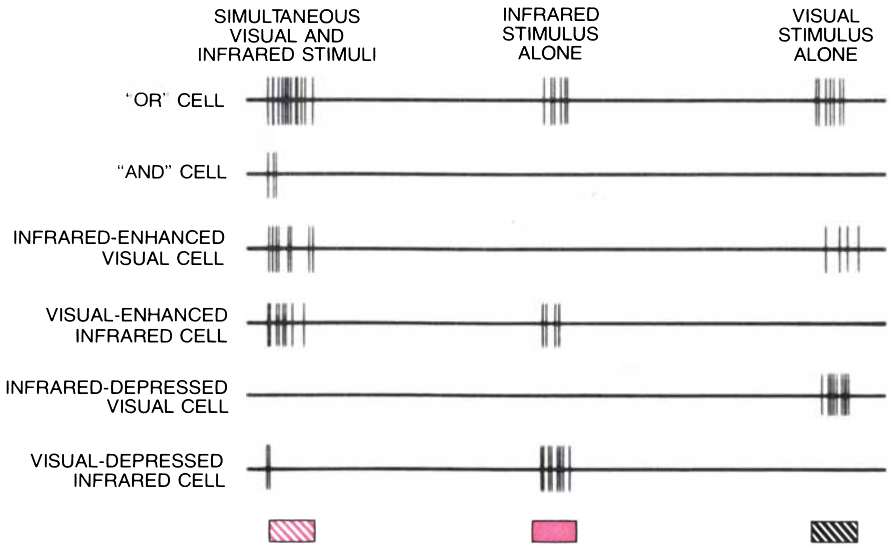
In a typical experiment we lowered a recording microelectrode into the optic tectum of a rattlesnake and advanced it until we could record the electrical responses of a single neuron. Our computer then generated a sequence of infrared, visual and combined infrared-visual stimuli and recorded the responses from the cell. We then advanced the microelectrode until another cell was encountered. In this way we sampled and tested several scores of neurons.

Our hunch about possible exotic modality interactions turned out to be well-founded. We discovered a number of novel infrared-visual interactions. The responses of the tectal cells, and therefore the cells themselves, fall naturally into six classes, two of which are represented by the "or" and "and" neurons. We confirmed that "or" cells respond to both infrared and visual stimuli presented alone as well as to combined infrared-visual stimuli. "And" cells respond only to simultaneous infrared-visual stimulation.

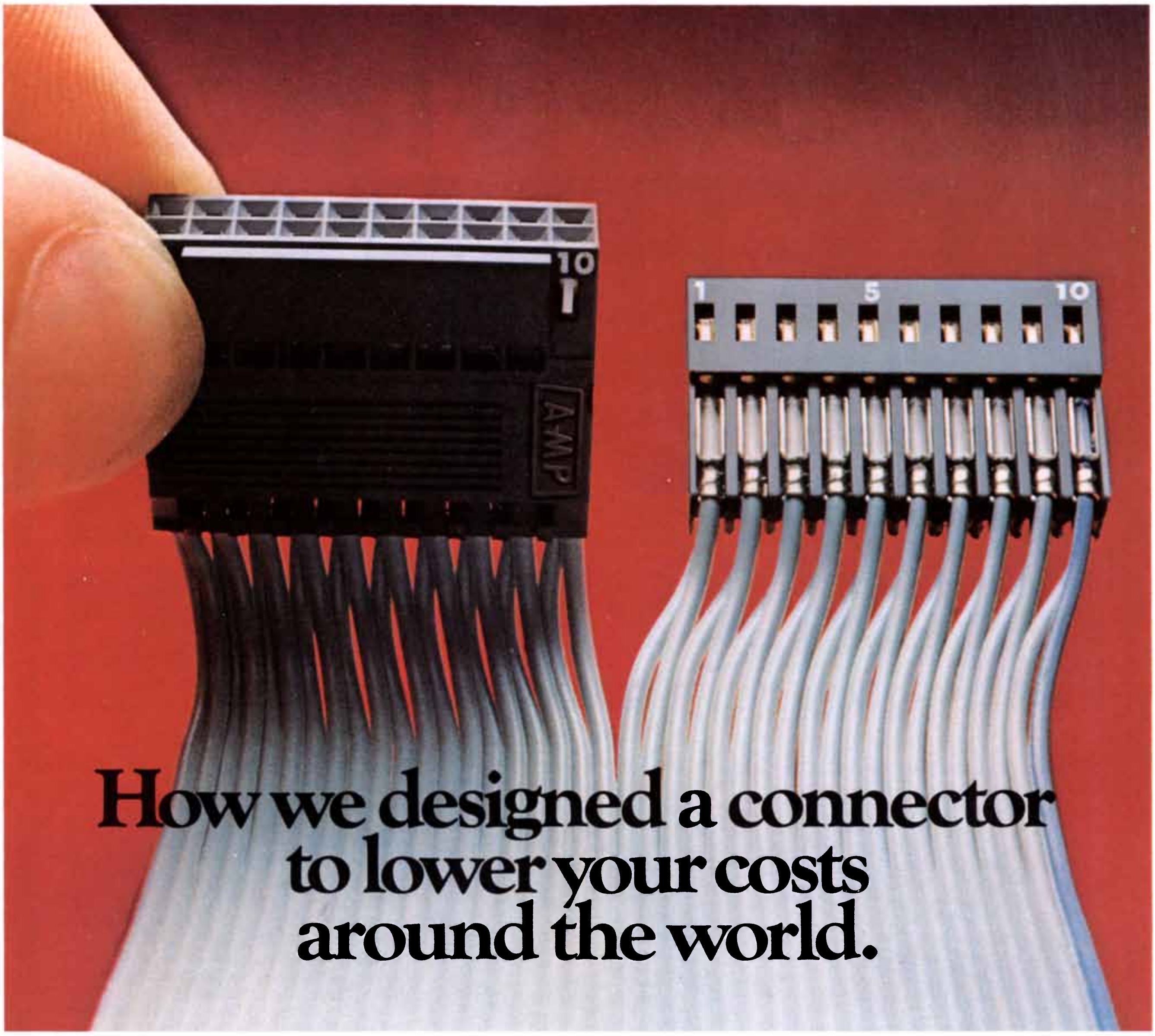
Tectal neurons of two other classes exhibit "enhancing" interactions of visual and infrared modalities. We describe such cells as "visual-enhanced infrared" cells or as "infrared-enhanced visual" cells, depending on the stimulus that does the enhancing. For example, a "visual-enhanced infrared" cell will show a moderate response (say four nerve impulses) to an infrared stimulus presented alone and will show no response to a visual stimulus. When infrared and visual stimuli are presented together, however, a larger response (eight impulses) is evoked. The excitatory influence of the visual stimulus, although the stimulus is not strong enough to generate a response by itself, can significantly enhance the cell's response to the infrared stimulus. Similarly, an "infrared-enhanced visual" cell will show no response to an infrared stimulus but will show an enhanced response to a visual stimulus when it is accompanied by an infrared stimulus.

Tectal neurons of the remaining two classes display inhibitory interactions of the infrared and the visual modalities. The neurons of one class, which we call "infrared-depressed visual" cells, respond vigorously to a visual stimulus presented alone but fail to respond or responded weakly to the combined infrared-visual stimulus. The inhibitory influence of the infrared stimulus reduces or completely suppresses the response to the visual one. On the other hand, with cells of this class the removal of a warm object or the introduction of a cold one will enhance visual responses. The neurons of the other class, in which an infrared response is inhibited by a visual stimulus, we call "visual-depressed infrared" cells.

What functional role might be played by the infrared-visual neurons we have identified? Let us try to answer the question by considering how the tectum might function when a rattlesnake is searching out its prey. A simplified description of such behavior might divide it into several processes: the detection and localization of a stimulus, the identification of the stimulus and the triggering of the appropriate orientation movement. An "attentional" process may also be involved in which the animal's attention is focused preferentially on one region of space; a stimulus in that



SIX CLASSES OF TECTAL CELLS that respond to infrared and visual stimuli have been identified. The rectangles on the bottom line show the time courses of combined infrared and visual stimuli (*color with white hatching*), infrared stimuli alone (*color*) and visual stimuli alone (*white hatching on black*). Tectal cells classed as "or," "and" and "enhanced" exhibit excitatory interactions of infrared and visual inputs. Two "depressed" classes of cells respond to only one of the two stimuli and show inhibitory interactions when both inputs are presented together.



If you're a progressive company, you need the flexibility to keep costs down yet maintain the same performance— anywhere in the world.

That's why if you use electrical connectors, it pays to specify AMP.

An example is our modular interconnection.

Designed for use with printed circuit boards, its building block simplicity has made it universally acceptable. Now it is particularly cost effective because of mass termination technology—making up to 50 connections in the time it took to make one.

As a result, this connector is helping to increase productivity for customers around the world.

Can we help you?

Whether you're looking for connectors or better ways to apply them, call us. We'll give you what the world expects from the industry's leader in specialized engineering. Products designed for performance. Products which help reduce the cost of making yours.

AMP has a better way.

For ways to save you money, for products and innovations that help improve productivity, we have the better ways.

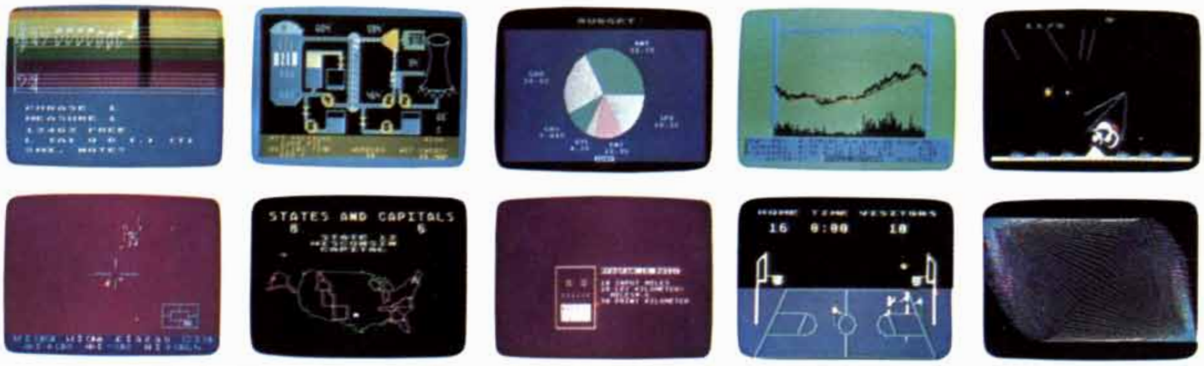
Write for our brochure, "Why AMP Means Productivity." AMP Incorporated, Harrisburg, PA 17105.

Auckland • Barcelona • Brussels • Buenos Aires • Dublin • Frankfurt • Harrisburg • Helsinki • s-Hertogenbosch • Kowloon • London • Luzern • Mexico City Montreal • Oslo • Paris • San Juan • Sao Paulo • Singapore • Stockholm • Sydney • Taipei • Turin • Toronto • Tokyo • Valley Forge • Viby • Vienna

AMP means productivity.

THE GRAPHIC DIFFERENCE

BETWEEN ATARI® COMPUTERS AND ALL OTHERS.



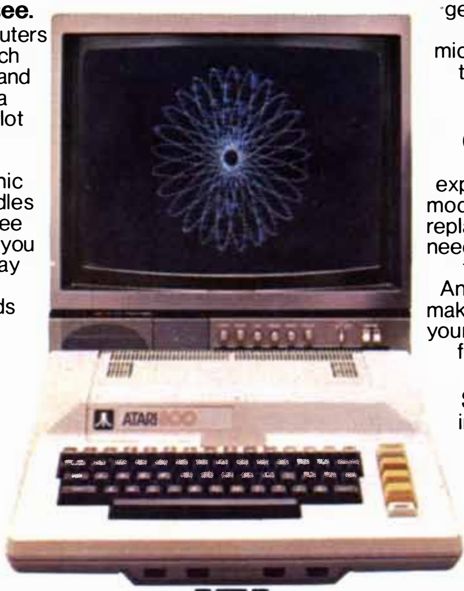
3.7 million reasons why the ATARI Home Computer is something to see.
The display screen used with our computers is composed of 192 horizontal lines, each containing 320 dots. Delivering color and luminosity instructions to each dot for a second requires 3.7 million cycles...a lot of work for the normal 6502 processor.

That's why the ATARI computer has equipped its 6502 with its own electronic assistant. It's called ANTIC, and it handles all the display work, leaving the 6502 free to handle the rest. What this means to you is uncompromisingly spectacular display capabilities without loss of computer power needed to carry out the demands of your program.

That's a quality you just don't find in ordinary personal computers. And it's one of the reasons some computer experts say that ATARI computers are so far ahead of their time.

There's more...which is what you'd expect from ATARI.

Language. The ATARI Personal Computer uses several programming languages to give the user maximum control of its extraordinary capabilities. PILOT, Microsoft BASIC,* and ATARI BASIC are understood and spoken by the ATARI computer. You'll also find our Assembler Editor cartridge indispensable for machine language programming.



Sound. An ATARI computer has four sound generators, or voices, activated by a separate microchip. This leaves the principal microprocessor chips free to perform other tasks. And you can take full advantage of this capability which is designed for easy programming.

Change. ATARI Home Computers have been designed to make change and expansion easy. The ATARI computer has a modular operating system* that can be easily replaced as new technology develops. If you need it, memory expansion requires no more than inserting additional RAM modules.* And the ATARI ROM cartridge system also makes it easy to change languages. In short, your ATARI computer won't be obsoleted by future developments...*beause it already incorporates the future.*

Sharing. To learn more about the amazing capabilities of ATARI computers, visit your local computer store for a demonstration. Or send for our Technical User's Notes, intended for the serious programmer. They are only $27 and contain a lot more information about our computers' special capabilities than most companies could tell. See your ATARI dealer, or send $30 ($27 plus $3 postage and handling), payable to ATARI, to Technical User's Notes, c/o ATARI Customer Service, 1340 Bordeaux Avenue, Sunnyvale, CA 94086.

ATARI®

*ATARI 800™ computer only.

region would trigger a movement of the head or the eyes but an equivalent stimulus in another region would not.

"Or" cells and "enhanced" cells would be useful in the detection process. "Or" cells would signal events in a particular region of space regardless of whether the events are perceived by way of the visual or infrared modalities or by way of both modalities at once. "Or" and enhanced cells respond more strongly to combined stimulation than to stimulation evoked by either visible or infrared radiation alone. Detection and localization would be improved by such excitatory interactions, particularly when neither modality is strongly excited. This might be the case if in the dim light of dusk a mouse was near the maximum detectable distance for the pit organ. Modality summation would occur at the correct tectal location because of the similar spatial organizations of the infrared and the visual modalities.

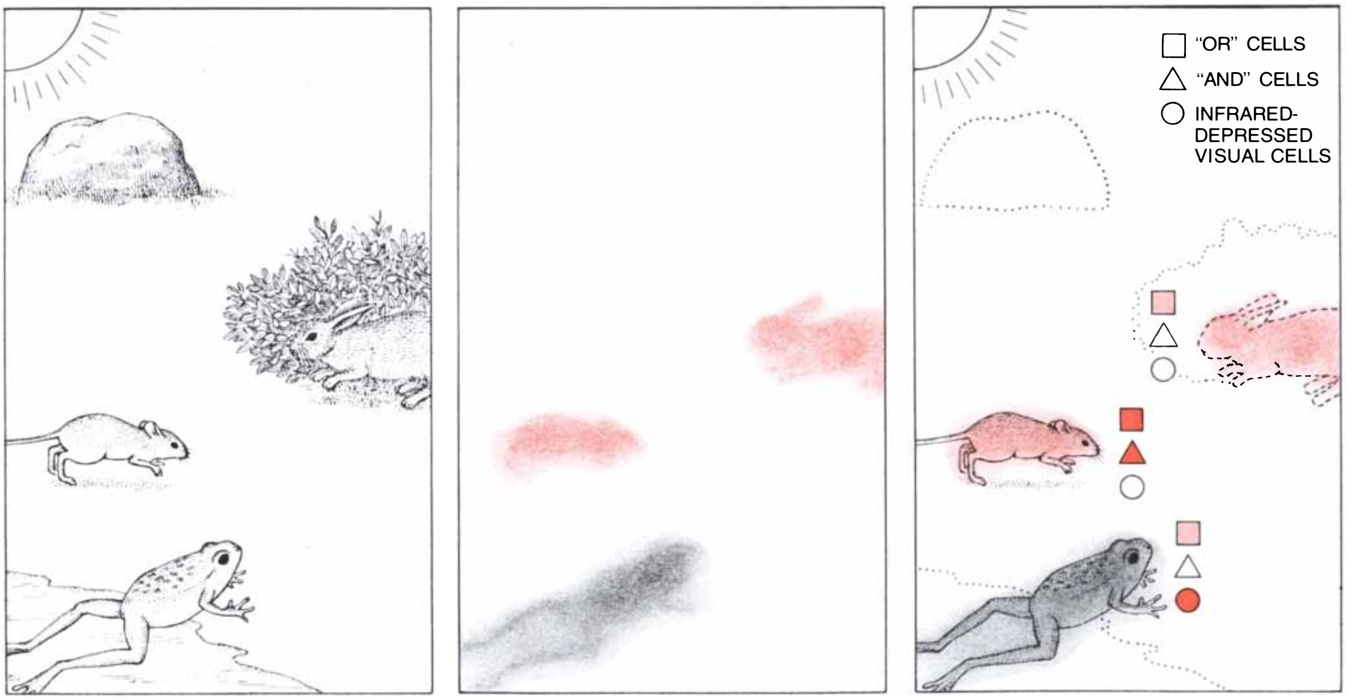
"Or" neurons and "enhanced" neurons might also be involved in an attentional process, one in which a warm, visible object to one side of a snake might catch the "attention" of the orientation machinery of the brain. The infrared component of the stimulus would generate an elevated excitatory state among "or" and enhanced units in a large tectal region because infrared receptive fields are fairly large. Such a heightened excitation could prime that region of the tectum for visual stimulation by the same object or a nearby one. The infrared stimulation would thereby serve to draw attention to the corresponding part of visual space.

Infrared-visual tectal interactions might also contribute to the identification process by stimulating "feature-detecting" neurons that respond only to specific combinations of infrared and visual stimuli. "And" cells, for example, are activated only by objects that can be detected by the pit organ and simultaneously by the eye. Such simultaneous stimulation causes "and" neurons to give a brief but high-frequency burst of impulses. We have found that the most effective stimulus for these cells is a small, warm, moving object. These neurons might be described whimsically as mouse detectors.

Infrared-depressed visual cells, on the other hand, respond best to thermally neutral or cool visual objects: leaves fluttering in a breeze or perhaps a frog jumping at a pond's edge. The visual activity of these cells would be depressed by the infrared radiation of warm-blooded animals. In such examples cross-modality interactions are essential for correct identification of the stimulus.

The cross-modality interactions we have described show that in the rattlesnake tectum the infrared and visual sense modalities are integrated to a significant degree. Could similar modality-combining interactions be operating in the optic tectums of other species? Are tactile and auditory signals in mammals combined in complex ways with visual input? We simply do not know. The experiments necessary to identify such interactions have not been done. We would be willing to bet, however, that the kinds of modality interactions we have observed in snakes are present and functionally important in the tectums of most other vertebrates.

It is ironic that so little is known about how the infrared sense is exploited by snakes in their natural habitat. No one knows whether rattlesnakes and pythons make use of their infrared detectors to hunt prey during the day, when their visual system is also operational. Nor is it known whether snakes rely on their infrared system to identify predators and to locate a comfortable resting place. These are interesting questions for the future. For the present it is impressive enough to know that in the pit vipers and the boid snakes the trigeminal skin sense has evolved into a sensitive distance sense that is much like vision itself.



HYPOTHETICAL WORLD VIEWS perceived by a rattlesnake are represented schematically. The picture at the left depicts a scene as it might be recorded by the snake's visual system. The picture in the middle is the same scene as it might be perceived through the snake's infrared system. Only objects that are warm (*color*) or cool (*gray*) and moving would be detected. Stationary objects such as the bush and the pond would not be represented because they do not stimulate the infrared system well. Even the sun-warmed rock would be absent from the infrared image because it is also stationary. The images would be blurred because of the poor imaging properties of the pit organ. The picture at the right suggests how the optic tectum might analyze the combined input of the two sensory systems. Responses of three of the six classes of bimodal tectal cells are represented by squares, triangles and circles. "Or" cells would be strongly activated (*solid color*) by the warm, clearly visible mouse and activated less intensely (*light color*) by the infrared component of the rabbit, which is hard to see against the bush, and by the visual component of the cool frog. "And" cells would be activated only by the mouse, which is both warm and clearly visible. The frog's cool skin would serve to disinhibit the "infrared-depressed visual" cells and cause them to respond strongly.

Cocaine

"Recreational" users who take the drug under controlled conditions often cannot distinguish it from other drugs or even from a placebo. Yet huge profits have made cocaine a major international commodity

by Craig Van Dyke and Robert Byck

A few hundredths of a gram of cocaine hydrochloride, chopped finely and arranged on a smooth surface into several lines, or rows of powder, can be snorted into the nose through a rolled piece of paper in a few seconds. The inhalation shortly gives rise to feelings of elation and a sense of clarity or power of thought, feelings that pass away for most people in about half an hour. Although the growing interest in cocaine in the U.S. and Western Europe is in part a consequence of this simple, hedonistic experience, the real importance of the drug derives from the interaction of social, economic and political factors: the current status of cocaine as a fashionable drug for "recreational" use, its artificially inflated price, the control of its distribution by criminal organizations, the stresses its barter and untaxed trade impose on the monetary and taxation systems, the strains it places on the resources of law enforcement, the threats to civil liberties posed by drug laws that are not based on a rational understanding of the drug's effects and the potential of the cocaine trade for corrupting officials and undermining respect for the law.

Estimates of the cost and consumption of cocaine are subject to the biases of the reporting source and should therefore be viewed with skepticism. According to a 1979 report by the White House Strategy Council on Drug Abuse, some 10 million Americans had taken cocaine within the preceding 12 months, compared with 10,000 people 20 years before. The National Narcotics Intelligence Consumers Committee (NNICC) has estimated that in 1979 between 25,000 and 31,000 kilograms of cocaine entered the U.S. illegally. During 1980 U.S. cocaine imports are estimated to have been between 40,000 and 48,000 kilograms. These quantities are based on estimates of the refining capacity of clandestine laboratories, on estimates of the proportions of the refined product that are sold in the domestic and foreign markets and particularly on estimates of the production yields of raw coca leaves in the source countries. Perhaps not surprisingly, the estimates of crop yield made in the source countries are in many cases smaller than U.S. estimates by a factor of two or more.

The illicit retail dealer can sell a white, crystalline substance consisting of from 10 to 85 percent cocaine and various other substances for between $100 and $140 per gram. The NNICC, from its estimate of 1980 cocaine imports, puts the retail value of the industry at between $27 and $32 billion. If the cocaine trade were included by *Fortune* in its list of the 500 largest industrial corporations, cocaine would rank seventh in volume of domestic sales, between the Ford Motor Company and the Gulf Oil Corporation. Based on U.S. estimates, the monetary value of Bolivia's cocaine exports may now surpass the value of the country's largest legal industry, tin. Colombia's more highly refined cocaine exports total about $1 billion annually, half the value of the coffee crop.

It is all too easy to suppose the physiological and social consequences of the use of cocaine are commensurate with its popularity and economic importance. Actually the assessment of the medical and psychological implications of short- and long-term cocaine usage has only recently begun. In 1975 the National Institute on Drug Abuse began a research project intended to define the pharmacology of cocaine in man. A number of investigators have since been engaged in various aspects of cocaine-related research, including a detailed description of the effects of the drug in man, its distribution and metabolism in the body, its reinforcing properties and its potential for abuse.

After six years of work on the problem our group at the Yale University School of Medicine has been able to describe with reasonable confidence the time course of the basic pharmacological effects that follow the introduction of cocaine into the body by various routes. We have also been able to associate the time course of cocaine concentration in the blood with replicatable measures of its psychological effects. Although cocaine is intensely pleasurable to some people, we have found that its ability to produce a unique "high" may be overrated: our subjects, all experienced cocaine users, could not distinguish a single dose of cocaine taken intranasally from the same quantity of the synthetic local anesthetic lidocaine. Investigators at the University of Chicago School of Medicine found that their subjects could not distinguish the immediate effects of intravenous cocaine from those of amphetamine, although at later times the differences between the drugs are apparent. Such results are the first steps toward distinguishing the almost overwhelming mythology that surrounds cocaine from reliable information about its effects.

At the same time there has been renewed interest in understanding the history of cocaine consumption as well as its significance to the Indian cultures of the Andes, including parts of present-day Colombia, Ecuador, Peru, Bolivia and Chile. Of course, the Andean social and pharmacological experience with the coca leaf cannot be compared directly with the current social experience with the more potent substance cocaine hydrochloride, but one can still hope to gain a cultural perspective on the drug. Such a perspective is at least as important to informed public policy as knowledge of the biochemical and pharmacological action of cocaine is.

Cocaine is an alkaloid, a member of a broad group of plant substances that also includes nicotine, caffeine and morphine. In nature cocaine is found in significant quantities only in the leaves of two species of the coca shrub. *Erythroxylum coca* requires a moist, tropical climate. According to Timothy Plowman of the Field Museum of Natural History in Chicago, *E. coca* may be native to the Peruvian Andes, although it now grows throughout the eastern highlands of the Andes in Ecuador, Peru and Bolivia. The concentration of alkaloidal cocaine in its leaves can be as high as 1.8 percent. *E. coca* was cultivated by the

COCA-LEAF CHEWING, represented here in a ceramic sculpture, has been an important part of the culture of the peoples of the Andes for at least 5,000 years. It is from the leaves of the coca plant that the alkaloid cocaine is extracted; chewers of the leaves absorb the substance directly. The sculpture depicts a man whose cheek is puffed out with coca leaves and who carries the implements associated with chewing in his hands. The small gourd in his left hand holds *llicta,* an alkaline powder he adds to the coca in his mouth by wetting the spatula in his right hand with saliva. The powder then adheres to the spatula. In 1979 David Paly of the Yale University School of Medicine showed that the addition of an alkaline substance to coca leaves increases the concentration of cocaine in the plasma of the blood by a factor of 10. The sculpture, which stands 19.7 centimeters high, is in the collection of the Museum of the American Indian in New York City. It has been attributed to Phase I of the Moche culture, which flourished on the northern coast of Peru from 100 B.C. to A.D. 100.

Incas, and it remains the primary source of cocaine for the illicit world trade.

E. novogranatense, the other cocaine-rich species of *Erythroxylum*, is cultivated in drier, mountainous regions of Colombia and along the Caribbean coast of South America. The *truxillense*, or "Trujillo," variety of the latter species is now grown on the northern coast of Peru and in the dry valley of the Marañón River, a tributary of the Amazon in northeastern Peru. Its leaves are harvested for legal export to the Stepan Chemical Company in Maywood, N.J., where the cocaine is extracted for controlled pharmaceutical purposes and the remaining leaf material is prepared as a flavoring for Coca-Cola.

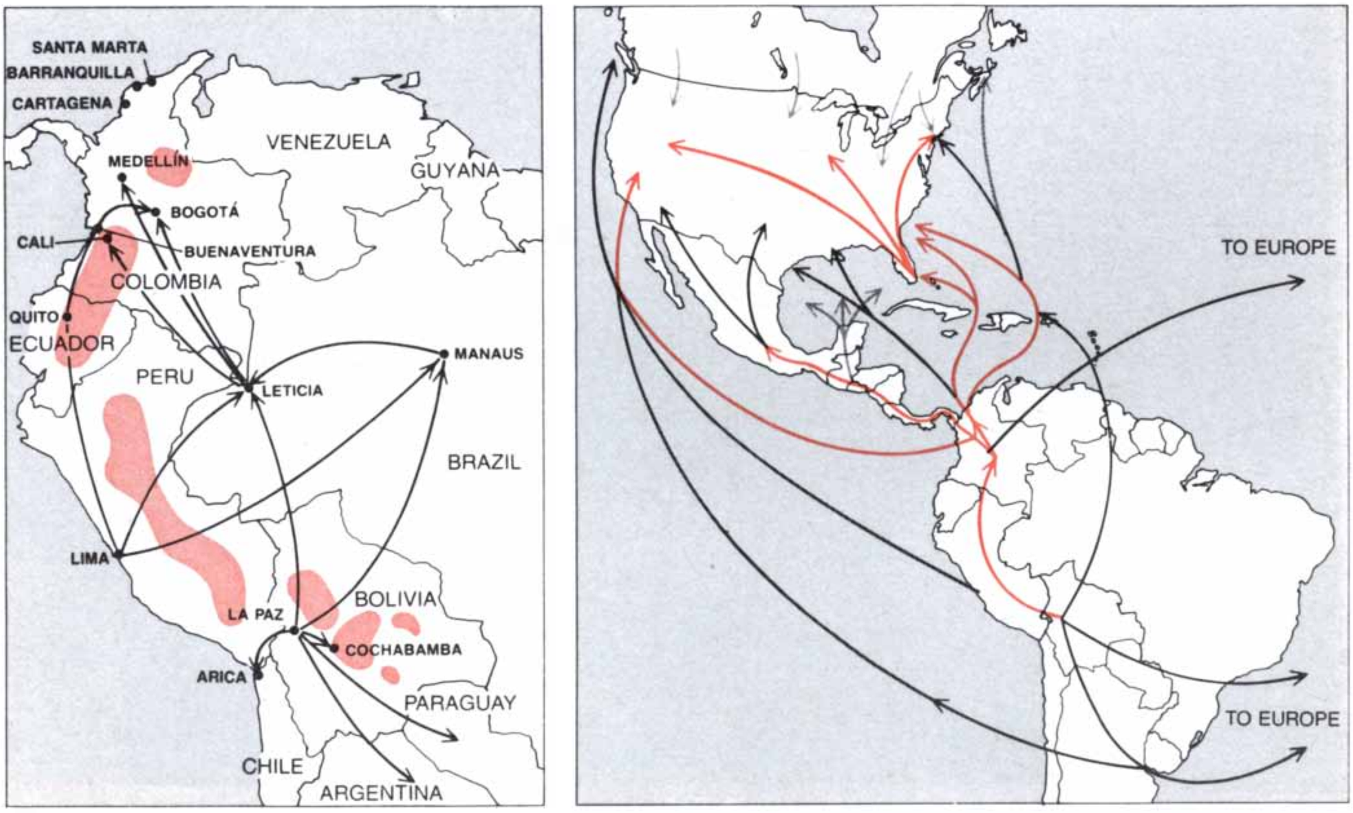
Recent archaeological findings in Ecuador indicate that human experience with cocaine dates back at least 5,000 years, long before the Inca empire was established. To the Incas coca was a plant of divine origin and a symbol of high social or political rank. According to one myth, the god Inti created coca to alleviate the hunger and thirst of the Incas, who believed themselves to be descendants of the gods. The Inca state controlled virtually every aspect of daily life, including the cultivation and use of coca. Coca was chewed primarily by the ruling classes, although on occasion it was disbursed to soldiers, workers or runners. Casual chewing of the leaves was considered a sacrilege.

The first reports of coca chewing to reach Europe were almost coincident with the discovery of the New World. The letters of Amerigo Vespucci, published in 1507, mention the Indian practice of chewing leaves and adding ashes to the cud. A major component of the ash is carbonate of lime, or calcium carbonate ($CaCO_3$), which intensifies the subjective effects of the coca leaves; among the Andean Indians the practice of adding lime continues today. Other European observers were first mystified, then repulsed and finally impressed by the Indian's use of coca. A manuscript completed in 1613 by Don Felipe Guaman Poma de Ayala represents coca chewing as an unauthorized social activity engaged in by the Indians when they were expected to be working.

Other accounts of coca chewing from explorers and chroniclers were more enthusiastic. Coca was reported to cause a striking increase in endurance, enabling men to do hard work with little food at high altitudes. By 1569 the Spanish entrepreneurs in South America had recognized the utility of coca in recruiting Indians for labor, and Philip II of Spain gave their practice official approval by declaring coca necessary to the well-being of the Indians.

If coca leaves retained their potency after drying and long travel as successfully as tobacco leaves and coffee beans do, the role of coca in European and North American society might be quite different from what it is today. Botanical specimens of coca arrived in Europe soon after the Spanish exploration, and coca leaves continued to be exported to Europe in small quantities throughout the 17th, 18th and early 19th centuries. In spite of the glowing descriptions of the effects of coca among the Indians, however, Europeans did not adopt the practice of chewing coca. In terms of its cultural importance the real crossing of the Atlantic by coca had to await the development of 19th-century methods for isolating chemical compounds. When the pure alkaloid and related substances were made available by these means, it was in a social context quite different from the one that prevailed when tobacco and coffee were taken up in the 16th century. The modern social history of cocaine has been determined in large part by the attitudes that formed in the late 19th century and by the reactions to those attitudes.

The German chemist Friedrich Gaed-



ILLICIT COCAINE TRADE begins on the eastern slopes of the Andes, where coca is harvested (*colored areas on map at left*). The harvest is processed into cocaine paste, and the cocaine paste moves by circuitous routes to clandestine laboratories primarily in Colombia (*arrows on map at left*), where it is converted into cocaine hydrochloride. The major international smuggling routes lead from port cities and back-country airstrips to various ports of entry in the U.S. and Europe. On the map at the right the routes with the largest volume of cocaine traffic are indicated by colored arrows; secondary routes are shown in black, and the remaining routes are shown in gray. When the cocaine hydrochloride reaches the countries in which it is distributed, it is cut to between 10 and 85 percent purity with various adulterants. An estimated 40 to 48 metric tons of cocaine entered the U.S. in 1980, worth some $27 to $32 billion at retail prices.

cke was probably the first to isolate alkaloidal cocaine from the coca leaf in 1855, when he prepared an oily substance and small crystals from the distilled residue of coca extract. Albert Niemann of the University of Göttingen was the first to characterize the substance chemically in 1859. Niemann reported that cocaine had a bitter taste and that, after a short time, it numbed the tongue. By 1880 Vassili von Anrep, a Russian nobleman and physician at the University of Würzburg, noted he could not feel a pinprick after he administered cocaine subcutaneously. Meanwhile, because cocaine seemed to have actions opposite to those of drugs that depress the central nervous system, it was studied by American physicians as a possible antidote to the habits engendered by morphine and alcohol.

In 1884 enthusiastic reports about cocaine came to the attention of Sigmund Freud, who was then a young house physician in the neurology section of a Viennese hospital. Although von Anrep was one of his instructors, Freud took an interest in cocaine not primarily as a local anesthetic but rather as a stimulant of the central nervous system and as an aid to overcoming morphine addiction. He ingested cocaine himself and published *On Coca,* a review of the literature on coca and cocaine.

On Coca stirred a ferment of interest in cocaine. The review expressed a highly favorable opinion of the drug, and in letters of the same period Freud rhapsodically described how "a small dose lifted me to the heights in a wonderful fashion." Although he suggested the therapeutically useful possibility that cocaine could reduce the pain of inflammation, it remained for his friend and associate Carl Koller to recognize the full implications of the properties of cocaine as a local anesthetic. Koller applied it to the eye during surgery and so established the critical link between local anesthetic properties and surgical procedure.

COCA SHRUB (*Erythroxylum coca*) flourishes in wet, tropical forests on the eastern slopes of the Andes at an elevation of between 500 and 1,500 meters. Alkaloidal cocaine in concentrations as high as 1.8 percent is found in the leaf of the shrub. Under cultivation the plant is trimmed to a height of about two meters so that the leaves can easily be harvested. Cocaine is also found in other species of the genus *Erythroxylum,* but *E. coca* is richest in the alkaloid.

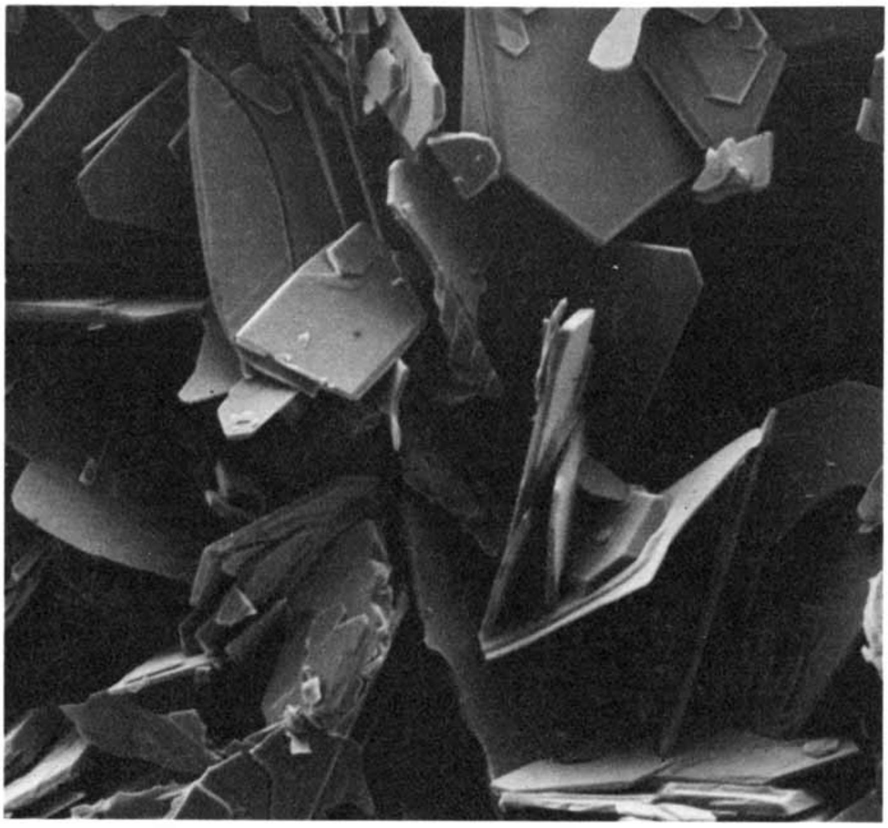
The attempt to show that cocaine and morphine could be employed antagonistically, so as to counter each other's effects, did not work out as well. The American surgeon William Stewart Halsted became one of the first victims of cocaine dependence as a result of experiments he carried out on himself to show that the drug could act as a "nerve blocker." Isolation aboard a schooner failed to control the habit, and Halsted was treated with morphine in the hope that its apparently opposite pharmacological effects would cancel his desire for cocaine. Unfortunately, to his psychological dependence on cocaine Halsted simply added a physical dependence on morphine, which lasted to the end of his life.

As the ineffectiveness of cocaine in treating morphine addiction came to be recognized and reports of serious preoccupation with cocaine began to circulate, Freud's enthusiasm for the drug began to wane. Although he continued to enjoy his own ingestion of the drug, he was shaken when, three years after he had begun his first investigations into cocaine, Albrecht Erlenmeyer accused him of having unleashed "the third scourge of humanity" (after alcohol and the opiates).

The descriptions of the effects of cocaine in man given by Freud and his associates were not superseded for nearly 90 years. Most subsequent scientific work on the topic has been concerned with establishing biochemical mechanisms that might account for Freud's observations; there has also been much hearsay, anecdote and invective. Interest in cocaine was sustained chiefly because of its properties as a local anesthetic, and as a result of this interest cocaine was one of the first alkaloids to be chemically synthesized. Richard Willstätter and his colleagues at the University of Munich achieved the synthesis in 1923, but the spatial structure of the cocaine molecule remained unknown until E. von Hardeggar and Hans Ott worked it out in 1955.

A local anesthetic blocks the conduction of a sensory nerve impulse when the drug is applied directly to the nerve. Whereas many substances can block nerve conduction by injuring or permanently destroying nervous tissue, a local anesthetic is unusual in that its effect is temporary and reversible. Many of the drugs with local anesthetic properties that have been synthesized since the turn of the century, such as procaine (Novocain) and lidocaine (Xylocaine), are structurally similar to cocaine and work by a similar mechanism.

A local anesthetic inhibits a nerve impulse by altering the membrane of the nerve cell. When the cell is in its resting state, there is an electric-potential difference of from 60 to 70 millivolts between the inside of the axon (the impulse-conducting filament that emerges from the body of the cell) and the fluid surrounding the axon. The axoplasm inside the axon is electrically negative with respect to the fluid outside the cell membrane. When the cell is stimulated, the signal called the action potential travels down



CRYSTALS OF COCAINE HYDROCHLORIDE are shown magnified 600 diameters in a scanning electron micrograph made by Alan Pooley of the Peabody Museum of Yale University and one of the authors (Byck). The nearly pure substance is seldom encountered by cocaine users. The crystals dissolve on mucous membranes and are absorbed into the bloodstream.

the axon as a wave of electrical depolarization. A slight depolarization of the axon membrane usually causes a large transient increase in the permeability of the membrane to positively charged sodium ions outside the cell. As the ions pass through channels in the membrane into the axoplasm, the potential difference between the axoplasm and the surrounding fluid is reduced. The movement of charge is a self-limiting process, and soon after the action potential is initiated potassium, calcium and other positively charged ions leak out of the axon and restore the electrochemical equilibrium of the cell.

It is not known precisely how a local anesthetic changes the axon membrane, but it is now thought that molecules of the anesthetic dissolve in the lipid matrix that composes the membrane and bind to receptor sites within the sodium channels. The presence of the anesthetic molecule at the receptor site probably interferes with the opening of the channel, thereby inhibiting the transfer of sodium ions across the membrane. Thus the depolarization of the axon is prevented and the nerve impulse is blocked. The action of the anesthetic is terminated by the breakdown of the anesthetic molecules and by their diffusion into the bloodstream.

Another property of cocaine that has been valued in clinical medicine is its tendency to constrict blood vessels when it is applied topically. Cocaine is the only local anesthetic that has this effect, and until recently it was the anesthetic of choice in eye surgery because it limits the flow of blood. It has since been found, however, that the reduced flow can damage the surface of the eye, and so cocaine is no longer recommended for use by ophthalmologists. It retains a role in surgery of the mucous membranes, such as those of the ear, nose and throat, as well as in procedures that require the passage of a tube through the nose or throat.

The constriction of blood vessels is one of several peripheral effects caused by the stimulation of the sympathetic nervous system. Drugs that act on sympathetic nerves to mimic such effects are called sympathomimetic drugs. Cocaine is a typical sympathomimetic agent in that it increases the heart rate, raises the blood pressure and, in large doses, increases the body temperature and dilates the pupils of the eyes.

The continual transmission of signals along the sympathetic nerve pathways is essential to life. In the axon the signal takes the form of an action potential, but when the action potential reaches the end of an axon, the signal is conveyed to the next nerve cell by a different mechanism, namely the release of a neurotransmitter such as dopamine or norepinephrine. The neurotransmitter is stored in vesicles in the axon near the synapse, the junction between the axon of one nerve cell and the dendrites of the next. When the axon fires, the vesicles fuse with the cell membrane, releasing the neurotransmitter into the synaptic cleft between the cells. The released molecules stimulate the succeeding nerve by attaching themselves to receptor sites on its dendrites.

Ordinarily the neurotransmitter molecules that do not stimulate the succeeding nerve cell are broken down by enzymes, diffuse into adjoining tissues or are pumped back into the cell terminal that released them. The pumping action is called reuptake. When cocaine molecules are present in the synaptic cleft, however, they inhibit the reuptake mechanism and so mimic the effects of the release of more neurotransmitter molecules. The neurotransmitter tends to remain in the cleft where it can continue to stimulate the receptors in the dendrites of the succeeding nerve cell.

It was once thought the psychologically stimulating effects of cocaine could be attributed solely to its ability to block the reuptake of norepinephrine. It is now known this is not the case. Many of the antidepressant drugs employed in psychiatry, such as amitriptyline (Elavil), block the reuptake of neurotransmitters, but they are not stimulants and they do not induce euphoria. On the other hand, there would seem to be little reason to suppose a local anesthetic such as lidocaine would induce euphoria either, although our investigations have shown that it does. A single dose of lidocaine, when administered into the nose, cannot be distinguished from the same quantity of cocaine. Moreover, other groups have shown that animals will work at a task to get injections of procaine just as they will for injections of cocaine. Since the molecules of cocaine, lidocaine and procaine are structurally similar, such findings may suggest merely that similar drugs cause similar experiences. They may also suggest, however, that certain kinds of drug-induced euphoria are generalized interpretations by the drug user of a rather wide range of unusual sensations caused by the interaction of a number of physiological and environmental factors.

The effects of a drug depend on the dosage, the form in which the drug is taken and the route by which it enters the body. Many effects also vary with the frequency with which the drug is taken: the body may accumulate the drug or develop a tolerance to it. The effects of psychoactive drugs also differ according to the expectations of the user, the setting or circumstances in which the drug is taken and the history and personality of the user. The strength of the last three variables in determining the overall effects of a drug makes it difficult to estimate just what portion of the cocaine experience should be attributed to the substance itself and what portion to ambience or expectation.

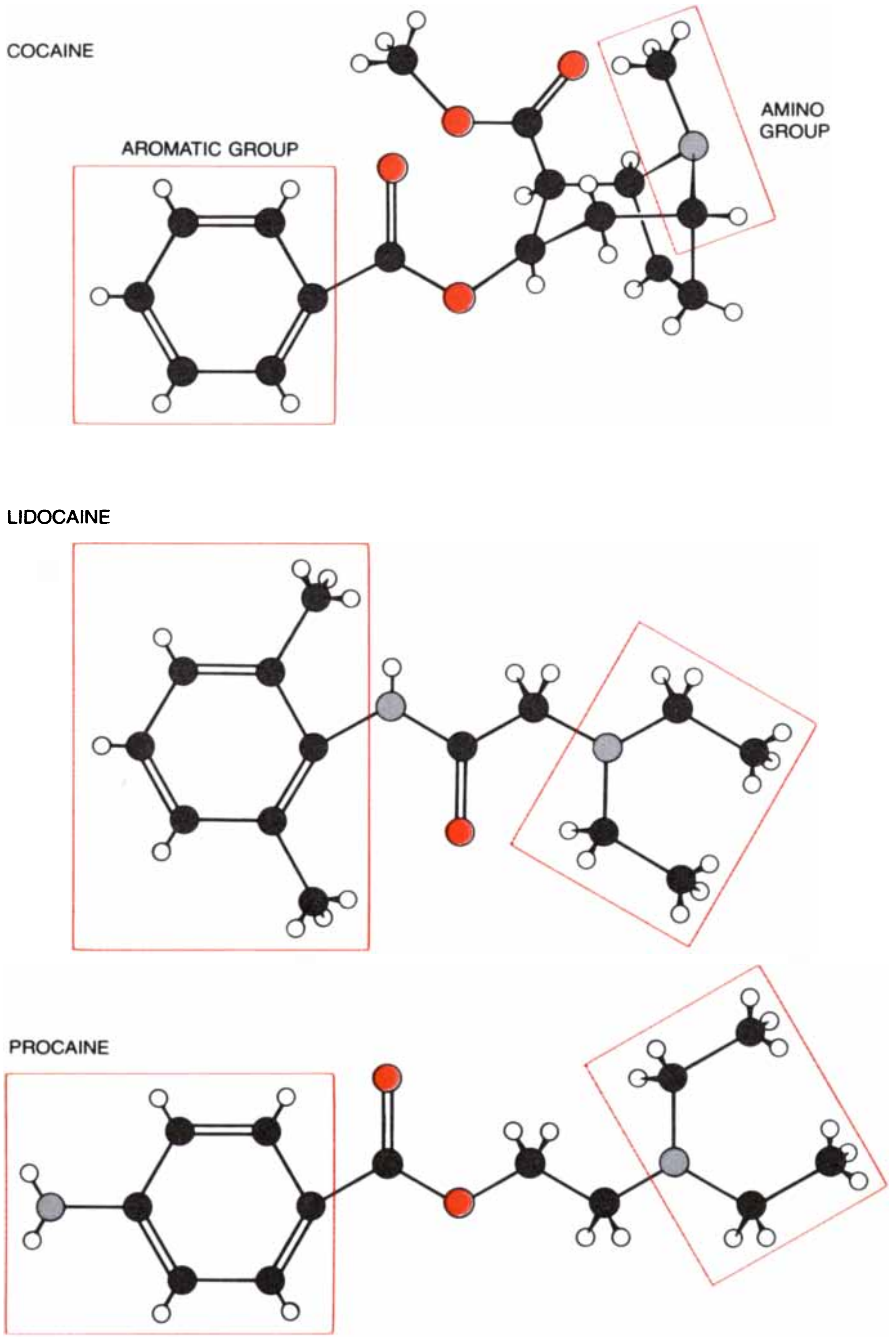
When cocaine is taken recreationally, expectation and mythology about the drug may generate subjective impressions that overwhelm any effects not consonant with the expectations. Moreover, the cocaine that is available on the street is likely to be adulterated, which complicates the pharmacology of the drug. The adulterant might be a cheaper stimulant such as amphetamine, a simple carbohydrate such as mannitol or lactose, a local anesthetic such as procaine or lidocaine or almost any other substance that is not immediately detectable by the cocaine buyer. On the other hand, when cocaine is given in controlled laboratory conditions, the stimuli that may heighten or modify its effects are ordinarily absent. Knowledge of the acute effects of cocaine must come from a blend of individual descriptions, which may be biased or fanciful, and from the reports of laboratories, which may desiccate the drug experience beyond recognition.

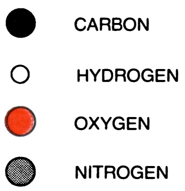
Even in the laboratory the effects of the user's expectations cannot be eliminated entirely. The ethics of research on drugs with human subjects allow only experienced users of cocaine to be employed in the studies of its effects. Such people may skew experimental results because their expectations about cocaine have already been formed in settings far different from those of the laboratory. To minimize such biases it is important that neither the experimenter nor the subject know what dosage (if any) has been given; to control for expectations a placebo must be interspersed with active doses of cocaine.

To establish a framework within which to organize our subjects' responses to cocaine we required a measure of the cocaine actually present in the body. In 1975 David Bailey and Peter Jatlow of the Yale School of Medicine developed a sensitive and specific assay for determining the concentration of cocaine in blood plasma. Their method employs a gas chromatograph, an instrument in which the molecules of a sample gas are forced through a column packed with an adsorbing material. Various kinds of molecules in the gas are adsorbed and released at different rates, so that a substance can be identified by the characteristic time it takes to pass through the column. To increase the sensitivity of the device to cocaine, Bailey and Jatlow installed a detector particularly sensitive to compounds that contain nitrogen, as cocaine does. The method made it possible for the first time to measure concentrations of cocaine in the blood as low as five nanograms per milliliter. The effects of doses that are commonly taken in recreational settings and before surgical procedures can thereby be traced.

Until the work of Bailey and Jatlow the determination of cocaine concentrations in blood was hindered by the presence in blood of the enzyme pseudocholinesterase, which is responsible for the breakdown of the neurotransmitter acetylcholine and which also breaks down cocaine. By adding fluoride ions to fresh plasma Bailey and Jatlow found that the action of the enzyme could be inhibited. The technical advance has in turn made possible the determination of blood-plasma levels of cocaine among coca chewers in South America, in settings that are not directly accessible to laboratory measuring instruments.

In the form of the hydrochloride salt, cocaine is readily absorbed through mucous membranes and enters the bloodstream. For unknown reasons the most popular route of administration is





STRUCTURE OF COCAINE resembles that of the synthetic local anesthetics lidocaine (Xylocaine) and procaine (Novocain). All three molecules incorporate an aromatic group that accounts for their solubility in fatty tissues, such as the brain, and an amino group, which includes a nitrogen atom. Lidocaine and procaine are among the adulterants sometimes added to the cocaine sold on the street. The authors found that when cocaine is administered intranasally, it cannot be distinguished from lidocaine even by experienced cocaine users.

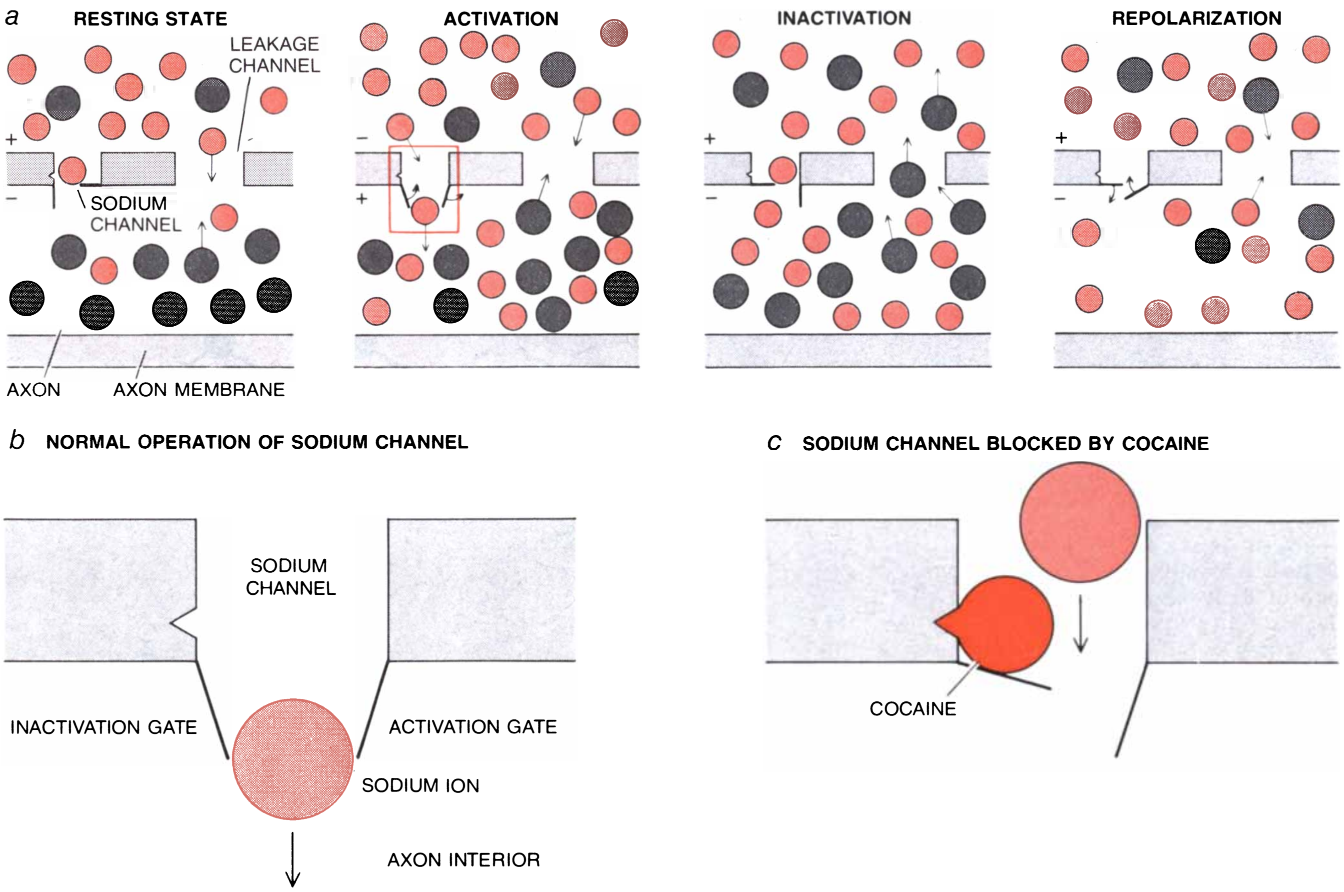
through the mucous membranes of the nose, although our group has demonstrated that crystals of cocaine hydrochloride are efficiently dissolved and absorbed from the gastrointestinal tract as well. Cocaine can also be introduced intravenously, or it can be absorbed through the lungs by smoking the alkaloidal substance, which users call free base. In South America an intermediate product in the manufacture of cocaine hydrochloride, called cocaine paste or coca paste, is now smoked by urban youths. Cocaine paste is approximately 30 to 90 percent free-base cocaine. In the U.S. amateur laboratory kits are sold for converting cocaine hydrochloride into free-base cocaine, which is less susceptible to heat decomposition. The free base is vaporized by hot gases generated in a device called a base pipe, and the vapors are inhaled.

Cocaine has been given intravenously in some laboratory studies because the moment when the drug is completely absorbed can be defined precisely. The concentration of cocaine in the blood peaks immediately after the injection and thereafter falls off with a half-life of from 45 to 90 minutes. Experimental subjects report intense euphoria, sometimes followed by a "crash," or extreme dysphoria, and a craving for more cocaine.

The effects of intravenous cocaine are so short-lived that many of the rather time-consuming tests of its effects cannot be carried out. We chose instead to examine the physiological and psychological effects of cocaine introduced into the nose. We did not ask our subjects to snort the cocaine crystals through a straw or a rolled dollar bill; nor did we follow the common street ritual of administering relatively small amounts of cocaine (about 30 milligrams) at regular intervals. Instead we gave single but higher doses from a nasal spray so that we could dissect the time course of the action of the drug.

The physiological effects of cocaine introduced in this manner are not striking. Not surprisingly, we found that higher doses induced higher blood-plasma levels, although, in apparent contradiction to the reputation of cocaine as a short-lived drug, the peak plasma concentrations are reached about an hour after the drug is taken. Doses of from 25 to 100 milligrams increase the heart rate by between eight and 10 beats per minute and raise both systolic and diastolic blood pressure by about 10 millimeters of mercury. These effects peak about 15 to 30 minutes after the drug is administered. At the relatively high dose of two milligrams per kilogram of body weight (or 140 milligrams for a person of average weight) there is a slight temperature increase, but the pupils of the eyes do not necessarily dilate.

Although cocaine, like all other local anesthetics, causes convulsions at very high doses, there is no discernible effect on the electroencephalogram at a dose of two milligrams per kilogram of body weight. In animals, however, repeated



LOCAL ANESTHETIC inhibits the transmission of sensory-nerve impulses by changing the permeability of the axon membrane to sodium ions. Normally the impulse propagates as a wave of depolarization passing along the axon membrane; during the depolarization the electric-potential difference between the axoplasm and the fluid surrounding the axon is momentarily altered. As is shown in the schematic diagram (*a*), a leakage channel allows positively charged ions both inside and outside the axon to cross the axon membrane in both directions. Nevertheless, the concentration of sodium ions (*colored circles*) is maintained at a higher level outside the axon than it is inside by a sodium pump (not shown), whereas the concentration of potassium, calcium and other ions (*black circles*) is greater inside the axon. (In the diagram the relative concentrations have been exaggerated for clarity.) In the resting state equilibrium is reached when the interior of the axon is negatively charged with respect to the exterior. When the nerve is stimulated (*b*), activation gates in the sodium channels of the membrane begin to open, allowing only the small sodium ions to migrate into the axoplasm. The voltage difference between the inside and the outside of the axon is thereby reduced. As the sodium channels are closed by inactivation gates, other ions move out of the axon to repolarize the membrane. In the presence of cocaine (*c*) the sodium ions cannot pass through the sodium channel, mainly because it is blocked by the cocaine but also because the inactivation gate is kept more tightly closed. Hence the nerve impulse is blocked.

Frederick Forsyth's Rolex is like his novels. Tough, accurate and very stylish.

Frederick Forsyth is not a prolific writer.

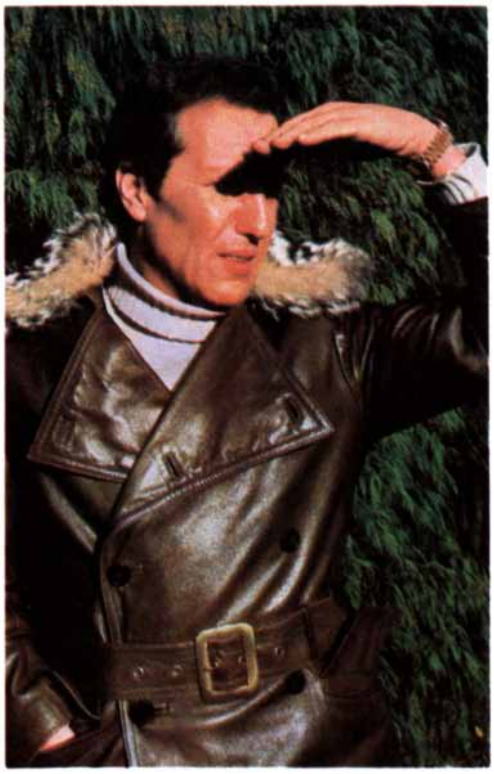
In fact, in the past twelve years he has completed just four full-length novels.

And yet *The Day of the Jackal, The Odessa File, The Dogs of War* and *The Devil's Alternative* have all become instant best-sellers around the world.

Already his first three books have been made into successful feature films.

Forsyth's writing is characterized by a blend of uncannily authentic detail and superb storytelling.

The facts are drawn from his own many experiences as a front-line war correspondent; the fiction, from something the craftsmen at Rolex appreciate only too well — a sense of style.



Frederick Forsyth wears a Rolex Oyster Day-Date in 18kt. gold, with matching President bracelet.

"It is very tough and well made," he says. And, it is also immensely practical.

"I can wear my Rolex all the time. I never have to take it off, even to use a chain saw. Nothing seems to bother it."

Apart from his Rolex, Frederick Forsyth is particularly pleased with the coat you see him wearing in the photograph.

He spotted it in a shop in London, and asked of what fur the collar was made.

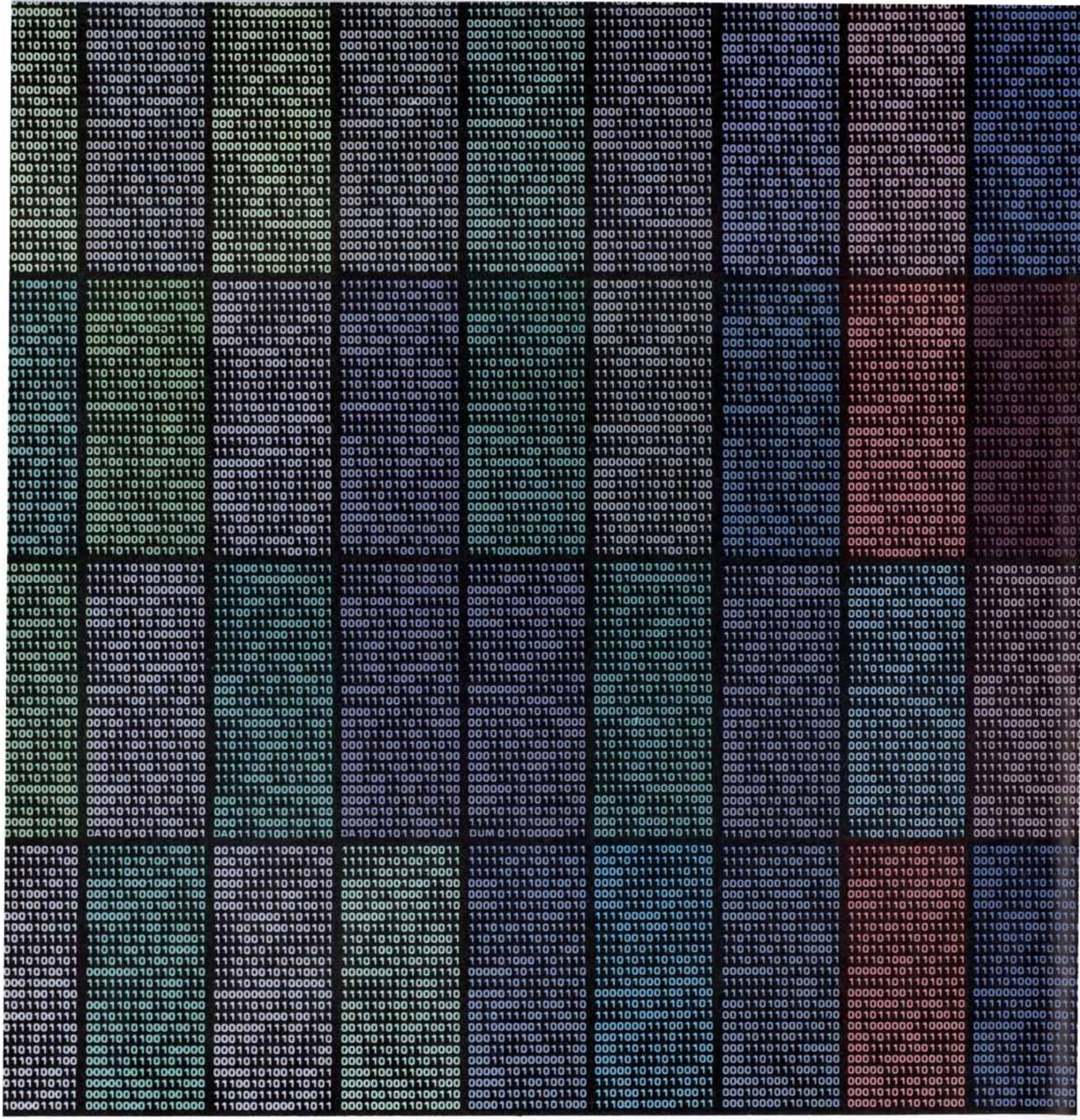
The assistant told him.

"Jackal."

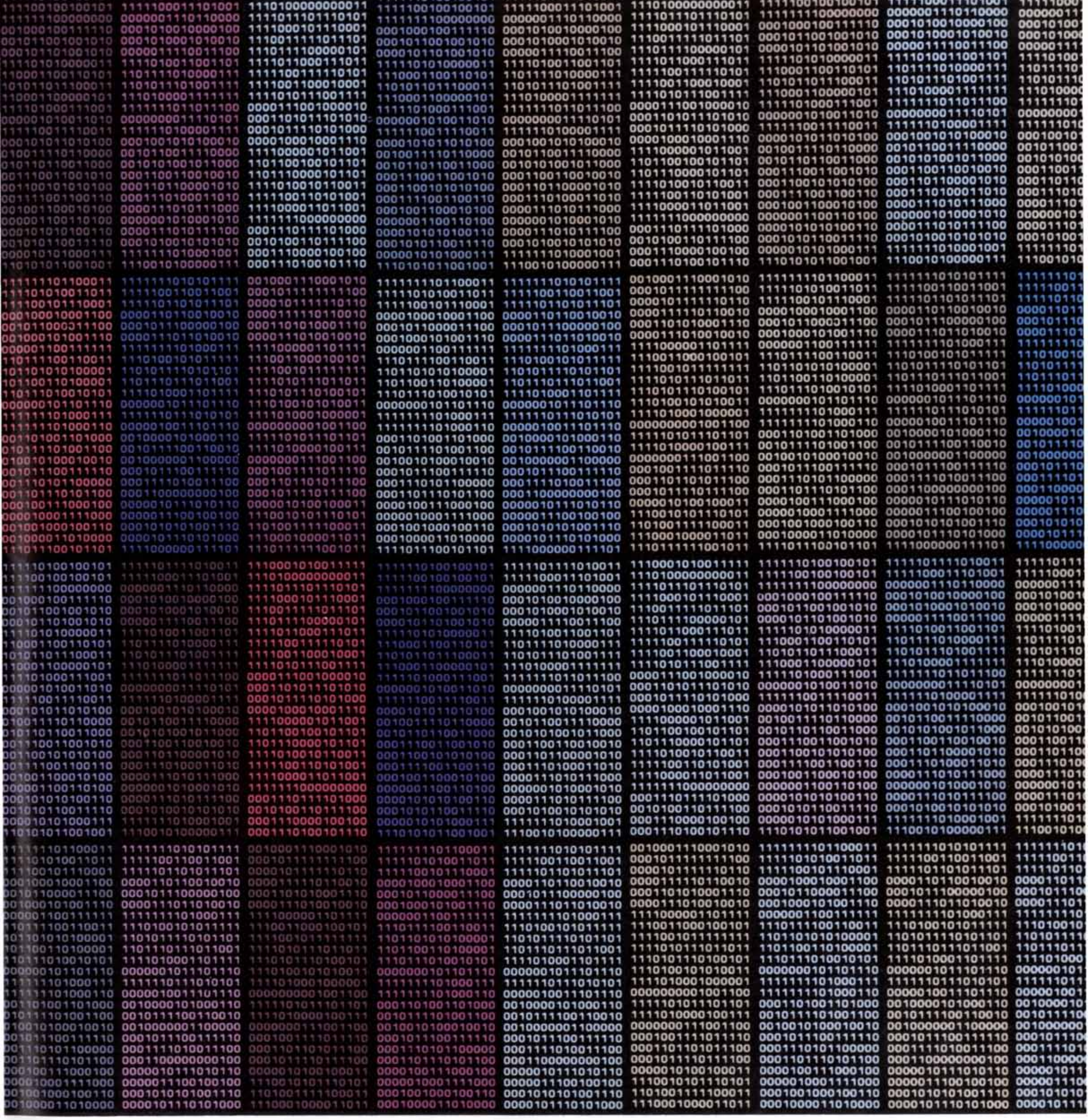
ROLEX

Pictured: The Rolex Day-Date Chronometer. Available in 18kt. gold, with matching President bracelet.

Write for brochure. Rolex Watch, U.S.A., Inc., Dept. 116, Rolex Building, 665 Fifth Avenue, New York, New York 10022-5383. World headquarters in Geneva. Other offices in Canada and major countries around the world.



These are the first four notes of Beethoven's Fifth Symphony.



Musical notes like these never existed before 3M invented the Digital Mastering System.

It uses computer technology to record every nuance of a piece of music in binary numbers. So that when reproduced, every note can be heard with such uncanny crispness and clarity, you can hear music the way Beethoven wanted it to be heard.

At 3M, by listening to people's needs, we've pioneered over 400 products to serve the needs of the communication arts field. We've developed everything from videocassettes to lithographers tape to photo offset plates.

All because at 3M, we're in the business of hearing. So let us hear from you.

3M hears you...

For your free 3M Communication Arts Brochure, write: Department 080703/ 3M, P.O. Box 4039, St. Paul, MN 55104.

Name________________________

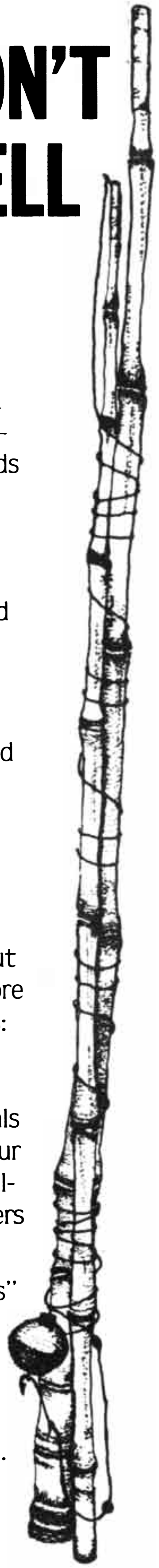
Address________________________

City____________ State & Zip______

Or call toll-free: 1-800-323-1718, Operator 368. (Illinois residents call 1-800-942-8881.)

3M

THE FISH WON'T BE BITING IN TWITCHELL CREEK TODAY,



OR TOMORROW, EITHER.

There was a time when with a little luck and a little patience, you could pull a fair-sized fish out of Twitchell Creek on Woods Lake in the Adirondacks.

But no more.

The fish are gone, along with the salamanders, ospreys, mayflies, swallows and myriad other creatures who once lived along the cool river banks.

They've been run off, or killed off by the rain, of all things. A deadly poisonous acid rain which has contaminated the water, choked the life out of the stream . . . and broken the delicate food chain on water and (we're finding out) on land as well.

All rain contains some acid, of course. But acid precipitation is different. And far more dangerous. Acid rain contains two killers: nitric acid and sulfuric acid which form when sulfur dioxide and nitrogen oxide mix with rain water. These two chemicals are being spewed by the ton-load into our air every day . . . emissions from the coal-burning power plants and industrial boilers our nation uses to keep going.

Fortunately, some lakes contain "buffers" . . . neutralizing agents which help lessen acid damage. But what of the others . . . in the Adirondack Mountains, in western Virginia, in the Great Smoky Mountains, throughout New England . . .

Who Will Stop The Rain?

The Izaak Walton League is working to do just that right now. The League was formed in 1922 by a handful of sportsmen who wanted to combat water pollution. And it endures today as a grassroots conservation organization composed of 50,000 fishermen, hikers, hunters and campers who speak out—and work hard—to protect wild America.

Congress passed a Clean Air Act in 1970—soon up for renewal—which set allowable limits for sulfur dioxide emissions from power plants. Some would like to relax those laws now . . . asserting the regulations will retard energy development. We disagree. We want stricter regulations to reduce these emissions still further. And we'll get them. Once and for all. Because there's something at stake here far greater than fishing. And that's life itself.

For more information on our activities, write:

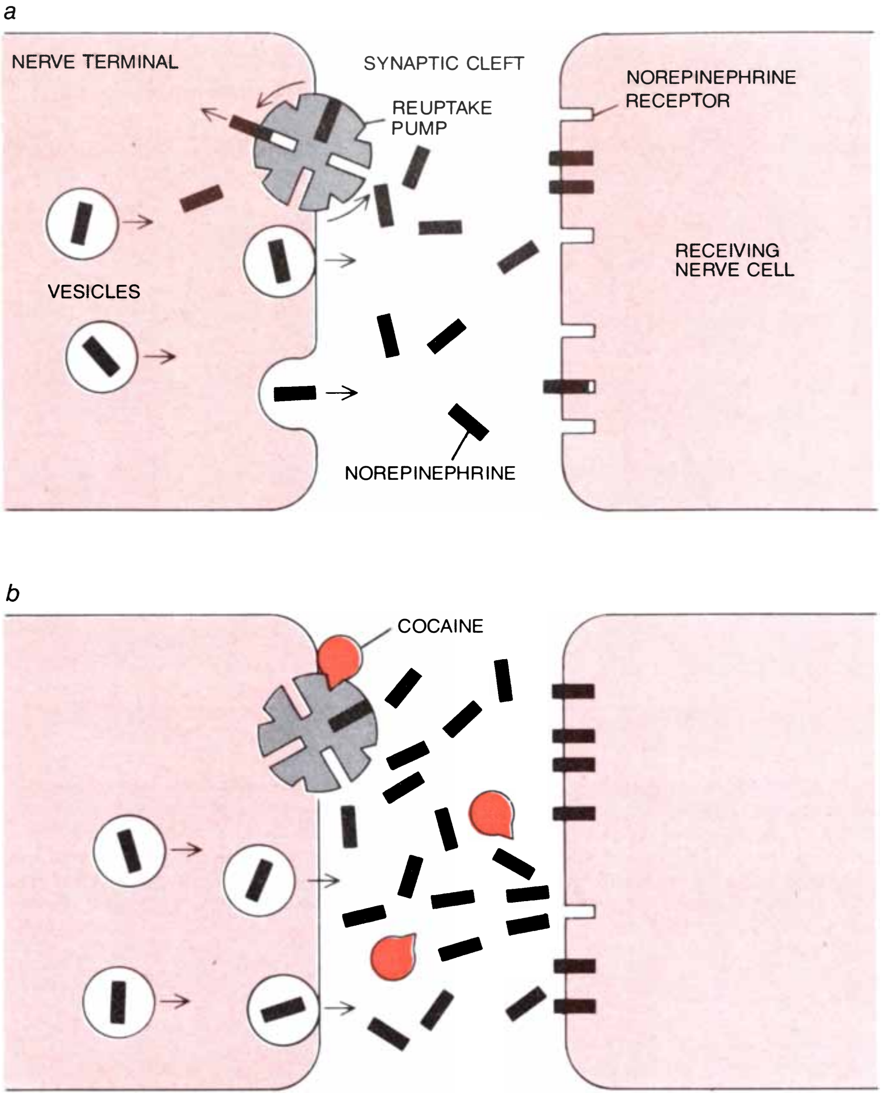
THE IZAAK WALTON LEAGUE OF AMERICA

Izaak Walton League
1800 North Kent Street
Arlington, Virginia 22209

small doses of cocaine may eventually lead to convulsions. Chronic cocaine use can lead to the abrupt onset of behavior that appears to be psychotic. Some investigators have argued that this may be the result of a "kindling effect," whereby a series of small doses sets off a response normally expected only from a much larger dose.

One of the main reasons for our interest in cocaine is its ability to reliably produce positive feelings in many people. In order to measure the psychological effects of cocaine we adopted several tests commonly employed in the study of mood and pain. In one such test the subject is asked to mark a point on a line that is labeled to represent six levels of euphoric or dysphoric feelings. Another test presents the subject with a list of adjectives, such as "lively," "suspicious," "anxious," "attentive," "sleepy" and "discouraged," together with a scale on which he can indicate the degree to which each adjective applies to his current feelings.

In experiments done in other laboratories subjects could not distinguish a small, 10-milligram dose of cocaine from a placebo. When 25 to 100 milligrams of cocaine was given intranasally, however, all the subjects reliably reported peak euphoria within 15 to 30 minutes. From 45 to 60 minutes after the 100-milligram dose two out of seven subjects experienced anxiety, depression, fatigue and a desire for more cocaine. There is often a crash or period of extreme discomfort after cocaine is smoked or injected, but it is less common after intranasal use. For the most part drugs are associated by the public with a single effect, but it is clear from both street and laboratory reports that the pleasurable effects of cocaine at relatively small doses are replaced by disturbing feelings at higher doses.



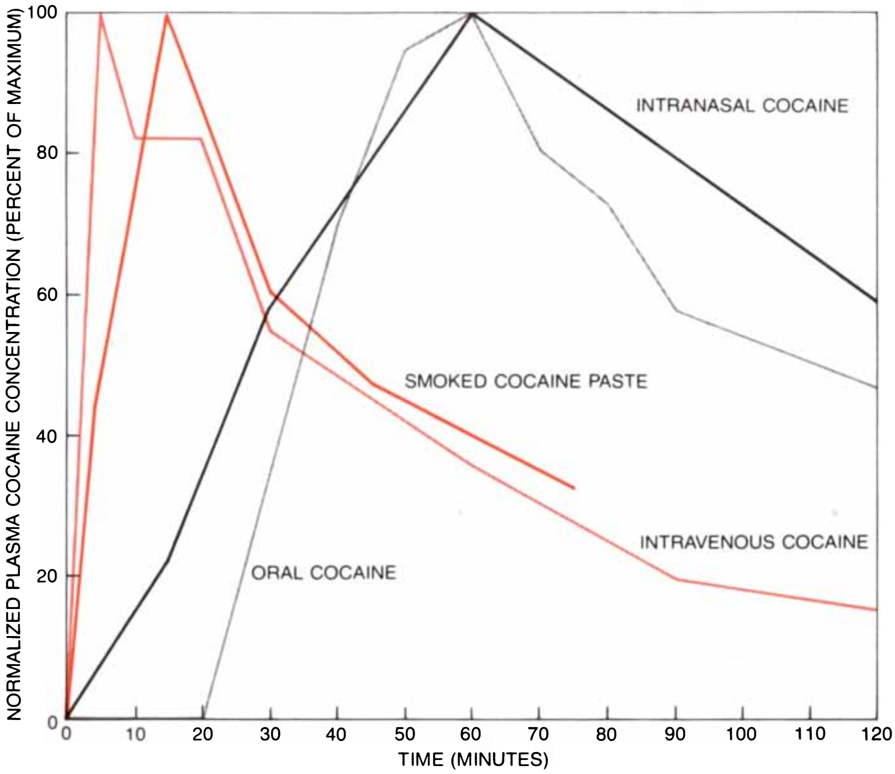
SYMPATHOMIMETIC ACTION of cocaine results from blocking the reuptake of neurotransmitters such as norepinephrine at synapses of the sympathetic nervous system. (The sympathetic nervous system controls such functions as heart rate and blood pressure.) When the molecules of the neurotransmitter are released from vesicles in the nerve terminal (*a*), the molecules cross the synaptic cleft and stimulate the succeeding nerve cell. Ordinarily some of the neurotransmitter molecules in the cleft are pumped back into the nerve that released them. In the presence of cocaine the action of the reuptake pump is blocked (*b*) and the stimulation by the neurotransmitter molecules increases as their concentration in the synaptic cleft builds up.

If the time course of the psychological effects of cocaine is compared with the time course of its concentration in the plasma of the blood, one finds that euphoria is most pronounced shortly before the plasma concentration has begun to fall. The euphoric effects disappear several hours before the plasma concentration returns to zero; a similar temporal relation between subjective feelings and plasma concentration can be seen in the effects of other drugs such as alcohol or the benzodiazepines. When cocaine is smoked or taken intravenously, discomfort can ensue when the plasma concentration is still half its peak level.

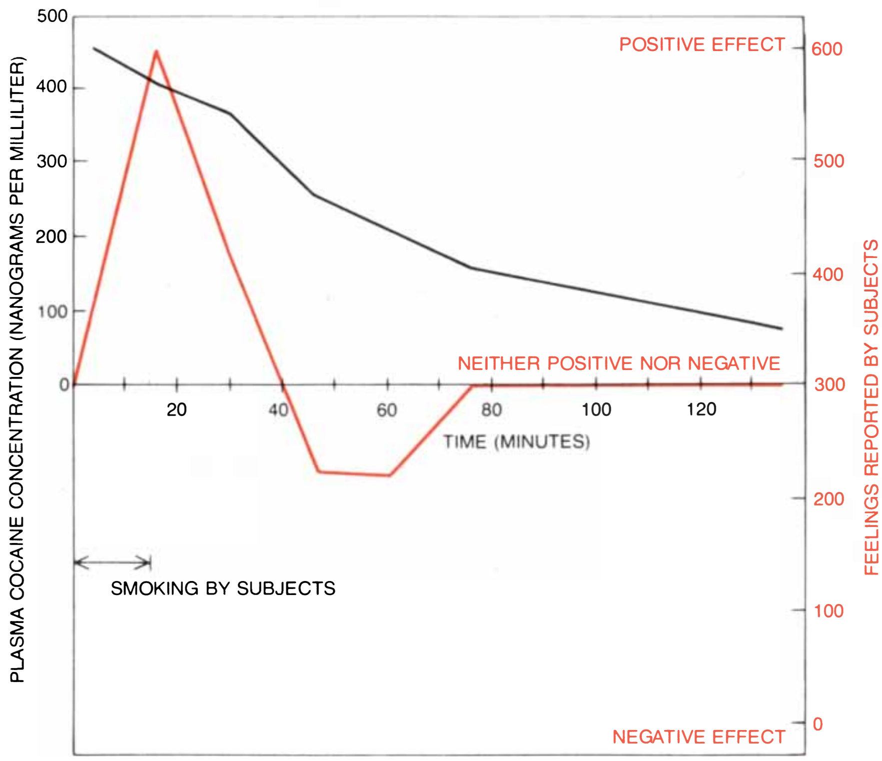
The lack of a match between the time course of plasma concentration and that of euphoric feeling suggests that the psychological effects of cocaine may be related to the rate of change of the plasma concentration rather than to its absolute level. Because plasma concentration changes much more abruptly when cocaine is smoked or injected than when it is snorted or taken by mouth, one would expect more potent psychological consequences from smoking or injecting if the rate of change is the controlling factor. Feelings reported by experimental subjects are in accord with this hypothesis. Alternatively, the rapid loss of euphoric feelings while the cocaine concentration in plasma remains high may result from decreased responsiveness of receptor sites once they have been occupied, so that there is the appearance of acute tolerance to the drug.

Another reason feelings of euphoria rise and fall so rapidly while the concentration of cocaine in the plasma remains high may be that cocaine is readily soluble in fatty tissue. Because of this solubility cocaine may be quickly taken up from the bloodstream into the brain; there is some evidence that the concentration of cocaine in the brain is much greater than the concentration in plasma. If the large initial concentration of cocaine in the plasma leads to a high concentration in the brain, subsequent equilibration would be achieved if the cocaine diffused out of the brain and into the plasma once again. The plasma concentration would thus remain high even after the cocaine began to flush out of the brain. Hence what is known about the relation between plasma concentration and euphoria is at least consistent with the hypothesis that the euphoria induced by cocaine is closely linked with its concentration in the brain.

The effects of a drug depend on the long-term behavior patterns of an individual with respect to the drug as well as on the discrete phenomena that can



TIME COURSE of cocaine concentration in the blood plasma is graphed for doses of cocaine taken into the body by four routes. The concentrations are normalized, so that all values are given as a percentage of the peak value. Once the peak concentration is reached cocaine has a half-life in the bloodstream of about an hour. Differences in the psychological effects of cocaine when it is administered by various routes are probably caused by differences in the time required for the plasma concentration to reach its peak value. Smoking or injecting cocaine leads to a faster rise in the plasma concentration of the drug than snorting or ingesting does.



DYSPHORIC FEELINGS followed the initial euphoria in experimental subjects who smoked cocaine paste, even though the concentration of cocaine in the plasma of the blood remained relatively high. The dysphoria is characterized by anxiety, depression, fatigue and a desire for more cocaine. The peak feelings for the subjects were probably reached shortly before the peak plasma concentration, but the first psychological measurements were made later than the plasma assay. Hence the temporal sequence of the peaks shown cannot be regarded as definitive.

be observed after a single dose. Chronic use of a drug can have both social and pharmacological consequences. If a person becomes totally involved in the acquisition and consumption of a drug, his entire life structure may change. When cocaine is taken regularly, it can cause sleeplessness and loss of appetite; after high doses or chronic consumption the user may experience an anxious paranoid state. Whether or not cocaine causes true psychosis remains an undecided question, but there is increasing evidence that hallucinations and paranoia can result if the drug is taken frequently in large doses.

Is cocaine addictive? The question appears to call for a simple answer, but a yes or no reply does not do justice to the tangle of medical definition, folk wisdom, legal classification and social recrimination that is summarized by the word "addictive." One longstanding medically accepted definition of the term is derived from the description of opiate effects. For a drug to be considered addictive a person must develop tolerance for it, in the sense that repeating the same dose causes a diminishing response. Moreover, the drug must lead to physical dependence, so that repeated doses are required to prevent the onset of a withdrawal syndrome.

According to this definition, cocaine is not addictive. Cocaine users can take the same dose every day and get the same effect. There are withdrawal signs detectable on the electroencephalogram and in sleep patterns, but they are quite undramatic when compared with the withdrawal syndromes associated with opiates, barbiturates or alcohol.

On the other hand, cocaine certainly is addicting within the broader sense of the term that has now been adopted by many pharmacologists. That is, cocaine is severely habit-forming. In this context the chronic consumption of cocaine through the nose must be distinguished from its regular injection and from free-base smoking. Dependence on intranasal cocaine manifests itself in a pattern of continued use while supplies are available and in simple abstention when supplies are lacking. The pattern of behavior is comparable to that experienced by many people with peanuts or potato chips. It may interfere with other activities of the individual, but it may be a source of enjoyment as well.

In contrast, the smoking or injection of cocaine can lead to almost continual consumption and drug-seeking behavior, destructive to personal competence and productivity. Free-base smoking is probably also dangerous to the respiratory system because it constricts the blood vessels of the lungs. The high price of street cocaine, together with the relatively large amounts of the substance necessary for injections or free-base smoking, undoubtedly function to limit the damage people might otherwise inflict on themselves.

A considerable bureaucratic constituency depends for its existence on the public belief that cocaine is unequivocally pernicious. The U.S. Drug Enforcement Administration employed 1,950 agents worldwide with an operating budget of $227 million during the fiscal year ending in September, 1981. In that fiscal year agents devoted 31 percent of their investigative man-hours to cocaine-related work. Operatives of the agency are active in nine South American countries, and in all these countries except Colombia (where marijuana trafficking is also seen as a problem) the main concern is cocaine.

The State Department, moreover, has exerted a great deal of diplomatic leverage in persuading the governments of these countries to adopt policies desired by the U.S. toward the growing of coca and toward the enforcement of laws pertaining to the cocaine trade. The U.S. has now signed treaties that commit Peru and Colombia to the eradication of coca, although no such agreements have been made with Bolivia. United Nations agreements to limit the increase of coca production in Peru and Bolivia have been tied to financial incentives, and they are accompanied by plans to cultivate substitute crops.

It is not clear whether the potential for the abuse of cocaine justifies the intensity of such efforts. Deaths from recreational cocaine use are rare. In Andean Indian societies blood-plasma levels comparable to those encountered among intranasal cocaine users are common, yet there is little evidence of physical harm from coca chewing. In such cultures the religious use of cocaine cannot be termed drug abuse, yet the zeal of eradication efforts may deprive the Indian of a traditional substance. The main threat to North American and European society from cocaine is the waste of human potential that could result from widespread consumption. Cocaine policy and regulations take little account of these conclusions, just as drug regulations in the past have been based neither on science nor on sense. We have legalized both alcohol and tobacco, which, although pleasurable to some people, are distinct hazards to both personal and public health.

One can learn from history that cocaine is a drug easily overused. One can learn from the present cocaine trade that if a demand is great enough, an industry will arise to meet it. Medically cocaine is a relatively safe drug, but in the hands of naive people it can lead to self-destructive behavior. Many of the questions surrounding the issue of cocaine can be answered by scientific investigation, but the final decisions about cocaine will be political and economic, not scientific.

THE AMATEUR SCIENTIST

Motors in which magnets attract other magnets in apparent perpetual motion

by Jearl Walker

Paul Monus of Cleveland State University builds delightful motors in which a magnet moves past a series of other magnets, sometimes apparently without end. Typically a small cylindrical magnet rolls along a track of magnets covered with a sheet of Plexiglas. The cylinder is propelled primarily by the interaction of the fields of the moving magnet and the stationary magnets, but the motion is maintained by the momentum of the moving magnet and on occasion by an electromagnet placed strategically near the track.

The motors are by-products of Monus' research on the efficiency of magnetically aided direct-current motors. In some of the by-product motors the aim is to have a cylindrical magnet move from a starting point to a definite stopping point. In others the aim is to have a motion initiated by the experimenter continue. My favorite motor is one where a cylindrical magnet rolls indefinitely in a circle over an array consisting of several permanent magnets and one electromagnet.

The simplest of Monus' motors has a linear track of 155 flat ceramic magnets. Each magnet is an inch long, 3/4 inch wide and 3/16 inch thick. The magnets lie in blocks of five across the width of the track, which consists of 31 such blocks. The blocks are sandwiched horizontally between two strips of soft iron, each an inch wide, 1/4 inch thick and 31 inches long. The blocks and the strips are held in place by horizontal brass bolts. (Brass has the advantage of not distorting the magnetic fields.) All the ceramic magnets are oriented in the same way, so that one of the iron strips is a south magnetic pole and the other is a north magnetic pole.

The ceramic magnets and the soft iron are in the class of materials termed ferromagnetic. Other types of materials display magnetic properties only when they are put into an existing magnetic field. A ferromagnetic material is magnetic on its own because of a special alignment of some of the electrons in its atoms. All electrons have intrinsic magnetic fields around them, almost as if they were tiny loops of current. Indeed, in investigations of magnetic materials the electrons are often visualized as being such small loops. To further visualize the magnetic field contributed by an electron the vector called a magnetic moment is assigned to it. All electrons have magnetic-moment vectors of the same size.

In materials that are not ferromagnetic the orientation of the vectors is random. Since the material is sure to have a great many electrons, each vector has a partner that points in the opposite direction. The electrons are paired off in this way in each atom. Thus each vector is canceled and the material has no net magnetic field. This cancellation of magnetic moments is the reason the human body is not magnetic in spite of its abundance of electrons.

The situation is different in a ferromagnetic material because of the mutual alignment of many of its electrons. Each atom has two electrons with no partner to cancel their magnetic moment. These atoms have a net magnetic moment. Regions where the magnetic moments of the atoms are aligned are called domains. Hence each domain has its own net magnetic moment.

A permanent magnet has a magnetic field (even if it is not in the presence of some other source of one) because many of the magnetic moments of the domains are aligned. If the moments of the domains are entirely aligned, the magnet is said to be saturated. If they are less than entirely aligned, the magnet has a weaker magnetic field. Monus' track is made of permanent magnets in which the alignment of the magnetic moments of the domains has generated a relatively strong magnetic field.

The domains in the strips of soft iron are less aligned. The iron is therefore said to be magnetically soft. When Monus attaches soft iron to the blocks of ceramic permanent magnets, the magnetic fields from the blocks aid in aligning the magnetic moments of the domains in the iron. The strips of iron then add to the magnetic field surrounding the track. They also help to smooth the transition in the magnetic field from one block of ceramic magnets to the next.

Over the track of magnets Monus fastens a flat sheet of Plexiglas. One end of the sheet is flush with the end of the track. The other end extends beyond the track and serves as a launching section for the cylindrical magnet. The sheet is three inches wide, 1/8 inch thick and 37 inches long. The overhang for the launching section is therefore about six inches.

A similar sheet of Plexiglas is mounted on the underside of the track; the ends are flush with those of the track. The entire structure is supported about two inches above a base made out of Plexiglas or wood. Other nonmagnetic materials could be substituted. The support is adjusted until the track is horizontal. (Monus also makes motors with tracks tilted from the horizontal, to which I shall be returning.)

The cylindrical magnet that is propelled by the magnetic field of the track is made out of three rings of ceramic magnet joined end to end; each ring is an inch in diameter and 1/4 inch thick. The rings are held together by a wood or plastic rod pushed through their central hole. They are positioned so that a north-pole face is adjacent to a south-pole one.

The roller is placed on the launching section of the Plexiglas cover with its magnetic field oriented opposite to the field from the track. Hence its south magnetic pole is on the same side as the track's north magnetic pole. When the roller is released, the magnetic field of the track prevents it from spinning around a vertical axis. The field also pulls the roller toward the track. The magnetic force on the roller and the roller's own momentum combine to cause it to roll along the track and to stay on the Plexiglas.

The horizontal propulsion is the result of the interaction of the roller's magnetic field with the change in the magnetic field through which the roller passes. Magnetic fields are often depicted with field lines. A stronger magnetic field is depicted with a denser collection of lines. The roller begins its trip in a region where the field lines are relatively far apart and is pulled into regions where the lines are progressively closer together. This gradient of field lines (representing the strength of the track's magnetic field) provides the force for the acceleration of the roller.

When the roller passes over the first block of ceramic magnets, it continues to accelerate as the magnetic field of the track propels it. At the midpoint of the track the propelling force diminishes and then vanishes. The roller does not stop, however, because of the momentum it has by the time it reaches the equilibrium point. Overshooting that point, it continues toward the far end of the track. The magnetic force on it now

NEW NISSAN STANZA

ONLY AT YOUR DATSUN DEALER.

32 MPG,* 5 PASSENGERS BIG.

Outstanding gas mileage, spine-tingling acceleration, room to pamper five adults. All in harmony, all in Nissan Stanza, the car everyone's trying to build.

SHAPED BY THE WIND TO RUN WITH IT.

From sloping nose to sharply angled windshield to contoured roofline, Nissan Stanza is aerodynamics at its slippery best.

A spirited new NAPS-X engine powers that wind-cheating wedge to a combination of acceleration and fuel economy that is virtually unprecedented in 5-passenger cars.

ADVANCED FRONT-WHEEL DRIVE.

Nissan Stanza's weight is distributed almost equally between front and rear wheels.

This balanced stance lightens steering effort and, combined with a precise rack-and-pinion system, lets you groove around curves as though on rails.

46	32
EPA EST. HWY	EPA EST. MPG

*EPA estimates for comparison, with standard 5-speed. Actual mpg may differ depending on speed, trip length and weather. Highway mpg will probably be less. California mpg lower.

YOU CAN CROSS YOUR LEGS IN THE BACK OF A NISSAN STANZA.

Both 2- and 4-door Nissan Stanza hatchbacks treat you and your luggage to room you'd never expect in an economy car: over 100 cubic feet, riding smoothly on a wide track, long wheelbase and

refined 4-wheel independent suspension.

Nissan has been building fuel-efficient cars since gasoline was 18¢ a gallon. But we're most proud of our quality. For we are driven to bring you more miles per gallon and more years per car.

DATSUN
WE ARE DRIVEN

© 1981 NMC–U.S.A.

Product of NISSAN

The $149[95] personal computer.

Introducing the Sinclair ZX81

If you're ever going to buy a personal computer, now is the time to do it.

The new Sinclair ZX81 is the most powerful, yet easy-to-use computer ever offered for anywhere near the price: only $149.95* completely assembled.

Don't let the price fool you. The ZX81 has just about everything you could ask for in a personal computer.

A breakthrough in personal computers

The ZX81 is a major advance over the original Sinclair ZX80—the world's largest selling personal computer and the first for under $200.

In fact, the ZX81's new 8K Extended BASIC offers features found only on computers costing two or three times as much.

Just look at what you get:

- Continuous display, including moving graphics
- Multi-dimensional string and numerical arrays
- Mathematical and scientific functions accurate to 8 decimal places
- Unique one-touch entry of key words like PRINT, RUN and LIST
- Automatic syntax error detection and easy editing
- Randomize function useful for both games and serious applications
- Built-in interface for ZX Printer
- 1K of memory expandable to 16K

The ZX81 is also very convenient to use. It hooks up to any television set to produce a clear 32-column by 24-line display. And you can use a regular cassette recorder to store and recall programs by name.

*Plus shipping and handling. Price includes connectors for TV and cassette, AC adaptor, and FREE manual.

If you already own a ZX80

The 8K Extended BASIC chip used in the ZX81 is available as a plug-in replacement for your ZX80 for only $39.95, plus shipping and handling—complete with new keyboard overlay and the ZX81 manual.

So in just a few minutes, with no special skills or tools required, you can upgrade your ZX80 to have all the powerful features of the ZX81. (You'll have everything except continuous display, but you can still use the PAUSE and SCROLL commands to get moving graphics.)

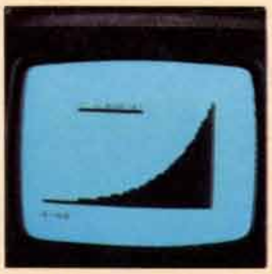
With the 8K BASIC chip, your ZX80 will also be equipped to use the ZX Printer and Sinclair software.

Order at no risk**

We'll give you 10 days to try out the ZX81. If you're not completely satisfied, just return it to Sinclair Research and we'll give you a full refund.

And if you have a problem with your ZX81, send it to Sinclair Research within 90 days and we'll repair or replace it at no charge.

**Does not apply to ZX81 kits.



NEW SOFTWARE: Sinclair has published pre-recorded programs on cassettes for your ZX81, or ZX80 with 8K BASIC. We're constantly coming out with new programs, so we'll send you our latest software catalog with your computer.

ZX PRINTER: The Sinclair ZX Printer will work with your ZX81, or ZX80 with 8K BASIC. It will be available in the near future and will cost less than $100.

16K MEMORY MODULE: Like any powerful, full fledged computer, the ZX81 is expandable. Sinclair's 16K memory module plugs right onto the back of your ZX81 (or ZX80, with or without 8K BASIC). Cost is $99.95, plus shipping and handling.

ZX81 MANUAL: The ZX81 comes with a comprehensive 164-page programming guide and operating manual designed for both beginners and experienced computer users. A $10.95 value, it's yours free with the ZX81.

The $99^{95} personal computer.

Introducing the ZX81 kit

If you really want to save money, and you enjoy building electronic kits, you can order the ZX81 in kit form for the incredible price of just $99.95.* It's the same, full-featured computer, only you put it together yourself. We'll send complete, easy-to-follow instructions on how you can assemble your ZX81 in just a few hours. All you have to supply is the soldering iron.

How to order

Sinclair Research is the world's largest manufacturer of personal computers.

The ZX81 represents the latest technology in microelectronics, and it picks up right where the ZX80 left off. Thousands are selling every week.

We urge you to place your order for the new ZX81 today. The sooner you order, the sooner you can start enjoying your own computer.

To order, simply call our toll free number, and use your MasterCard or VISA.

To order by mail, please use the coupon. And send your check or money order. We regret that we cannot accept purchase orders or C.O.D.'s.

CALL 800-543-3000. Ask for operator #509. In Ohio call 800-582-1364. In Canada call 513-729-4300. Ask for operator #509. Phones open 24 hours a day, 7 days a week. Have your MasterCard or VISA ready.

These numbers are for orders only. For information, you must write to Sinclair Research Ltd., 2 Sinclair Plaza, Nashua, NH 03061.

sinclair

AD CODE 03SA	PRICE†	QTY.	AMOUNT
ZX81	$149.95		
ZX81 Kit	99.95		
8K BASIC chip (for ZX80)	39.95		
16K Memory Module (for ZX81 or ZX80)	99.95		
Shipping and Handling	4.95		$4.95
		TOTAL	

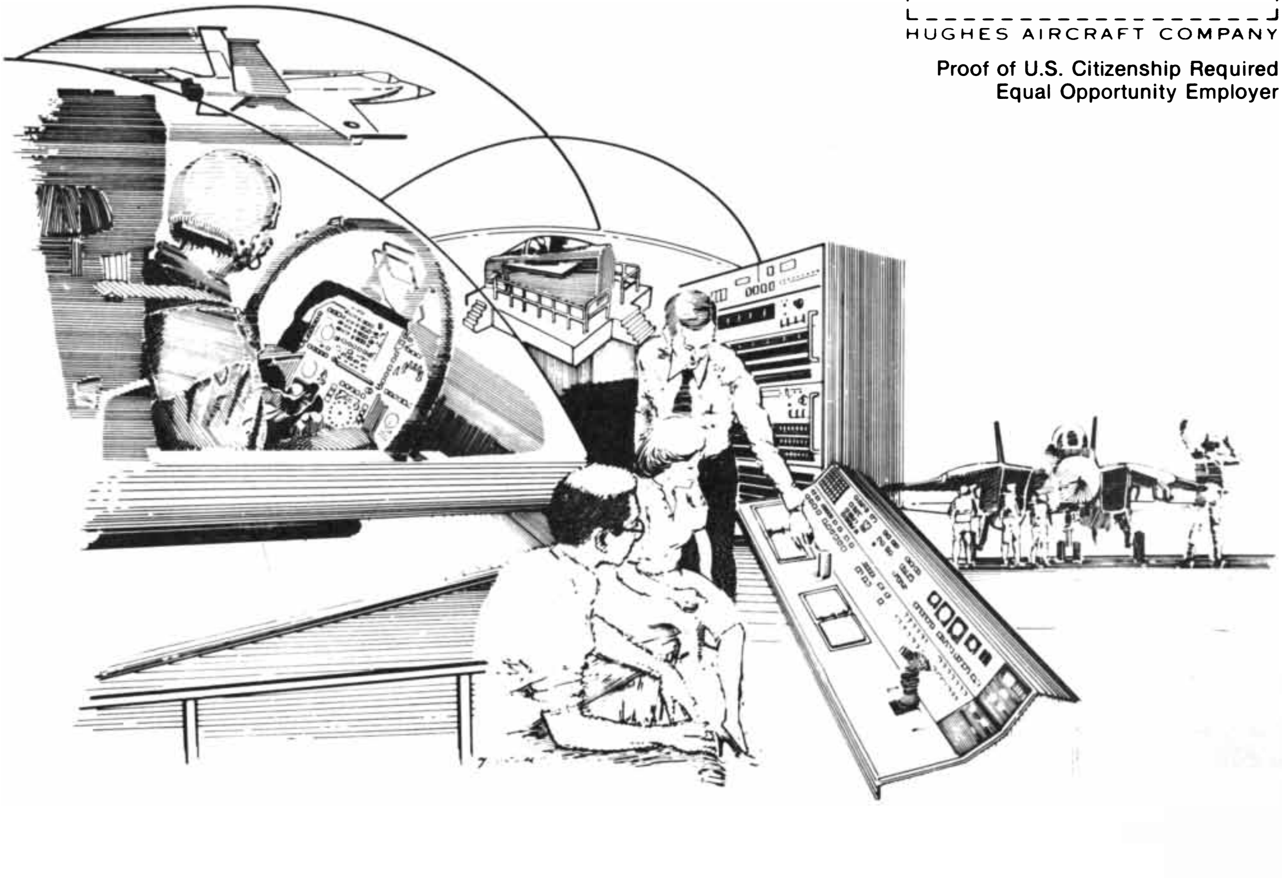
MAIL TO: Sinclair Research Ltd., One Sinclair Plaza, Nashua, NH 03061.

NAME ______________________

ADDRESS ______________________

CITY/STATE/ZIP ______________________

†U.S. Dollars



is opposite to its motion, and it gradually slows down.

Suppose the Plexiglas extends beyond the far end of the track. The roller moves out beyond the last block of magnets before it is brought to a stop by the magnetic force. Just how far it moves beyond the last block depends on how much energy it loses during its travel. If the loss is insignificant, the roller moves as far beyond the last block as it was ahead of the first block on launching. This behavior, of course, is due to the conservation of energy. The roller has a certain amount of potential energy when it is placed in the magnetic field of the track. As the track's magnetic field accelerates the roller the potential energy is gradually transformed into kinetic energy, although the total energy of the roller remains the same.

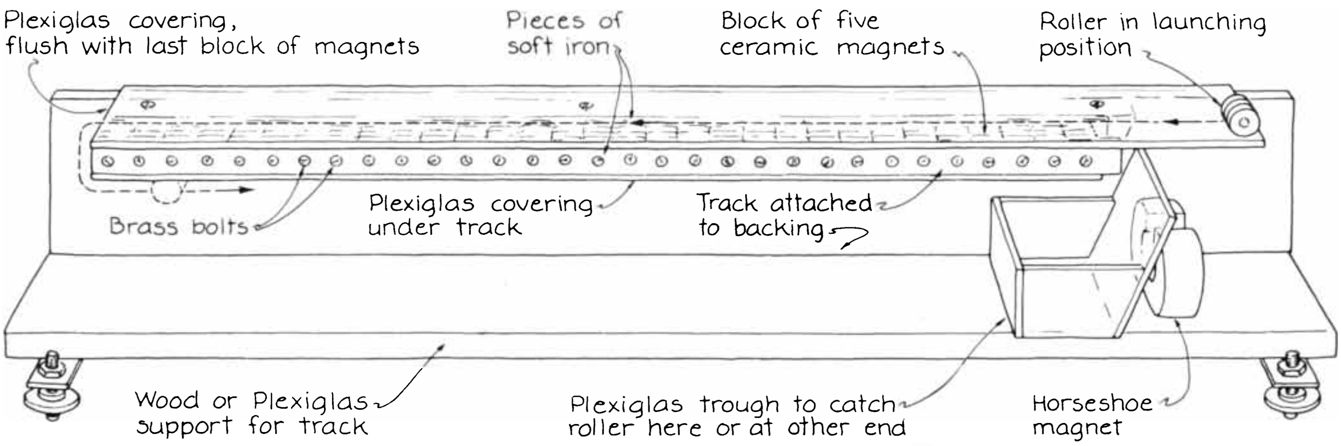
The kinetic energy is greatest and the roller's speed is highest as the roller passes through the equilibrium point of the track. Thereafter the magnetic field of the track gradually slows the roller and the kinetic energy is transformed back into potential energy. The transformation is complete when the distance between the roller and the last block of magnets is the same as that between the roller and the first block.

As a result of these forces the roller rolls back and forth along the track, continuously exchanging its potential energy for kinetic. The motion would continue forever if something did not sap the roller's energy. That something is mostly the tiny friction between the roller and the Plexiglas. The roller also loses energy because as it travels through the track's magnetic field it disrupts the existing alignment of the domains in the iron. In both cases the lost energy ends up as a tiny amount of heat, either in the surface of the Plexiglas and the roller (because of friction) or in the magnets (because of the jostling of adjacent domains).

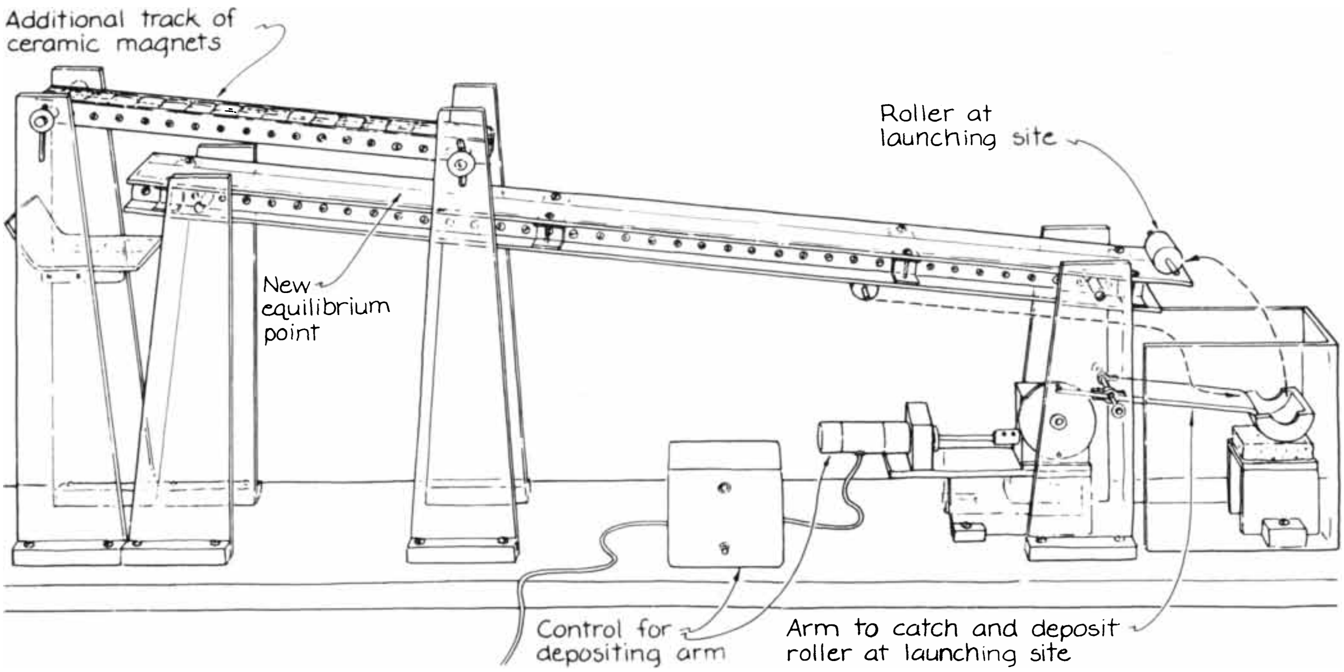
On each trip down the track the roller therefore has slightly less total energy. The distance by which it overshoots the end blocks decreases until it no longer overshoots at all. Finally the energy of motion is exhausted and the roller comes to a stop at the equilibrium point of the track.

Monus has devised a way to avoid a symmetrical back-and-forth motion of the roller. As I have indicated, the far end of the Plexiglas is flush with the end of the track. When the roller reaches this end, it drops over it. Now any one of several things can happen. At a low speed the roller is held by the magnets of the track, so that it goes around the end and is held against the sheet of Plexiglas on the bottom of the track. It now rolls back under the track toward the launching section. There the Plexiglas ends, and the roller may roll up around the end and bump into the underside of the launching section. This sudden loss of energy will probably stop the motion.

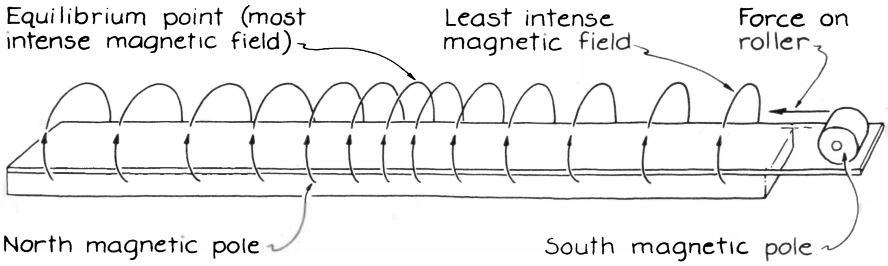
If the roller is moving fairly fast when



Paul Monus' arrangement for running a roller on a linear track



A double track for pulling the roller up a grade



The magnetic field along the linear track

it drops over the far end of the track, it falls onto the base that supports the entire apparatus. If the base is not too far below the track, the track's magnetic field still pulls the roller backward along the base. If the roller is too far below the track, it might not receive enough propulsive force to keep it moving. It would stop soon after falling.

Monus has added two more features to this basic structure. First, to protect the base from the impact of a falling roller he sometimes rigs up a Plexiglas container to catch the roller. The container can be at the far end of the track from the launching section, in which case it catches a fast roller that does not return along the bottom of the track. Alternatively the container can be at the near end of the track under the launching section, in which case it catches a slow roller that does return along the bottom of the track. Second, he has positioned a horseshoe permanent magnet on the base just under the launching section, partly to help the roller returning along the bottom of the track fall off at the end of its run.

Monus has experimented with this additional magnet in order to establish the right initial speed for the roller. He has also experimented with the initial distance between the roller and the first block of ceramic magnets. A greater distance means that the roller goes down the track too fast and falls off without returning. A lesser distance means that the roller rounds the far end and returns to the launching end along the underside of the track. A small distance means that it returns on the top side.

This type of simple magnetic motor need not be horizontal. When the far end of the track is elevated, progressively larger accelerations are needed in the initial launching to get the roller to the far end. Monus has designed another kind of motor to help the roller climb the tilted track. Above the far section of the track he has mounted another track of ceramic magnets, made the same way the main track is and having the same orientation. The function of the second track is to shift the equilibrium point away from the midpoint of the main track toward the far end. As a result the roller accelerates longer on the main track and goes farther before it begins to slow down. It can now climb a modest slope.

Monus says the positioning of the additional track is critical. He mounts it with brass bolts that run horizontally through slots in the supporting pillars of Plexiglas. By adjusting the position of the bolts in the slots he finds the optimum separation between the two tracks and the best angle for the additional track.

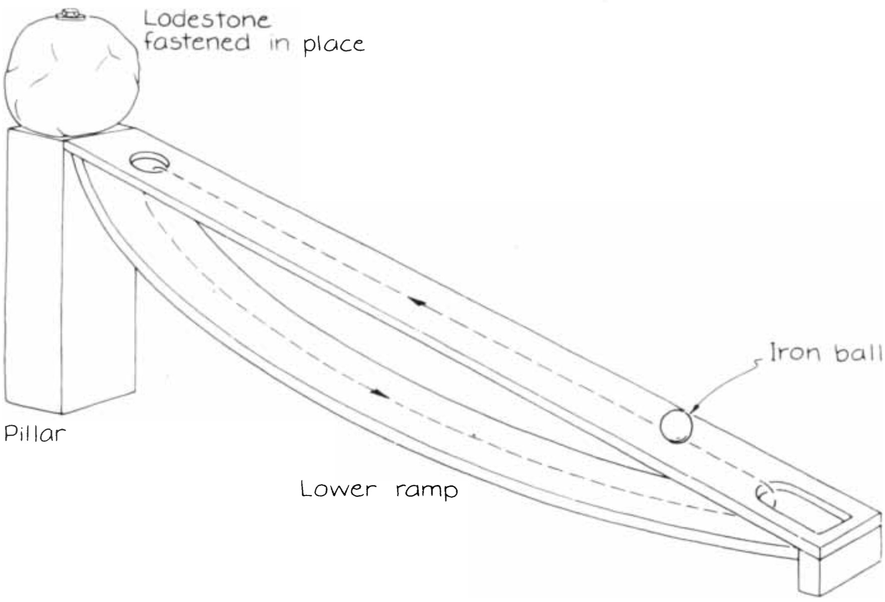
When everything is properly adjusted, the roller rounds the far end of the tilted main track and returns along the underside to the lower end. As the roller descends it gains speed. In this motor the Plexiglas sheet mounted under the track extends slightly beyond the end of the track. Because of this additional length and because of the gain in speed from the descent the roller falls off the end of the track. Here there is an apparatus to catch the roller and return it to its starting position.

The roller is caught in a Plexiglas bucket covered with plastic foam to protect it from the impact of the falling roller. The bucket is mounted on a pivoted arm actuated by an electric motor. When the roller falls into the bucket and depresses it, the arm is rotated upward, depositing the roller on the launching section. The entire journey up the tilted track and down along the underside is repeated. The motion continues as long as the arm driven by the electric motor redeposits the roller on the launching section.

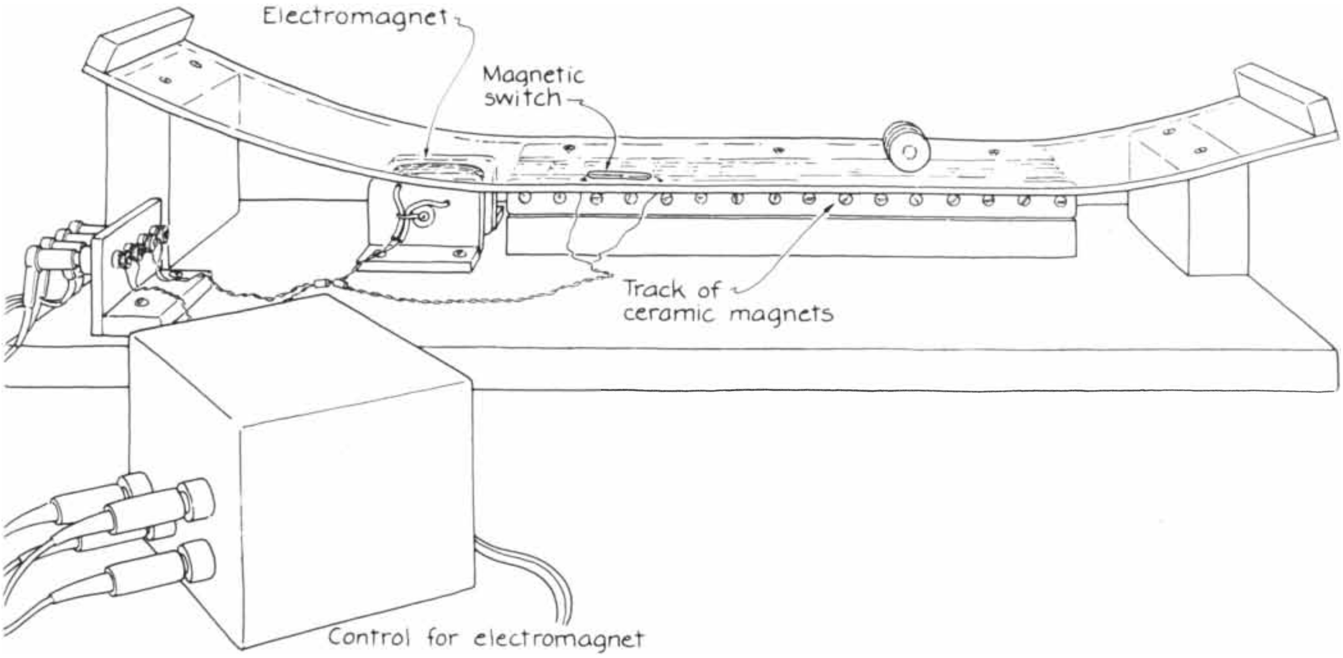
This kind of apparatus, like other motors designed by Monus, immediately suggests "perpetual motion." Actually, of course, the motion continues only because the motor-driven arm lifts the roller back to its launching position. Continuous cycling of the roller could not be achieved without such an input of energy.

In 1570 Johannes Taisnierus, a Jesuit priest, believed a similar apparatus he had designed would run perpetually. In his scheme a lodestone (a piece of magnetic mineral) is mounted at the top of a pillar to which a ramp runs from the ground. An iron ball is placed at the bottom of the ramp. If the magnetism of the lodestone is strong enough, it pulls the ball up the ramp. Near the top of the ramp the ball comes to a hole, falls through it and is returned by another ramp to its starting point. Taisnierus believed the cycle would be repeated forever. Of course, even if the lodestone were strong enough, frictional losses and others would soon bring the iron ball to a halt.

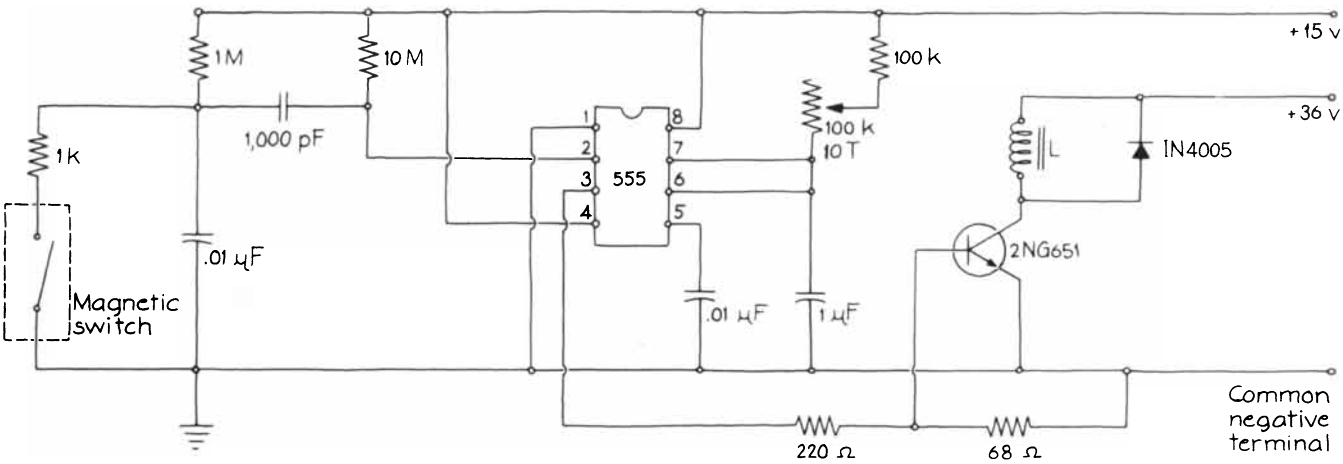
Monus has investigated the losses in his type of motor with a horizontal track of magnets to which he has added Plexiglas curved upward at both ends. At one end of the track he placed an electromagnet. Along the track he mounted a magnetic switch that would send current to the electromagnet. When the roller passes the switch, there is a brief delay and then the electromagnet is turned on. Therefore the electromagnet is energized just as the roller approaches the end of the track. Although the magnetic



How Johannes Taisnierus envisioned "perpetual motion" in the 16th century



The magnetic switch and electromagnet that make up the energy lost by the roller



Circuitry of the electromagnet

field of the track and the pull of gravity are both acting to slow the roller at that point, the pull of the electromagnet is strong enough to keep the roller moving up the track.

Without this additional pull the roller would merely roll back and forth over the track until it had lost all its kinetic energy. The purpose of the electromagnet is of course to supply the additional energy on each cycle. Monus adjusts the amount of current available to the electromagnet to match the energy the roller loses during a cycle. As a result the roller keeps going indefinitely. By measuring the current to the electromagnet he measures the losses of the system.

The circuit for the magnetic switch and the electromagnet is shown in the lower illustration above. Two voltage sources (of 15 and 36 volts respectively) are needed. Their negative terminals are connected to a common ground. The potentiometer is critical to the success of the circuit. Monus has installed one that has 10 turns (it is identified by the *10 T* in the diagram). This resistor controls the electromagnet. If the field from the electromagnet is too weak, the motion of the roller soon damps out. If the field is too strong, the roller is pulled so vigorously by the electromagnet that it bangs into the stop at the end of the track.

The most amusing of Monus' motors is one in which the roller travels over a horizontal track of permanent magnets laid out in the approximate shape of a horseshoe. The track consists of three main sections. The first is a short, straight length in which the magnets are mounted as they are in the linear motors I have already described. The second section is semicircular. The third section is an electromagnet triggered by a magnetic switch on the semicircular section. Over all these magnets is a sheet of Plexiglas.

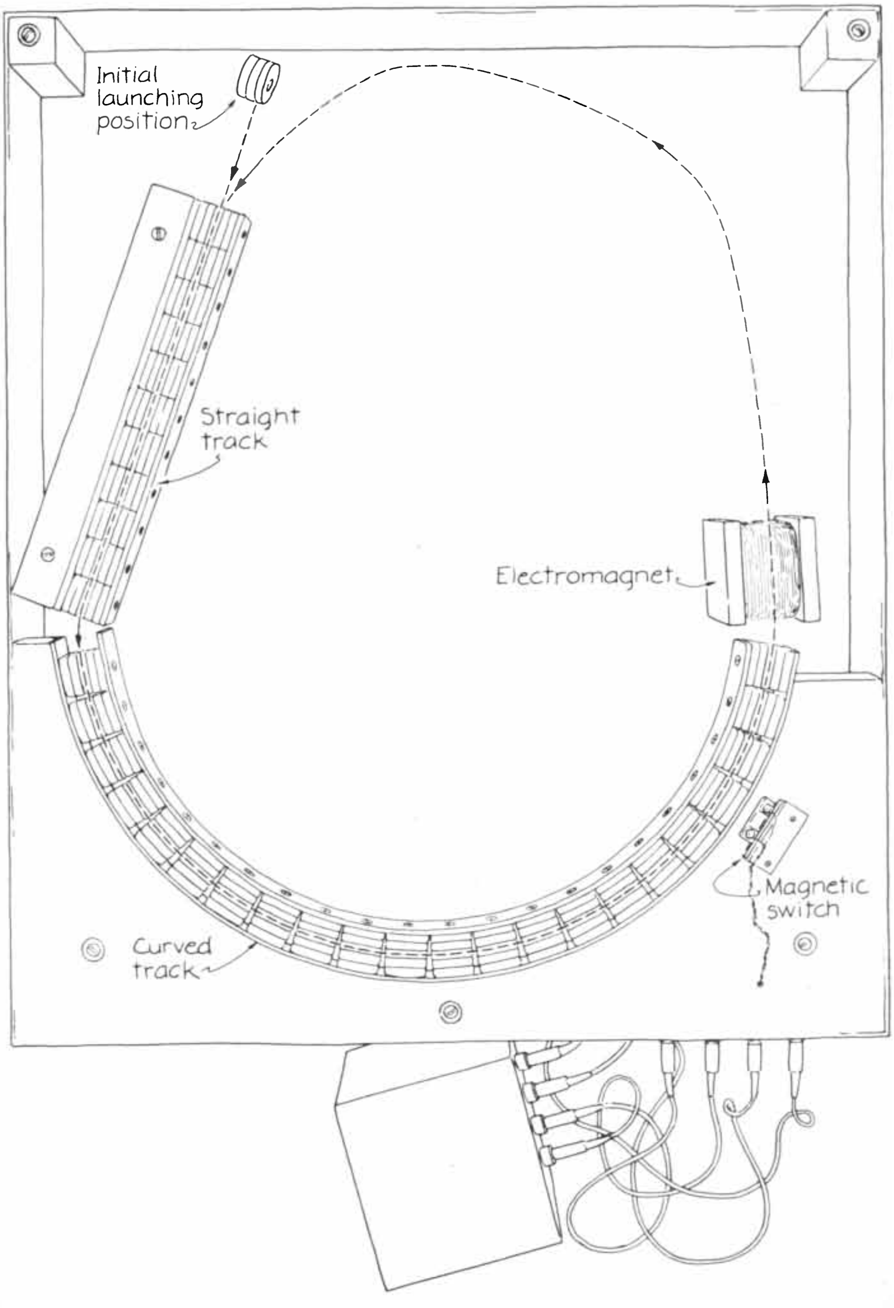
The roller is released a few inches from the starting end of the first section. As in the linear motors it accelerates onto the track, overshoots the equilibrium point and continues past the last block of magnets in the straight section. There it is pulled onto the curved track. Again it accelerates toward the equilibrium point of the section, overshoots it and emerges from the other end of the section.

Before the roller has a chance to be pulled back onto the curved track it rolls over the electromagnet triggered by a magnetic switch on the track. The pulse of current to the electromagnet creates a magnetic field and by that means pulls the roller toward the electromagnet. This pull also supplies fresh energy to the roller, adjusting for the energy it lost as it traveled.

The roller overshoots the electromagnet, which loses its field as the pulse of the current shuts off, and continues on out across the sheet of Plexiglas. As it does so it is pulled over toward the first section of track where it began and is accelerated onto that track again. The entire cycle is repeated. I have watched the thing for long periods because the motion is mesmerizing. The roller accelerates by an invisible force, whips around a curved track, is set free and then with a jerk is recaptured by the invisible force.

The most mystifying of Monus' creations is a Plexiglas wheel that spins indefinitely. On its rim are five blocks consisting of two ceramic magnets each. The blocks are evenly spaced along the perimeter. The wheel rotates on a horizontal axle supported by two Plexiglas pillars. Parallel to the plane of the wheel and immediately adjacent to it is a larger stationary sheet of Plexiglas. On this sheet is a curved row of ceramic magnets at a slightly larger radius from the center than the magnets on the rim of the wheel. At a smaller radius is a shorter curved row of magnets.

Once the wheel has been set spinning one of the rim magnets at the bottom of the wheel is pulled by the longer row of magnets on the stationary sheet. This rim magnet moves up to and then along the curved row, further turning the wheel. The magnetic force on the rim magnet diminishes once the magnet



A roller that goes on and on, thanks to the electromagnet

SCIENCE/SCOPE

The Advanced Medium-Range Air-to-Air Missile is undergoing full-scale development at Hughes after a 33-month evaluation of two competing designs. It employs the latest missile technology and is more capable, more reliable, and easier to maintain. The contract awarded to Hughes by the U.S. Air Force calls for 94 test missiles to be built, with options for 924 operational missiles and future options for developing second-source or follow-on missile production. The radar-guided AMRAAM will replace the AIM-7 Sparrow now in use. It will be carried by Air Force F-15 and F-16 fighters and Navy F-14 and F/A-18 aircraft. Ultimately 20,000 missiles are expected to be built for the Air Force and Navy.

An advanced electron-beam lithography system is being developed at Hughes to make Very High Speed Integrated Circuits (VHSIC), the "super chips" that will give military systems a tenfold increase in data processing capability. The high-speed lithography system will focus beams of electrons to "write" circuit patterns with submicron dimensions. These patterns will be converted into transistors and interconnects smaller than any now in production. VHSIC chips will be faster, more reliable, use less power, and have more memory than ordinary chips. They will be used in the air, at sea, on the ground, and in space.

Better windows for infrared sensors may be forthcoming after more research into a new fabrication process. Hughes scientists have made discs of fluorohafnate and fluorozirconate glasses by pressing glass pieces under low pressure (1024 psi) and high temperature (340°C). The process offers two important benefits. First, infrared glass compositions, which tend to crystallize when large batches are cooled from the melt, can now be formed into large optical elements up to 30 centimeters in diameter. Second, because the discs are cast into their final form, they have neither surface strains due to grinding nor polishing impurities, both of which reduce infrared transparency.

An ultramodern facility spanning 1.75 million square feet will be the showcase where outstanding Hughes engineering will combine with advanced manufacturing techniques and production processes. Our complex is nearly completed, so we're looking for experienced and graduating engineers to work on such programs as: infrared thermal imaging systems, laser rangefinders and designators, and missile launching and guidance systems. Send your resume to Dan O'Daly, Hughes Electro-Optical and Data Systems Manufacturing, Professional Employment, P.O. Box 924, Dept. SSA, El Segundo, CA 90245. Equal opportunity employer.

A California bank is among the pioneer users of business communications via satellite for transmitting voice, video, and high-speed data communications. Crocker National Bank will obtain the service from Satellite Business Systems beginning in July. The communications network will enable computers to exchange information much faster than is presently possible. It also will cut the bank's travel expenses because conferences can be held between cities via television. The SBS satellites were built by Hughes.

Creating a new world with electronics

HUGHES

HUGHES AIRCRAFT COMPANY
CULVER CITY, CALIFORNIA 90230

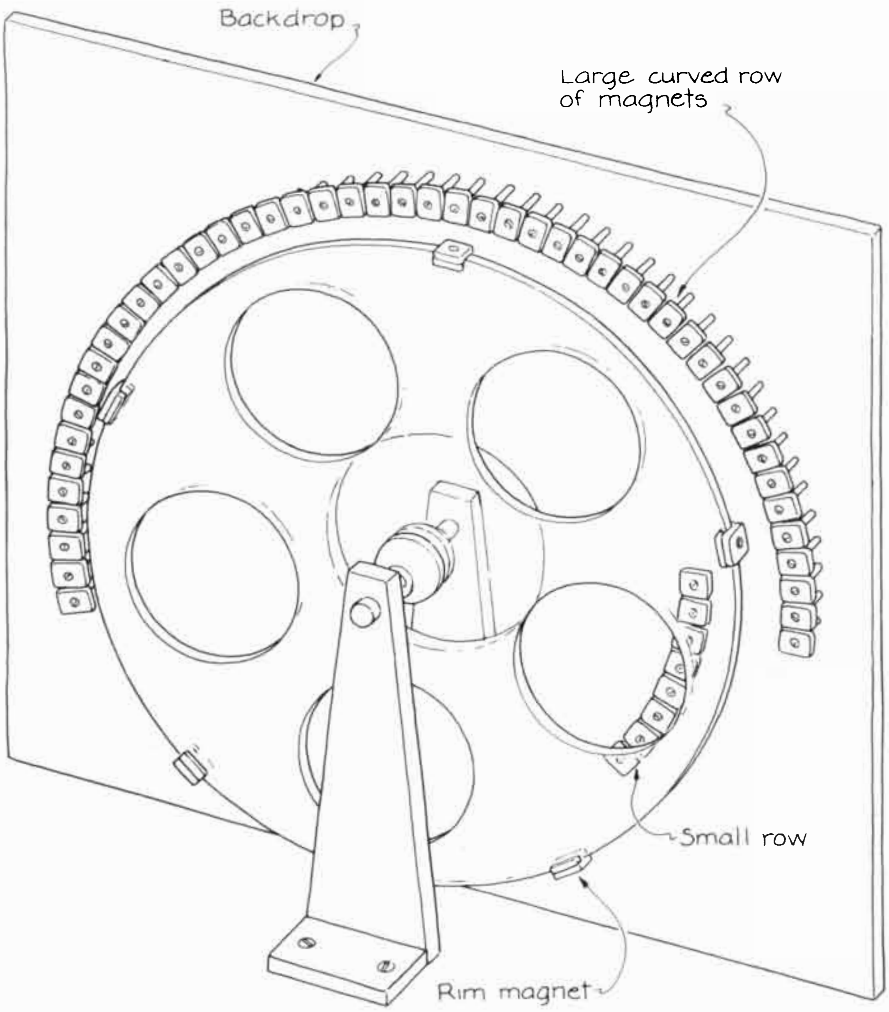
(213) 670-1515 EXTENSION 5964

has gone along part of the stationary row, but by then another rim magnet is being pulled along the row. Further impetus is given to the wheel when a rim magnet nears the end of the longer curved row and approaches the shorter curved row. Just as in the linear motor the two rows of magnets pull the moving magnet along. The rim magnet passes through the equilibrium point of the two rows and continues on past it.

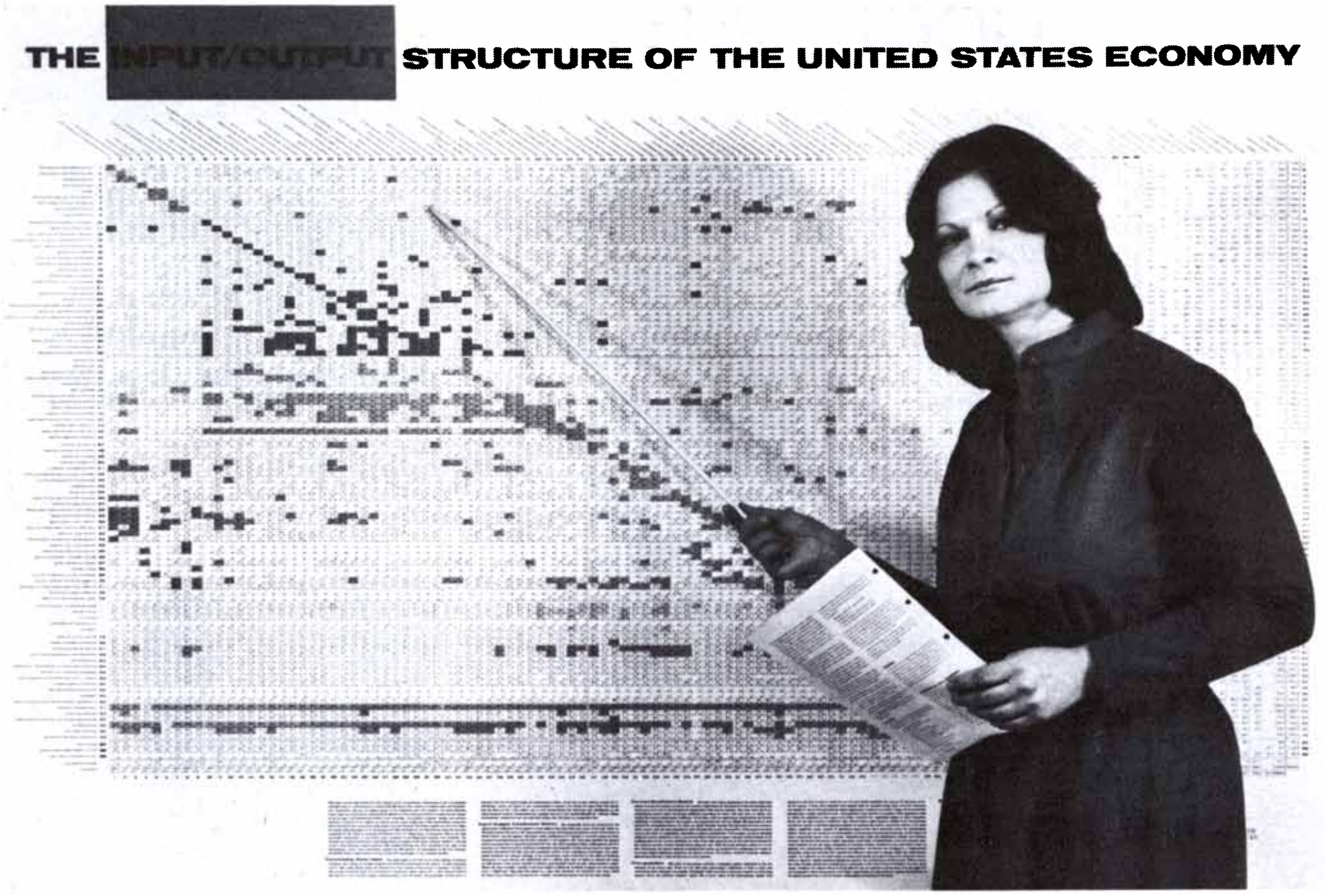
Although the wheel continues to turn indefinitely without visible assistance other than the magnets, its perpetual motion is an illusion. An electromagnet is hidden in the wood base of the machine. As in some of the other motors a magnetic switch triggers a pulse of current to the electromagnet. The electromagnet then pulls on a rim magnet as it emerges from the space between the two stationary curved rows of magnets. The pull adds energy to the wheel, compensating for its small energy losses.

The motor looks even more like a perpetual-motion machine because it has no telltale power line going to a wall plug. The power supply is a capacitor that stores enough energy from the electromagnetic fields already filling the room from the house current. Their energy is picked up by an antenna also concealed in the base; the oscillating current induced in the antenna is rectified and supplied to the capacitor. After a while the capacitor holds enough energy to generate a modest magnetic field in the electromagnet whenever the magnetic switch triggers it.

Monus bought his ceramic magnets from the Permag Magnetics Corporation, 2960 South Avenue, Toledo, Ohio 43609. The magnets are Indiana General brands, types Indox 1-F-1201 for the flat magnets and Indox 1-F-1403 for the cylindrical ones. Many other types and brands of permanent magnets can be substituted, although you might have to experiment with the size of the motors. Monus' magnetic switches came from Radio Shack: minireed switch 275–1610. Other brands will serve. If you would like detailed information and drawings describing Monus' motors, write to him at Box 421, Willoughby, Ohio 44094. The booklet costs $9.99 plus $1.50 for postage and handling. (Ohio residents should add the appropriate sales tax.) Monus can also send parts and building kits.



An arrangement for a wheel made to rotate by magnets



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__________

*Charts shipped international require one fold. 3 82

Make check payable to SCIENTIFIC AMERICAN. Please add sales tax for California, Illinois, Massachusetts, Michigan, Ohio and New York.

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

METAMAGICAL THEMAS

BIOCHEMISTRY. Albert L. Lehninger. H. S. Worth, 1970.

BIOCHEMISTRY. Lubert Stryer. W. H. Freeman and Company, 1975.

THE MOLECULAR BIOLOGY OF THE GENE. James D. Watson. W. A. Benjamin, 1976.

ILLEGAL MIGRATION

MANPOWER AND IMMIGRATION POLICIES IN THE UNITED STATES. David S. North and Allen Le Bel. National Commission for Manpower Policy, Special Report No. 20, 1978.

U.S. IMMIGRATION: A POLICY ANALYSIS. Charles B. Keely. The Population Council, 1979.

LABOR, CLASS AND THE INTERNATIONAL SYSTEM. Alejandro Portes and John Walton. Academic Press, 1981.

THE SEARCH FOR INTERMEDIATE VECTOR BOSONS

UNIFIED THEORIES OF ELEMENTARY-PARTICLE INTERACTION. Steven Weinberg in *Scientific American,* Vol. 231, No. 1, pages 50–59; July, 1974.

THE DETECTION OF NEUTRAL WEAK CURRENTS. David B. Cline, Alfred K. Mann and Carlo Rubbia in *Scientific American,* Vol. 231, No. 6, pages 108–119; December, 1974.

THE NEXT GENERATION OF PARTICLE ACCELERATORS. Robert R. Wilson in *Scientific American,* Vol. 242, No. 1, pages 42–57; January, 1980.

ANTIPROTON-PROTON COLLIDERS AND INTERMEDIATE VECTOR BOSONS. David Cline and Carlo Rubbia in *Physics Today,* Vol. 33, No. 8, pages 44–52; August, 1980.

RINGS OF THE GULF STREAM

THE BIOLOGY OF COLD-CORE RINGS. Peter Wiebe in *Oceanus,* Vol. 19, No. 3, pages 69–76; Spring, 1976.

DISTRIBUTION AND MOVEMENT OF GULF STREAM RINGS. David Y. Lai and Philip L. Richardson in *The Journal of Physical Oceanography,* Vol. 7, No. 5, pages 670–683; September, 1977.

PATTERNS AND PROCESSES IN THE TIME-SPACE SCALES OF PLANKTON DISTRIBUTIONS. L. R. Haury, J. A. McGowan and P. H. Wiebe in *Spatial Pattern in Plankton Communities: NATO Conference on Marine Biology,* edited by John H. Steele. Plenum Press, 1978.

GULF STREAM COLD-CORE RINGS: THEIR PHYSICS, CHEMISTRY, AND BIOLOGY. The Ring Group in *Science,* Vol. 212, No. 4499, pages 1091–1100; June 5, 1981.

ONCOGENES

CANCER: SCIENCE AND SOCIETY. John Cairns. W. H. Freeman and Company, 1978.

THE MOLECULAR BIOLOGY OF RNA TUMOR VIRUSES: A PHYSICIAN'S GUIDE. J. Michael Bishop in *The New England Journal of Medicine,* Vol. 303, No. 12, pages 675–682; September 18, 1980.

PROTEINS PHOSPHORYLATED BY THE RSV TRANSFORMING FUNCTION. T. Hunter in *Cell,* Vol. 22, No. 3, pages 647–648; December, 1980.

ENEMIES WITHIN: THE GENESIS OF RETROVIRUS ONCOGENES. J. Michael Bishop in *Cell,* Vol. 23, No. 1, pages 5–6; January, 1981.

LASER APPLICATIONS IN MANUFACTURING

CO_2 LASERS: EFFECTS AND APPLICATIONS. W. W. Duley. Academic Press, 1976.

HIGH POWER LASER WELDING—1978. Conrad M. Banas in *Optical Engineering,* Vol. 17, No. 3, page 210; May–June, 1978.

BASIC CHARACTERISTICS OF LARGE OUTPUT HIGH-ENERGY DENSITY HEAT SOURCES. Y. Arata in *Proceedings of the International Laser Processing Conference,* Anaheim, Calif., November 16–17, 1981.

SUPERCLUSTERS AND VOIDS IN THE DISTRIBUTION OF GALAXIES

THE COMA/A1367 SUPERCLUSTER AND ITS ENVIRONS. Stephen A. Gregory and Laird A. Thompson in *The Astrophysical Journal,* Vol. 222, No. 3, pages 784–799; June 15, 1978.

THE COSMIC TAPESTRY. Guido Chincarini and Herbert J. Rood in *Sky and Telescope,* Vol. 59, No. 5, pages 364–371; May, 1980.

THE PERSEUS SUPERCLUSTER. Stephen A. Gregory, Laird A. Thompson and William G. Tifft in *The Astrophysical Journal,* Vol. 243, No. 2, Part 1, pages 411–426; January 15, 1981.

THE INFRARED "VISION" OF SNAKES

RATTLESNAKES: THEIR HABITS, LIFE HISTORIES AND INFLUENCE ON MANKIND. Laurence M. Klauber. The University of California Press, 1972.

THERMORECEPTORS IN SNAKES. P. H. Hartline in *Electroreceptors and Other Receptors in Lower Vertebrates,* edited by A. Fessard. Springer-Verlag, 1974.

SPATIAL PROPERTIES OF PRIMARY INFRARED SENSORY NEURONS IN CROTALIDAE. John A. DeSalvo and Peter H. Hartline in *Brain Research,* Vol. 142, No. 2, pages 338–342; February 24, 1978.

ANATOMICAL AND PHYSIOLOGICAL LOCALIZATION OF VISUAL AND INFRARED CELL LAYERS IN THE TECTUM OF PIT VIPERS. Leonard Kass, Michael S. Loop and Peter H. Hartline in *The Journal of Comparative Neurology,* Vol. 182, No. 4, pages 811–820; December 15, 1978.

CONNECTIONS OF THE TECTUM OF THE RATTLESNAKE *CROTALUS VIRIDIS:* AN HRP STUDY. Edward R. Gruberg, Earl Kicliter, Eric A. Newman, Leonard Kass and Peter H. Hartline in *The Journal of Comparative Neurology,* Vol. 188, No. 1, pages 31–41; November 1, 1979.

SPATIAL SHARPENING BY SECOND-ORDER TRIGEMINAL NEURONS IN CROTALINE INFRARED SYSTEM. L. R. Stanford and P. H. Hartline in *Brain Research,* Vol. 185, No. 1, pages 115–123; March 3, 1980.

COCAINE

THE AMERICAN DISEASE: ORIGINS OF NARCOTIC CONTROL. David F. Musto. Yale University Press, 1973.

COCAINE PAPERS: SIGMUND FREUD. Edited with an introduction by Robert Byck. New American Library, 1975.

COCAINE, 1977: NATIONAL INSTITUTE ON DRUG ABUSE MONOGRAPH NO. 13. Edited by Robert C. Petersen and Richard C. Stillman. Department of Health, Education, and Welfare, 1977.

COCAINE 1980: PROCEEDINGS OF INTERNATIONAL SEMINAR ON COCA AND COCAINE. Edited by F. Raul Jeri. Pacific Press, Lima, Peru, 1980.

DRUG ADDICTION AND DRUG ABUSE. Jerome H. Jaffe in *The Pharmacological Basis of Therapeutics,* edited by Alfred Goodman Gilman, Louis S. Goodman and Alfred Gilman. Macmillan Publishing Co., Inc., 1980.

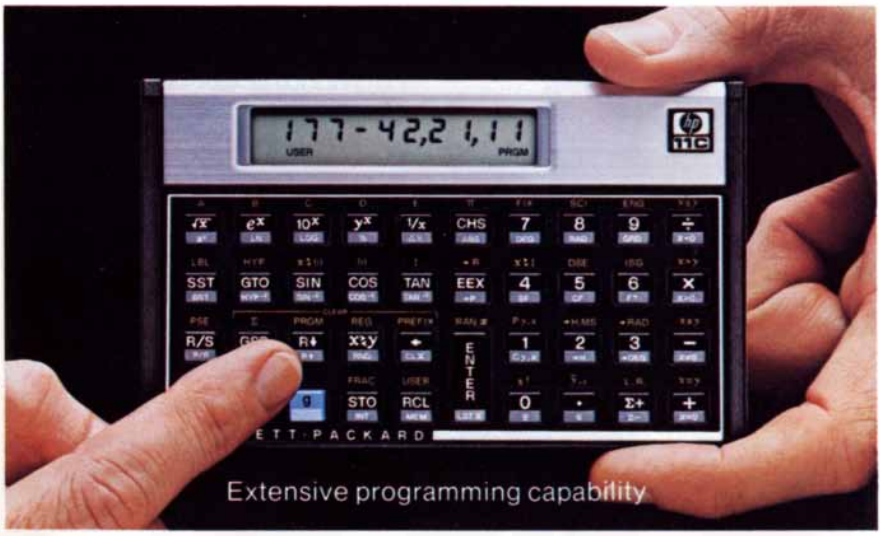
LOCAL ANESTHETICS. J. Murdoch Ritchie and Nicholas M. Greene in *The Pharmacological Basis of Therapeutics,* edited by Alfred Goodman Gilman, Louis S. Goodman and Alfred Gilman. Macmillan Publishing Co., Inc., 1980.

COCA AND COCAINE—1981: A SPECIAL ISSUE. Edited by L. Rivier and J. G. Bruhn in *Journal of Ethnopharmacology,* Vol. 3, Nos. 2 and 3; March/May, 1981.

THE AMATEUR SCIENTIST

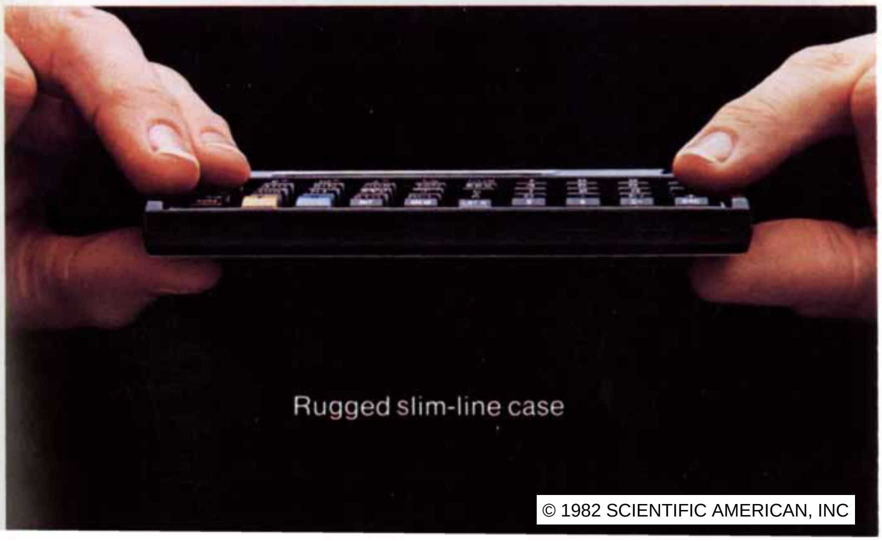
MODERN MAGNETISM. L. F. Bates. Cambridge University Press, 1963.

Introducing the new HP-11C. The whole is greater than the sum of its parts.



Extensive programming capability

Horizontal design with LCD display



Rugged slim-line case

Leave it to Hewlett-Packard to create an ingenious combination of built-in pre-programmed functions, a 200-line memory, and computer-like programming, including subroutines, indirect addressing, flags, and insert/delete editing—in a new, rugged, slim-line case.

And that's not all. The HP-11C's Continuous Memory remembers your programs and data even when the calculator is turned off. The calculator is powered by standard, disposable long-life batteries, making it both portable and convenient to operate.

Sure, other calculators may offer you some of these features. But Hewlett-Packard gives you all of them. The new HP-11C, built to the exacting standards of quality and reliability that you've come to expect from Hewlett-Packard.

To see the HP-11C in action, visit your nearest Hewlett-Packard dealer. For details and the address of the dealer in your area, call toll free: **800-547-3400,** Dept. 236B, except Hawaii & Alaska. In Oregon, call 758-1010. Or write Hewlett-Packard, Corvallis, OR 97330, Dept. 236B.

Suggested retail price, $125, excluding applicable state and local taxes—Continental U.S.A., Alaska, and Hawaii. 611/09

HEWLETT PACKARD

POSSIBLY THE ONLY MOTORCYCLE COMPANY WITH A PAST WORTH BUILDING ON.

Each year, for more years than many motorcyclists now care to remember, another wave of revolutionary motorcycles rolls into America.

Machines whose advanced engines, it is purported, will propel a rider not merely down the road but into a whole new era of motorcycling.

An era which, with increasing predictability, lasts all of one model year. When yet another wave of "new era" machines rolls ashore.

There is one motorcycle company, however, whose technology is not so transient. A company not so anxious to scrap its last year's work. Or even its last decade's.

The Bavarian Motor Works of Munich, Germany.

AN ENGINEERING HERITAGE THAT DIDN'T BEGIN THIS YEAR.

Before a motorcycle company can have a heritage, it must first find something worth building on.

While there are any number of manufacturers still searching for this elusive building block, BMW introduced in 1923 what many aficionados consider to be the perfect mechanical foundation for a motorcycle.

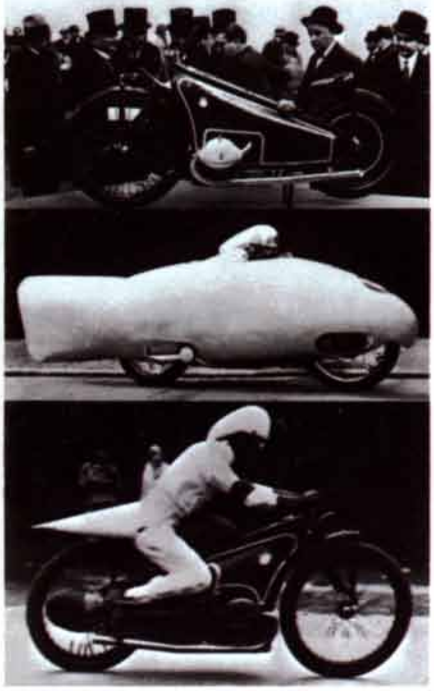
The horizontally opposed twin-cylinder engine.

A springboard which placed the BMW engineers on a path of continuous refinement—as opposed to continuous revampment—ever since.

SIMPLICITY IN AN AGE OF COMPLICATIONS.

"In typical BMW fashion," observed Cycle Guide recently, "the BMW engineers have brought their motor into the Eighties without resorting to needless complexity."

Indeed, the design of the BMW twin has grown not increasingly complex with



TOP: BMW is the rage of the 1930 Paris Motor Show.
CENTER: In 1937 BMW again builds the fastest motorcycle in the world, clocking 279.5 kph.
BOTTOM: BMW's set 9 world records in 1932.

the years but ingeniously simple —the hallmark of a brilliantly engineered machine.

It is the simple genius of the opposed twin which makes a BMW significantly less expensive to maintain and repair than most motorcycles.

For that matter, this simplicity also makes a BMW far less likely to need repair in the first place.

With the cost of a motorcycle mechanic now hovering at about $24 an hour; with the price of a mere tune-up of certain multi-cylinder motors upward of $100, it's easy to understand why the BMW twin is becoming more appreciated by the day.

The rewards of this remarkable engine, however, are not purely financial.

THE JOYS OF BEING ONE OF THE FEW MOVING PARTS ON A BMW.

Cycle Guide flatly states that "no engine configuration known to man is better at aiding the low 'cg' required in a flick-left-flick-right bike than the opposed twin."

Small wonder then that BMW's are renowned as extraordinary handling machines.

For, in contrast to motorcycles encumbered by the top-heaviness of their own engines, the BMW's cornering prowess is actually enhanced by its motor design.

Providing its rider with a control so exhilarating as to inspire the motorcycle editor of AutoWeek, Michael Jordan, to write of one BMW:

"Long after a clumsy Japanese hyperbike would have wrestled you into exhaustion, the R 100 RS urges you to uncharted apexes; the only limit to your enthusiasm is the amount of gas in the bike's 5.8-gallon fuel tank."

1982 MOTORCYCLES THAT WON'T BE DATED BY 1983.

There is an inescapable irony in how quickly so-called "motorcycles of the future" become things of the past.

While BMW's, machines whose basic engine design has remained unchanged for over a half-century, continue to endure.

It is this timelessness of design, of course, that marks the difference between a motorcycle that is perceived as a classic. And one that simply is not.

A perception which, on any given day, can be measured with the dollars-and-cents yardstick of the NADA's Used-Motorcycle Handbook. Where, over the past 3 years, BMW's have retained the astonishing average of 77.7% of their original purchase price.

Which provides additional proof that the best way to determine a motorcycle's future is to simply look at its past.

THE LEGENDARY MOTORCYCLES OF GERMANY.

BMW