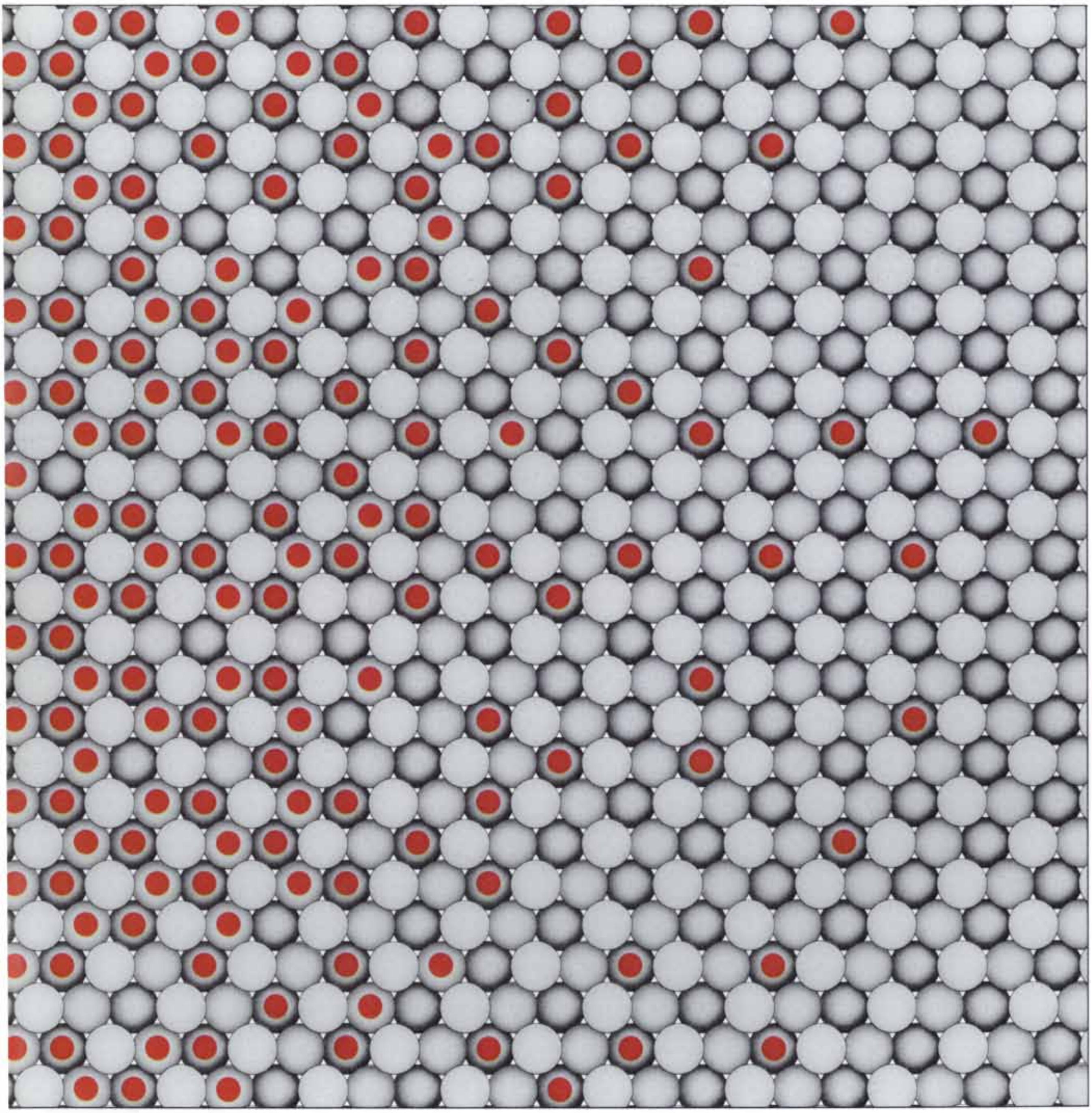
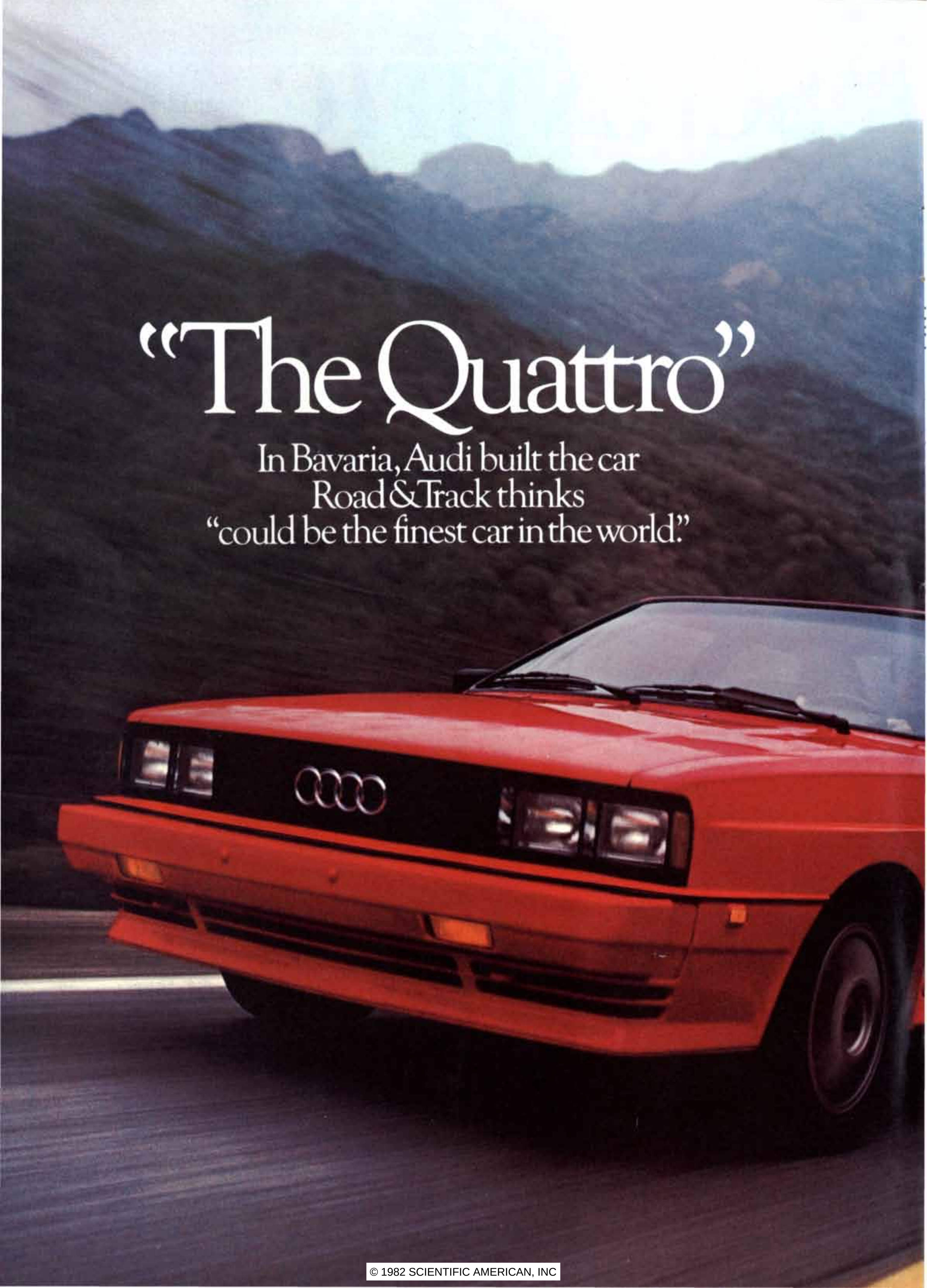
SCIENTIFIC AMERICAN

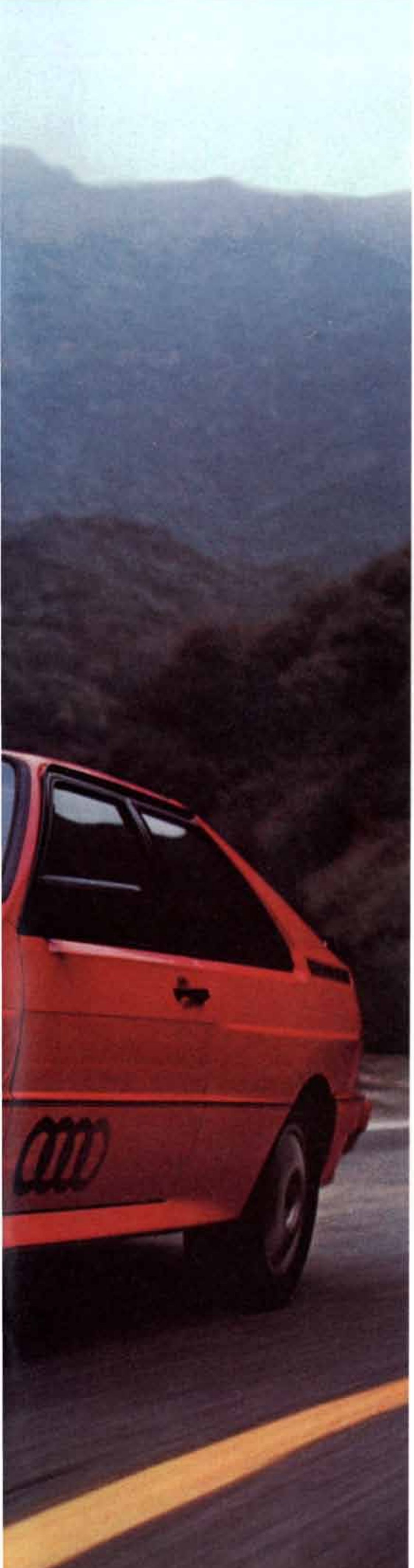


SURFACE DIFFUSION

$2.00

August 1982


"The Quattro"
In Bavaria, Audi built the car
Road & Track thinks
"could be the finest car in the world."



THE GRAPES OF REMY MARTIN. THE PRIVILEGED FEW.

For over 250 years, the house of Remy Martin has refused to compromise.

All of Remy Martin's cognac is made exclusively with the grapes from the two best districts of Cognac: Grande Champagne and Petite Champagne.

Of the six grape-growing areas within the Region, these two areas are indisputably the finest—and the only areas Remy Martin deems suitable for the making of its cognac.

But Remy Martin does not owe the quality of its cognac to these vineyards alone.

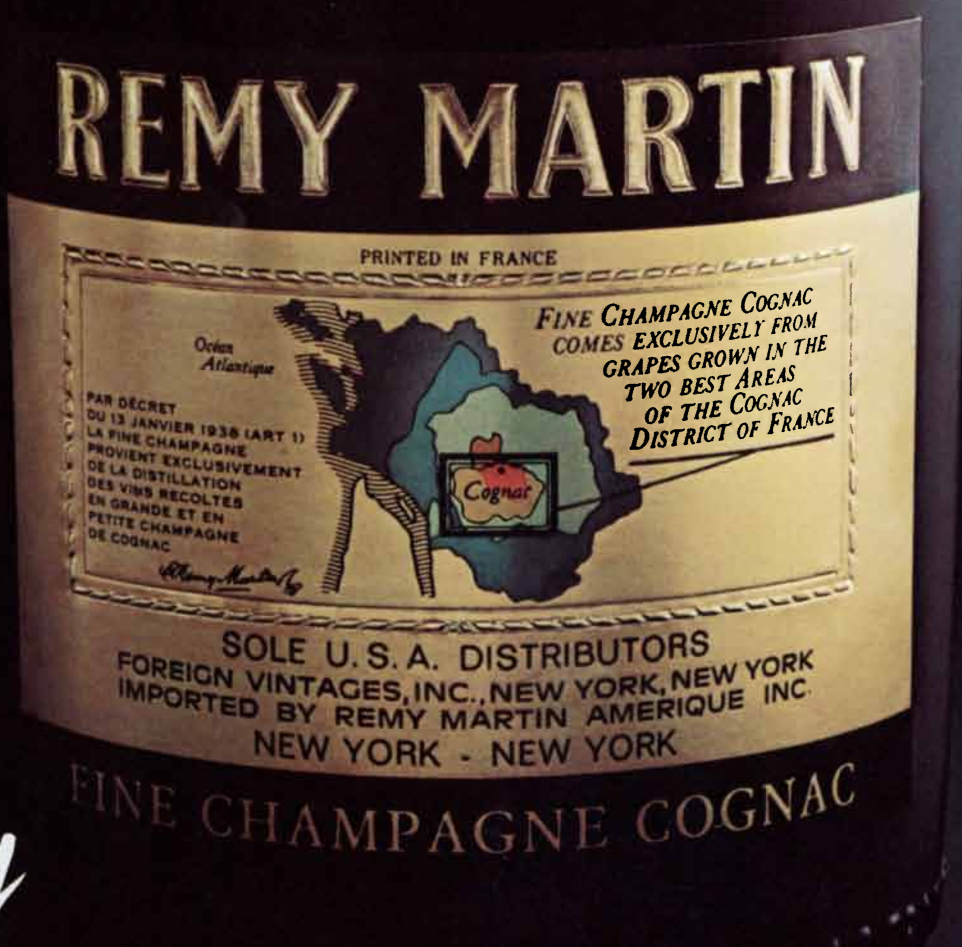
From time immemorial, Remy Martin has used only the finest oak from the majestic Limousin Forest in the making of its casks and cooperage.

It has used, exclusively, the traditional copper, onion-shaped "Charentais pot-still" to ensure the perfect and proper distillation of the young white cognac. A still that's smaller in volume than the stills used by every other major cognac house.

Yet it is, above all, the skilled craftsmanship and pride of the cognac-maker, who in harmony with nature, makes Remy Martin the finest of cognacs.

Doubtless, there will forever be less costly, less time-consuming, less painstaking ways to make cognac.

But the proud house of Remy Martin will continue to create cognac as it always has in the past; cognac of exceptional body, bouquet and character. Remy Martin. Cognac without compromise.



Rémy

THE FIRST NAME IN COGNAC SINCE 1724

EXCLUSIVELY FINE CHAMPAGNE COGNAC FROM THE TWO BEST DISTRICTS OF THE COGNAC REGION

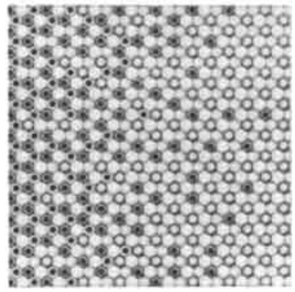
Sole U.S.A. Distributor Foreign Vintages, Inc. New York, N.Y. 80 Proof.

Established 1845 **SCIENTIFIC AMERICAN** August 1982 Volume 247 Number 2

ARTICLES

DEPARTMENTS

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



THE COVER

The illustration on the cover is a schematic view of the diffusion of adsorbed atoms on a single facet of a perfect metallic crystal (see "Surface Diffusion," by Robert Gomer, page 98). The crystal is made up of tungsten atoms, shown in gray, which are packed in the ordered array called a body-centered-cubic lattice. The atoms exposed at the surface of the crystal have a regular configuration characteristic of the orientation of the plane of the surface with respect to the lattice structure. The plane shown is a rugged one at the atomic scale. The diffusion of atoms on the surface proceeds more slowly than it would on a smoother one. Adsorbed atoms at the surface, shown in red, move from regions of high concentration to those of low concentration (*left to right*). The atoms tend first to occupy the sites on the tungsten surface where their energy can be minimized, namely in the deepest wells. The shallower energy wells are then filled as the adsorbed atoms continue to spread over the crystal plane. Diffusion on a single crystal plane can be observed and recorded at high resolution with the remarkable instrument called the field-emission microscope; the magnification of the microscope can be as great as a million diameters.

THE ILLUSTRATIONS

Cover illustration by Gabor Kiss

Page	Source
36	Space Systems Division, General Electric Company
38–42	Allen Beechel
45–46	Kennecott Minerals Company
47	Karen A. Couch
48	Joseph V. Scaletti
49	Thomas L. Kieft and Gregory P. Minion
50	James A. Brierley
51	John A. Baross, Oregon State University
52	Karen A. Couch
53	Gerald W. Strandberg, Oak Ridge National Laboratory
55	Renzo Sancisi, Kapteyn Laboratory
56–57	Walken Graphics
58	Walken Graphics (*top*), Klaas S. de Boer and Blair D. Savage (*bottom*)
59–62	Walken Graphics
87	David J. Lim, Ohio State University College of Medicine (*top*); Acarology Laboratory, Ohio State University (*bottom*)
88	Ernesto O. Hoffmann, Louisiana State University School of Medicine
89–94	Bunji Tagawa
99	James Franck Institute, University of Chicago (*left*); Gabor Kiss (*right*)
100–105	Jerome Kuhl
106	James Franck Institute
108–109	Jerome Kuhl
111–113	Tom Prentiss
114–116	Ilil Arbel
119	Lawrence E. Gilbert, University of Texas at Austin
123	Earth Satellite Corporation
124	Andrew Tomko
125	Ian G. Gass, Open University
126–130	Andrew Tomko
132	Houghton Library for Rare Books, Harvard University
134	National Central Library, Florence
135	Ilil Arbel
136	Vatican Secret Archives
138	National Central Library, Florence
140	The New York Public Library, Astor, Lenox and Tilden Foundations
142–143	Vatican Secret Archives
144	Jearl Walker
145–148	Michael Goodman

LETTERS

Sirs:

Gordon Moyer ["The Gregorian Calendar," by Gordon Moyer; SCIENTIFIC AMERICAN, May] states that the calendar established by Pope Gregory XIII in 1582 "affords a highly satisfactory compromise between essential accuracy and much-desired simplicity." The accuracy is impressive and the simplicity of introducing leap centuries to hold down the drift of the date of the vernal equinox is elegant, but there the simplicity of our calendar ends. The months vary from 28 to 31 days, they are misnamed from September through December, one cannot calculate in any simple way the number of days elapsing from one month to another, or know on what day of the month a given day of the week will fall without consulting a calendar, etcetera. Also, in contrast to the "grand cycle" Moyer describes of 400 years between the years that exactly repeat identical calendars, a simple calendar would repeat them monthly for the days of the week and annually for the days of the month.

A truly convenient system was overlooked by those whose passion for astronomical accuracy and religious tradition left them indifferent to arithmetic simplicity. Within 1.24+ of the number of days in a tropical year exists a number, 364, having three extraordinarily convenient factors: 7, 4 and 13. They suggest a year made up of 13 months, each month of exactly four seven-day weeks. All one has to do is have the last day of the year, the 365th, be a correction day. In leap year it would be two days. This would become the interim day, or two days, for ending one year and beginning the next. This interim/correction period would not be part of any week or month. The factors 4 and 13 also specify the quarters of the year, each being 13 weeks long.

Two further changes would assist the manipulation of calendar time 'in one's head': day names and month names should be number derivatives. For starters Monday becomes Oneday, Tuesday Two'sday or Twosday... Thursday (once Thor's day and day four of the new week) becomes Four'sday (Forsday?); Fiveday (Friday) would inevitably be corrupted to Fiday, etcetera. More fun and games are in store for number-naming the months, only this time we don't make the ninth month the seventh, the 10th the eighth, etcetera, as the present calendar does. If, following the Latin, we call the 13th month Tredecember and transpose Christmas (currently the 359th day of a nonleap year) to the new system, it would become Twosday, the 22nd of Tredecember, and would remain Twosday every year, including leap years and leap centuries. All the change-ups around holidays falling on different days of the week with the attendant adjustments for weekends versus weekdays would end.

Now, that's a simple calendar. With a little practice, calculating days and dates is no more difficult than counting change. There's only one thing wrong with it. It has no more chance of happening than I have of becoming Pope.

VERNON ROWLAND, M.D.

Professor of Psychiatry
Case Western Reserve University
Cleveland, Ohio

Sirs:

In his article on the Gregorian calendar Gordon Moyer failed to mention that the scholars who devised that calendar adopted the same system employed by the Jewish calendar.

Some 3,000 years ago Moses faced the identical problem in determining the date of Passover. He had to justify the lunar and tropical years because Passover, the most important festival in the Jewish calendar, must be in the spring. His solution was to fix it at "the fourteenth day of the first month" (Numbers 9:5). The "first month" was the new moon following the vernal equinox....

MERYL SILVERSTEIN

Brooklyn, N.Y.

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 ______

______

______

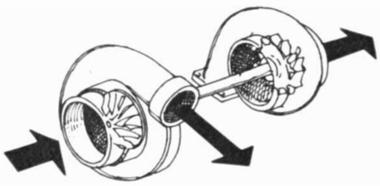
Old Address ______

______

______

A car that captures energy lost in the tailpipe and recycles it to give you "free" extra power: The Peugeot turbocharged diesel.

Two trucks lumber up a long, steep hill, their exhausts bellowing. A car noses out from behind them. It looks safe to pass. Or is it? The car is a diesel, and a diesel's strong suit is fuel economy, not swift acceleration.



The exhaust-driven turbocharger blows a 600-mph whirlwind that "packs" the engine with 13% more horsepower, 35% more torque.

The driver presses down on the accelerator pedal... and starts a fascinating chain of events.

Under the hood of the car, a tiny turbine wheel begins to spin ever faster; it will eventually reach a speed of up to 100,000 revolutions a minute.

A 600-mph whirlwind is created that literally stuffs the combustion chambers with power.

The pressure on top of each piston rises to nearly 13 *tons.*

Gathering momentum, the car surges past one truck, and then the other, and soon dwindles to a dot on the horizon.

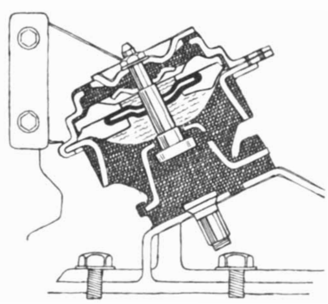
Zero to 50 in 11.2 seconds

This was no ordinary diesel. It was the Peugeot 505 S Turbodiesel.

Exhaust gases that normally would be wasted have been harnessed by a turbocharger to provide "free" extra power: 13 percent more horsepower, 35 percent more torque.

This 5-passenger sedan, weighing a ton and a half, will reach 50 miles an hour in 11.2 seconds.

The Peugeot 505 S Turbodiesel has extra power on demand for passing situations like this.



The engine has its very own suspension system. Tiny hydro-elastic shock absorbers like this are one reason why the 505 S Turbodiesel feels so smooth.

Although the performance of the Peugeot Turbodiesel yields little to comparable gasoline-engined cars, it has not sacrificed any of its diesel miserliness with fuel: [27] EPA estimated mpg, 36 estimated highway mpg.* Proof that you do not have to switch to a "downsized" car to live with today's bloated fuel costs.

Important note: 85 percent of the diesel engines sold in the U.S. are actually gasoline engines converted to burn diesel fuel. This is not Peugeot's way. Every Peugeot diesel is a *true* diesel. It is designed from scratch to withstand the enormous internal pressures of diesel combustion: the Turbodiesel's compression ratio is two and one-half times that of a gasoline engine. Engine blocks are designed with heavy structural ribs. Reinforced crankshafts, pistons, and connecting rods are fitted. The Turbodiesel's upper piston rings are coated with ultra-hard molybdenum.

The Peugeot Lion

The lion has represented Peugeot since 1858. It was Peugeot that built the first car ever sold commercially (1891)...the world's first diesel-powered car (1922)...and the first high-rpm diesel engine (1967).

A suspension of awesome efficiency

Normally, shock absorbers are designed to provide either a soft ride *or* good handling...but not both at the same time. The Peugeot shock absorbers, however, are awesomely efficient damping devices. They will swallow large bumps without breaking stride *and* deliver excellent handling.

Cut apart a Peugeot shock absorber and you will discover its secret: the piston contains *four times as many valves* as a normal shock absorber.

Every car is tested

After the car comes off the assembly line, it is put to a test that is extremely rare today. Every single Peugeot is taken to a special track where it is tested by one of a 65-member team of expert drivers. It must meet Peugeot's standards or it is not permitted to leave the factory.

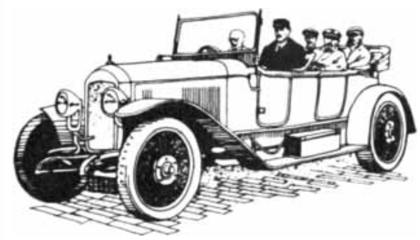
Peugeot has little tolerance for a poorly made car.

The meaning of "S"

Peugeot does not believe in nibbling you to death with extra-cost options. When you spend $16,175** for a 505S Turbodiesel, this is the "S" (for Special) equipment that comes with the car:

Factory-installed air conditioning
Electric windows
Electric sunroof
Cruise control
Digital AM/FM 40-watt stereo radio with cassette player, scan tuning, and four speakers
Automatic electric antenna
Central door-locking—a twist of the key locks all four doors and fuel-filler door
Alloy wheels
Multi-adjustable driver's seat.

So fully equipped is the 505S that the only other ways to spend your money are by ordering a 3-speed automatic transmission ($370) in place of the manual 5-speed overdrive gearbox, and by specifying metallic paint ($295).



The world's first diesel-powered car, built by Peugeot in 1922.

See your dealer for details of Peugeot's limited warranties.

*With 5-speed manual transmission. Use EPA estimated mpg for comparison purposes. The mileage you get may vary with trip length, speed, weather, and condition of car. Actual highway mileage will probably be less.

** Based on P.O.E. manufacturer's suggested retail prices. Actual prices may vary according to local dealer. Destination charges ($255), state and local taxes, dealer preparation, if any, and license fees are extra.

For name of nearest Peugeot dealer, call **1-800-447-2882** toll free in the continental U.S. In Illinois, 1-800-322-4400.

Please send me the Peugeot Facts Kit, complete with brochures, road tests by leading car magazines, plus list of my nearest dealers.

☐ Please include details on Overseas Delivery

Name__________

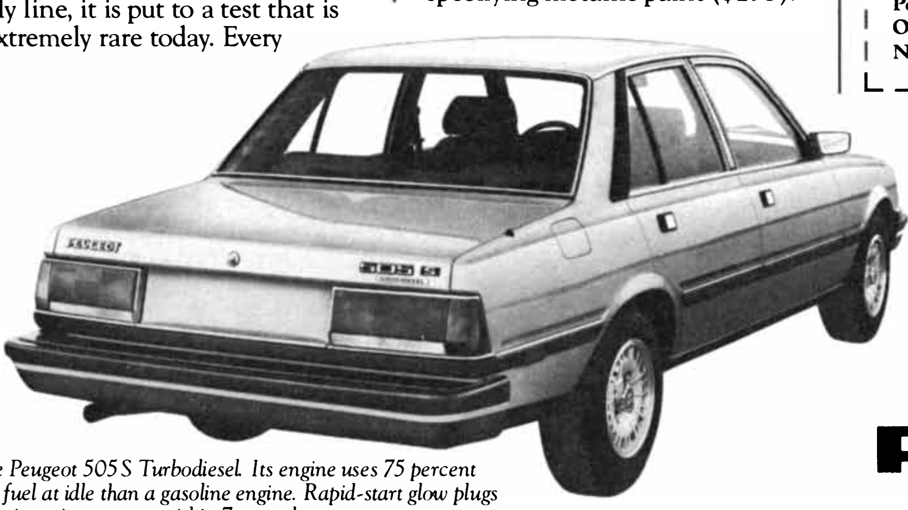
Address__________

City__________

State______ Zip______

Peugeot Motors of America, Inc., One Peugeot Plaza, Lyndhurst, New Jersey 07071

54-000SA-0008



The Peugeot 505S Turbodiesel. Its engine uses 75 percent less fuel at idle than a gasoline engine. Rapid-start glow plugs permit engine to start within 7 seconds.

PEUGEOT

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

AUGUST, 1932: "When some wise historian of the remote future undertakes to interpret our time, the early decades of the 20th century will call for a special comment, because it was then that physicists began an attack on the atom that resulted in discoveries that changed the whole character of chemistry and engineering. To that historian the experiment conducted by Drs. John D. Cockcroft and Ernest T. S. Walton of Cambridge University, which resulted in the splitting asunder of lithium atoms and the reuniting of their shattered nuclei in new combinations, will be singled out as an example of the method that finally led to the voluntary transmutation of the elements, to the controlled release of the energy that holds matter together and thus to a revelation of the whole plan and method of creation. Drs. Cockcroft and Walton are not the only ones who have devised an electrical method of hurling particles at atoms. Drs. Merle A. Tuve, Lawrence R. Hafstad and O. Dahl of the Carnegie Institution of Washington have been experimenting for many months with protons to which energies of 2,000,000 and 3,000,000 volts have been imparted."

"With the discovery of the neutron, science has one more physical entity to juggle with. The possibility of dense matter being formed out of neutrons might help to explain the nature of the 'white dwarf' stars. These stars are thousands of times denser than water and are supposed to be at extremely high temperatures. If they had a core of neutrons, they could be dense and yet cool. The neutron may have remarkable possibilities in astrophysical theory."

"Fliers have now crossed the North Atlantic by the great-circle route, by the Faroes and Iceland, by Newfoundland and Ireland and by the Azores, and they have crossed the South Atlantic from Africa to Brazil. The east-to-west crossing has proved to be more difficult on the whole than that from the American continents to Europe and Africa; the southern crossing is least troubled by storms and fog. It naturally follows that one of the most seriously considered services should be one across the South Atlantic—from Europe to Africa and thence across the ocean to Brazil and return. Among the competitors for first honors in establishing trans-Atlantic air services it is therefore likely that either the Italians or the Germans will launch their services first, the former with airplanes and the latter with an airship and with airplanes."

SCIENTIFIC AMERICAN

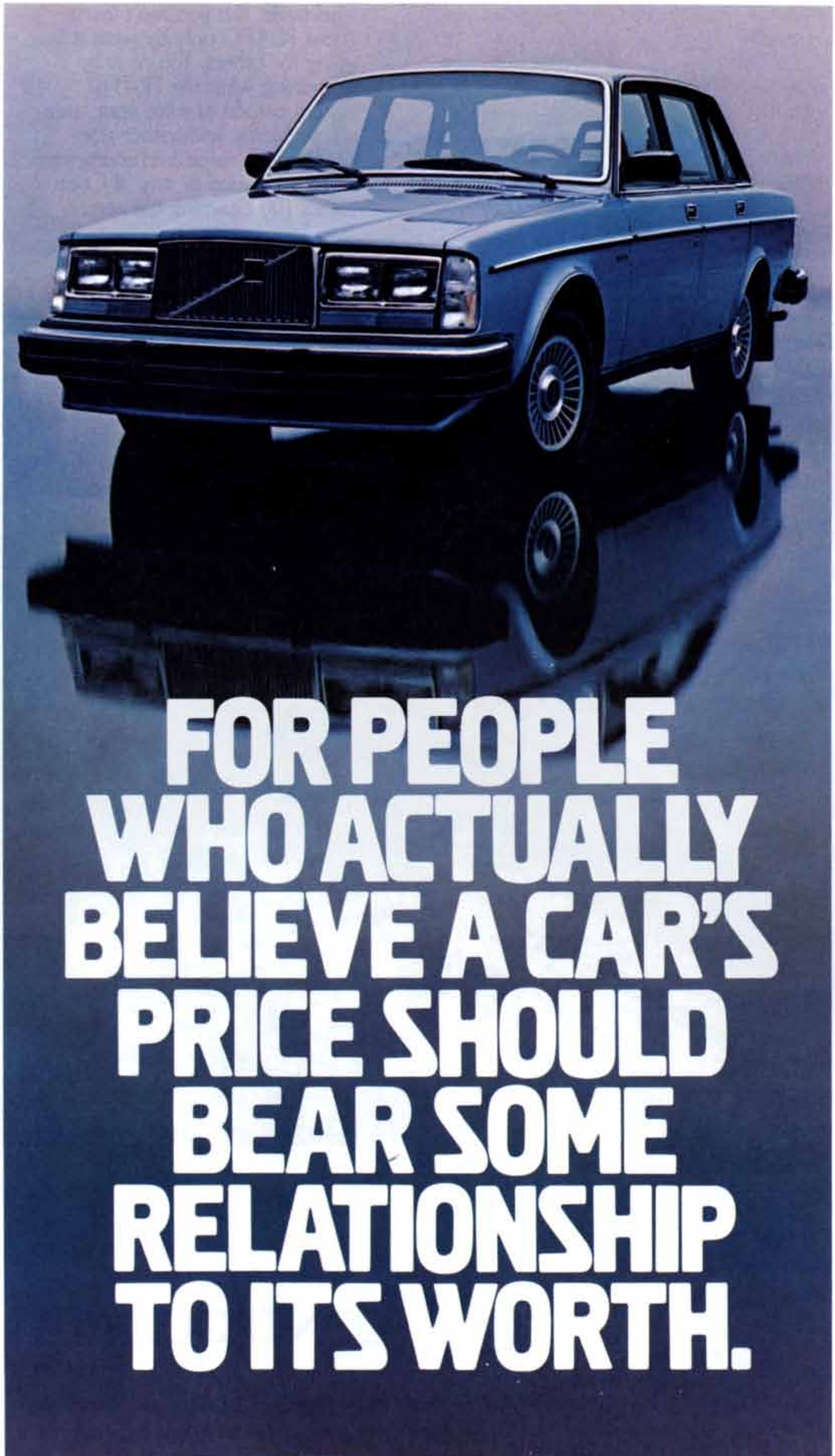
AUGUST, 1882: "On Pearl Street in New York, under the shadow of the Third Avenue Elevated Railroad, is an iron-front building, originally put up for commercial purposes, that for a year or more has been in process of preparation for a central electric lighting station under the Edison system. The beginning of this great work was indicated by the laying of underground conductors around every block in that portion of the city bounded on the east by the East River, on the west by Nassau Street, on the north by Spruce and Ferry streets and Peck Slip and on the south by Wall Street. This district included 946 consumers, whose premises are already wired. The number of lamps to be used in connection with these wires is 14,311. Now the street conductors are laid, the service conductors are put in, the buildings are wired, the dynamos with their attached engines are in place and the district and central stations are fully equipped, and we have no doubt that before this paper meets the eye of the reader the district will have been illuminated."

"Robert Koch has endeavored to ascertain what agents are able to destroy the spores of bacilli, how they behave toward the microphytes most easily destroyed, such as the molds, ferments and micrococci, and whether they suffice at least to arrest the development of these organisms in liquids favorable to their multiplication. His results with phenol, thymol and salicylic acid have been unfavorable. Sulphurous acid and zinc chloride have also failed to destroy all the germs of infection. Chlorine, bromine and mercuric chloride gave the best results; solutions of mercuric chloride, nitrate or sulphate diluted to one part in 1,000 destroy spores in 10 minutes."

"William Ramsay of University College, Bristol, believes there is a probability that our sense of smell is excited by vibrations of a lower period than those that give rise to the sense of light or heat. These vibrations are conveyed by gaseous molecules to the surface network of nerves in the nasal cavity. The differences among smells are caused by the rate and by the nature of such vibrations, just as the difference among tones of musical sounds depend on the rate and on the nature of the vibration, that nature being influenced by the number and pitch of the harmonics. It is to radiant heat, Mr. Ramsay thinks, that we must look for indications of harmonics of the fundamental vibrations that are, according to his theory, the cause of smell. Prof. Tyndall has shown the power odors have of absorbing heat rays. There is no doubt that by refracting such heat rays by means of a rock-salt prism, after they have passed through an atmosphere of odor certain portions of the heat spectrum show colder spaces, each corresponding to the particular rate of vibration that is absorbed by the vapor. By measuring the position of such gaps in the heat spectrum, calculating the particular vibration of the rays at such gaps and referring them to their fundamental, we should arrive at the rate of vibration of the molecule that causes smell."

"A correspondent who has lately gone over the territory devastated by the great fire in the forests of Michigan last fall says his observations are conclusive that phenomena aside from the ordinary conditions of combustion were developed. In the first place the fire created at least two centers that had the essential features of storms, especially the spiral winds. The evidences are confirmatory of the belief that this storm center, after it became fully developed, consisted of a heated body of air or gas in a state of combustion, constantly fed by the smoke and vapor driven to the center by the whirling winds and the gases generated in the combustion of the pines and other resinous woods. This body of air or burning gas by its heat acquired an ascensive force, but by the rapid forward motion of the fire it was sucked forward, actually preceding the fire proper. It is evident that this body was of intense heat, possibly as great as 400° Fahr., at which point oxygen and carbon unite. That such a body of luminous vapor existed, detached from the fire, is asserted by many who saw it from a distance."

"At the recent meeting of the New York State Teachers' Association the report of the Standing Committee on the Condition of Education showed that, notwithstanding the steady increase in the population of the state, the number of children in daily attendance in the public schools is declining. The decrease was attributed by the chairman of the committee to 'the increased demands made by manufacturing interests,' by which was meant, we presume, an increased employment of children in factories. At other teachers' gatherings this summer the same condition of things has been noted and variously commented upon as being more or less visible throughout the country, and more or less to be deplored. The general feeling seems to be that the schoolmaster is losing his grip and that the country is likely to suffer."



© 1981 VOLVO OF AMERICA CORPORATION. ELECTRIC SIDE MIRRORS OPTIONAL

There are, fortunately, still some people in the world who can discern the difference between a car that's worth the money and one that's outrageously overpriced.

If you're one of them, you'll appreciate the Volvo GL.

It's an automobile that not only exhibits extraordinary workmanship and care of construction. It abounds in the refinements you'd expect in a fine prestige car: air conditioning, sun roof, power windows, orthopedically-designed front bucket seats, and a driver's seat that automatically heats up when the temperature falls below 57°. (All of these things, incidentally, are standard.)

In one place, and one place only, does the Volvo GL fall short of other select imported luxury cars. In price.

Which simply means it's now possible for you to purchase a fine European luxury sedan.

Without being taken for a ride.

VOLVO

A car you can believe in.

THE AUTHORS

ROGER REVELLE ("Carbon Dioxide and World Climate") is professor of science and public policy at the University of California at San Diego. After he had worked for many years as an oceanographer his interests expanded to include numerous interactions of science, the biological environment and economic development. He was graduated from Pomona College in 1929 and went on to obtain his Ph.D. from the University of California in 1936. In the same year he joined the faculty of the Scripps Institution of Oceanography as an instructor, ultimately becoming professor of oceanography. From 1951 to 1964 he was director of the Scripps Institution. In 1963 and 1964 he was university dean of research. In 1964 he went to Harvard University as Richard Saltonstall Professor of Population Policy. From 1964 to 1976 he was director of the Center for Population Studies at Harvard. He returned to San Diego in 1976. Revelle has served in many public positions, including that of science adviser to the Secretary of the Interior from 1961 to 1963. He was the first chairman of the Board on Science and Technology for Development of the National Academy of Sciences and was a member of the U.S. delegation to the United Nations Conference on Science and Technology for Development. In 1974 he was president of the American Association for the Advancement of Science. His interest in the role of carbon dioxide in the environment goes back to his graduate-school days. Revelle writes: "My first major scientific paper concerned the role of carbonate and bicarbonate ions and dissolved carbonic acid in controlling hydrogen-ion concentration in the ocean."

CORALE L. BRIERLEY ("Microbiological Mining") is president of Advanced Mineral Technologies, Inc., a company she helped to found this year. She received a bachelor's degree in biology at the New Mexico Institute of Mining and Technology in 1968 and a master's degree in chemistry from the same institution in 1971. She joined the New Mexico Bureau of Mines in 1971 as a chemical microbiologist and remained in that position until earlier this year, taking time off to get a Ph.D. in environmental sciences (granted last year) from the University of Texas at Dallas.

KLAAS S. DE BOER and BLAIR D. SAVAGE ("The Coronas of Galaxies") are astrophysicists who share an interest in the diffuse gases that occupy the space between the stars in spiral galaxies such as our own. De Boer received his university education at the University of Groningen. From 1971 to 1974 he was a member of the faculty at Groningen. From 1974 to 1977 he was a member of the Netherlands Space Research Group and served part-time on the Astronomical Netherlands Satellite ultraviolet team. In 1978 he came to the U.S. as research associate at the University of Wisconsin at Madison. In 1981 de Boer joined the faculty of the University of Tübingen in Germany. Savage's bachelor's degree in engineering (1964) is from Cornell University. His M.A. (1966) and Ph.D. (1967) are from Princeton University. In 1967 and 1968 he was research associate at Princeton. In 1968 he went to the University of Wisconsin at Madison, where he became professor in 1978. Savage and de Boer collaborated on the work that led to the present article while both were at Wisconsin.

PAUL D. BUISSERET ("Allergy") is associate professor of medicine at the Louisiana State University School of Medicine. Born in Shropshire, he got his medical degree at the University of Leeds in 1964. He then served an internship and residency and spent some years in general practice. From 1971 to 1976 he was on the staff of St. Mary's Hospital in London. In the same period he was a member of the faculty of the Wright-Fleming Institute of Microbiology. From 1976 to 1979 he was on the faculty of Guy's Medical School in London, where he did work on the subject of the present article. In 1979 Buisseret moved to Louisiana State; his work there has been primarily in pulmonary physiology and medicine.

ROBERT GOMER ("Surface Diffusion") is professor of chemistry and director of the James Franck Institute of the University of Chicago. He is a native of Austria whose family moved to Scotland in 1938 soon after Austria's annexation by Germany. After a short stay in Scotland the family came to the U.S., where Gomer obtained his B.A. at Pomona College in 1944 and his Ph.D. from the University of Rochester in 1949. After a year of postgraduate work at Harvard University he joined the faculty at Chicago. His work has mainly focused on the phenomena of diffusion and adsorption; he has also done work in electrochemistry. He writes that "after hearing a seminar account of E. W. Müller's field-emission microscope I built one and became active in developing this instrument as a tool for surface studies." Gomer describes himself as an "outdoor intellectual" much of whose vacation time is spent skiing and hiking in Austria.

LAWRENCE E. GILBERT ("The Coevolution of a Butterfly and a Vine") is associate professor of zoology and director of the Brackenridge Field Laboratory of the University of Texas at Austin. His B.A. (1966) is from the University of Texas. After graduation he spent a year at the University of Oxford on a Fulbright Fellowship. He got his Ph.D. from Stanford University in 1971. In that year he joined the faculty at Texas. His main research interest is the social and demographic organization of insects that feed on plants. Gilbert writes: "My interests in insect-plant relations probably derive from helping my grandfather with his honeybee business in southern Texas each summer during my youth."

IAN G. GASS ("Ophiolites") is professor of earth sciences at the Open University in England. Born in England, he spent his early childhood in Burma. His original academic interest was history, but his university career was interrupted by World War II and in the course of four years of military service he became interested in geology. When the war ended he returned to the University of Leeds to study geology, receiving his B.Sc. in 1952. After being graduated he served in the geological surveys of the Sudan and Cyprus. In each case he was involuntarily retired when the country achieved independence. He writes that he then "decided to try academic life as the prospects of being repeatedly retired from the diminishing British colonial empire did not appeal." In 1960 he obtained his Ph.D. from Leeds. After teaching briefly at the University of Leicester, Gass joined the faculty at Leeds and remained there until 1969, when he moved to the Open University.

OWEN GINGERICH ("The Galileo Affair") is professor of astronomy and the history of science at Harvard University and a member of the staff of the Center for Astrophysics of the Harvard College Observatory and the Smithsonian Astrophysical Observatory. His B.A. (1951) is from Goshen College. His M.A. (1953) and Ph.D. (1962) are from Harvard. From 1955 to 1958 he was a member of the faculty of the American University in Beirut. In 1958 and 1959 he was lecturer in astronomy at Wellesley College. In 1960 he went to Harvard as instructor, becoming professor in 1969. His research interests have been divided between astrophysics and the history of astronomy. Much of his astrophysical work has dealt with the atmosphere of the sun. His historical research has centered on the work of Johannes Kepler and Nicolaus Copernicus. Gingerich notes that his work on the Galileo affair has been helped by discussions with the Reverend Joseph Clark and the Reverend Ernan McMullin. Gingerich also acknowledges the assistance of George V. Coyne, S.J., the director of the Vatican Observatory.

METAMAGICAL THEMAS

Undercut, Flaunt, Hruska, behavioral evolution and other games of strategy

by Douglas R. Hofstadter

In the summer of 1962 Robert Boeninger and I, both young mathematics students at Stanford, were riding in a bus in Germany on the way back from a trip to Prague when boredom set in. To pass the time we invented a curious number game. The rules were simple but the game was tricky because it involved trying to "psych out" your opponent in devious ways. The rules we initially made up went like this. The game would consist of 10 turns. On each turn we would each choose a number in secret and then we would compare them. One of us, however, would choose a number (always an integer) in the range 1 to 5 and the other would choose a number in the range 2 to 6. Each of us would get to "keep" our own number, that is, to add it to our score—provided the numbers did not differ by 1. If, however, they did differ by exactly 1, then the player with the *lower* of the two numbers collected *both* of them. Thus if I said 2 and Robert said 3, well then, I would get five points and poor Robert would get none. Very jolly! At least until I said 5 and Robert said 4.

It seemed amusing to have the ranges not quite coincide, since it is hard to sort out who really has the advantage. One's first impression might be that the "larger," or 2-to-6, player has an advantage, but that is nicely counterbalanced by the fact that if you name 6, you are running the risk of being undercut by your opponent's 5, whereas your 6 cannot undercut anything. Moreover, the "smaller" player can always name 1 safely without any risk of being undercut. Although the asymmetry seemed charming, we soon decided that having equal ranges—both 1 to 5—was probably preferable. And that was the way we played the game, which I shall here call Undercut.

Competition was pretty fierce. The lovely thing about the game was how level on level of "outpsyching" could pile up in our minds. For instance, I could tease Robert by choosing 4 a few times in a row, trying to lure him into saying 3, and just at that moment plan to switch my move on him, jumping to 2 and outfoxing him.

When I returned to Stanford from Europe that fall, I was eager to get a computer to play the game. My friend Charles Brenner had recently written a program that compiled frequencies of letters and letter groups (actually trigrams) in a piece of text in English (or any other language) and then, using a random-number generator, produced a pseudo-English output whose trigram frequencies were the same as those of the input text sample. I had been most impressed by the way seemingly deep patterns of English could be so aptly captured by such an algorithm, and I saw how the concept could be adapted to a game-playing program. In particular I was taken by the idea of getting a program to detect patterns in the move sequences of its opponent and then use them to generate predictions—in short, having the computer itself try to "outpsych" its opponent. It would be all the better if the opponent program were also trying to do something similar to my program.

I have vivid memories of standing over the printer from which the output would spill forth, watching the progress of games emerge line by line. We would have our programs play games of several hundred turns, thereby putting them to a serious test. My program often started out on a losing tack, given that it had not yet "smelled" any patterns in the opposing program's behavior. Sooner or later, though, there would be a moment when my program would appear to catch the scent and would make a decisive undercut or two in a row. Then I would see it begin to surge forward, often leaping quickly into the lead and wiping the opponent out. It gave me a feeling of overwhelming power: the power of insight defeating raw strength. It reminds me now of one of my favorite book titles: *Chess for Fun and Chess for Blood* (by Edward Lasker). This title captures exactly that subtle blend of goodwill and rivalry one feels in a highly competitive game with friends.

Since then I have realized how universal, how primitive, such a feeling is. It is probably the most engrossing aspect of all sports, that feeling of pitting two strategies and watching them fight it out. Even dogs seem to enjoy this kind of sporting competition. When I play a chasing game with my friend Shandy the airedale, I detect that he has a precise sense for how well I can anticipate his moves, and in his dodging tactics he always stays one ply—one level of trickery—ahead of me. Whenever I think I have caught on to his pattern, he somehow senses it and just at that moment shifts his strategy so that I wind up lunging for a dog that is not there.

Oh, to be sure, he does let me win sometimes, just to keep me interested. He even has the instinct of teasing, dropping his prized stick or ball right in front of me, acting nonchalant about it and coolly daring me to make a move for it. But he has it all calculated out. He knows how quick I am, how quick he is and what my patterns of trying to fake him out are.

What is more, Shandy often seems to come up with new ways of shifting his strategy, so that I cannot simply catch on to the metapattern of his strategy shifts and thereby outwit him. No, there is something extremely cunning in his dog's mind, and clearly the joyous exercise of that native intelligence reflects a deeper quality of dogs and people in general, namely the enormous evolutionary advantage intelligence seems to confer on beings that have it in this dog-eat-dog, people-eat-people world.

Back to Undercut. One day Jon Peterson, a math graduate student who used to hang around the "comp center," challenged my program with his own. He said he had used game theory in his program. At first I wasn't worried, but when I pitted my program against his, I soon saw there was good cause for worry. It was not that my program got trounced by his; it just never caught on to any patterns and simply wound up more or less tying his program every time. This was baffling. Jon explained that he had computed appropriate weights for each choice, 1 through 5, weights having nothing to do with the opponent's strategy, only with the payoff for each set of possibilities. Here is the "payoff matrix" he was talking about:

ME \ JON	1	2	3	4	5
1	0	−3	2	3	4
2	3	0	−5	2	3
3	−2	5	0	−7	2
4	−3	−2	7	0	−9
5	−4	−3	−2	9	0

The matrix shows, for each combination of numbers, how much Jon stands

to gain with respect to me. Note the "antisymmetry": the fact that each number, when it is reflected across the diagonal of zeros, changes sign, signaling that what is good for me is bad for him. And of course the zeros down the diagonal mean that when we name identical numbers, it does neither of us any good (other than carrying us one turn closer to the end of the game).

Since the game is completely symmetrical for the two players, there can be no winning strategy; otherwise both players could use it and be assured of beating each other. Nevertheless, there is an *optimal* strategy, according to game theory, that in a statistical sense will guarantee you long-term parity with your opponent. That strategy is based on assigning statistical weights to the five numbers. To find those weights you have to solve five simultaneous homogeneous linear equations. Each equation is based on making your opponent's expectation equal to zero. If Jon's weights for playing 1, 2, 3, 4 and 5 are respectively a, b, c, d and e, then *my* expectation, when I choose, say, 3, will be $-2a + 5b - 7d + 2e$ (read straight off the third row of the payoff matrix). Set this expectation at zero and you have one of the five equations. The other four arise analogously. The system to solve is thus

1. $-3b + 2c + 3d + 4e = 0$
2. $3a - 5c + 2d + 3e = 0$
3. $-2a + 5b - 7d + 2e = 0$
4. $-3a - 2b + 7c - 9e = 0$
5. $-4a - 3b - 2c + 9d = 0$

This amounts to inverting a 4×4 matrix. Jon had done so, and he came up with the weights 10, 26, 13, 16 and 1 respectively for choosing 1, 2, 3, 4 and 5. Thus according to game theory an optimal player should play 5 very seldom: one time out of 66. And 2 should be your commonest choice. It would do little good, however, to play 10 1's in a row, followed by 26 2's, 13 3's and so on. One must choose completely at random, given these weights.

Imagine a 66-sided die 10 of whose faces have the number 1 on them, one of whose faces has the number 5 on it, and so on. You must throw such a die (or a suitable computational simulation thereof) on each move. In other words, when you play according to this strategy, you must avoid any and all patterned behavior. No matter how tempting it may be, you must not yield. Even if your opponent plays 5 a dozen times in a row, you must totally ignore it and merely keep on throwing your 66-sided die obliviously. That is the way Jon's program played and why my program found nothing to pick up on. If Jon's program had ever given in to temptation and tried to outguess me, my program would probably have picked up some pattern and twisted it around to work against his. But his program knew nothing of temptation or teasing. It just played blindly on, and the longer the game was, the more surely Jon's program would break even. If it won, so much the better for it, but it had only a 50–50 chance of doing so. That's the "optimal strategy" for you!

It was a humiliating and infuriating experience for me to watch my program, with all its "intelligence," struggle to overcome the blind randomness of Jon's program, but there was no way out. I was disappointed to learn that in some sense the most intelligent strategy of all not only was dumb; it paid no attention to the enemy's moves. Something about this seemed directly contrary to the original aim of Undercut, which was to have the players trying to psych each other out to increasingly deep levels.

When I saw the game so completely demolished by game theory, I abandoned it. Recently, however, I have returned to thinking about such games in which patterns in one's play can be taken advantage of, even if game theory can in some theoretical sense find the optimal strategy. There is still something curiously compelling about the teasing and flirting and other ploys that arise in these games, something that vividly recalls strategies in evolution and even seems relevant to many current political situations.

Furthermore, there is something strikingly academic and bookish about undertaking a purely game-theoretical strategy when one is playing against a human opponent, particularly in the face of "teasing" strategies. Obviously human beings have goals in life more complex than merely winning a game, and this fact determines a lot about how they play a game. Impatience and audacity, for instance, are both important psychological elements in human game playing, and an optimal strategy in the ordinary game-theoretical sense does not take them into account. Hence I feel games of this kind are still important models of how people and larger organizations tackle complex challenges and threats.

Let me describe, therefore, some recent variations on Undercut I have been experimenting with. They all involve extending the degree to which one can go out on a limb by baiting one's opponent. My purpose was to encourage teasing, which means that one player flaunts a pattern for a while, in effect saying, "I dare you—just try an undercut!" In order to encourage this kind of pattern-flaunting it seemed reasonable to award bonus points to any pattern that is not picked up by the opponent. Let us call this version of the game Flaunt.

Suppose you and I are playing Flaunt. I say 4, you say 1. As in Undercut, I get four points, you get one. Now, on my next turn suppose I again say 4 and you say 2. If we were playing Undercut, I would again get four. In Flaunt, however, repetitions are rewarded. Therefore I am given the product of my two numbers: 4×4, or 16. Now suppose on my next turn I again play 4 and you again play 2. My daring earns me $4 \times 4 \times 4$, or 64, points, while you get 2×2, or four. Thus in these three turns I have gained $4 + 16 + 64$, or 84 points, to your $1 + 2 + 4$, or seven. Of course, you have not been oblivious to my prancing about; you have merely been biding your time. Now you make your move, a 3, hoping to undercut me. Too bad, I chose 2! I get five points, you get nothing. Sorry, sucker.

But suppose I had been so dumb as to let you catch me at this. If I had indeed said 4 this time, I would have been hoping for 256 points. But since you undercut my pattern with your 3, you get a high reward for it, namely 259 points (your three points plus my 256).

Flaunt is a theme amenable to many variations. The one I have just presented is the simplest. If you like, more complex patterned behavior can be rewarded. I am not sure of the best way to do this, and so what follows—the game I call Superflaunt—is only one possible way to reward pattern-flaunting. Suppose instead of playing 1–2–2 against my 4–4–4 moves you had played 2, then 1, then 2. You might well have had a good reason for doing so. Maybe it was the continuation of a pattern for *you* and was worth your while keeping up for the moment. If your preceding four moves had been 2–1–2–1, then these last three moves would have continued that pattern. Depending on how the scoring is done, extending your own established pattern might be more worthwhile than undermining my relatively new one. If 2–1–2–1–2–1–2 is worth the product of its elements, then that amounts to 16 points. (Actually it is worth 16 only if it was preceded by another 2–1, but that is beside the point.) By the time you have picked up on my 4–4 pattern maybe it seems worth it for you to let me have my third 4 while you name one more 2, thinking that this will lull me a bit and that at that moment you will suddenly strike and undercut me.

What should constitute a pattern in this game of Superflaunt? At the moment I am inclined to limit it to one fairly simple definition, although more complex definitions might be possible. The main idea is that a pattern exists when in a given situation you do what you did the last time you were in "that situation." Hence it all hinges on what is meant by the notion of "same situation." Let us say you have just played x and are about to play y. We shall say you are creating a pattern provided that the most recent time you played x you also followed it with y. If, for instance, your last seven moves had been 3–4–1–5–3–4–1, then to have your pattern continue you must play 5, and after that you must play 3, 4, 1, 5, 3, 4, 1 and so on. When you first establish the sequence 3–4–1–5, you of course get no bonus points, because until the repetitions start there is no pattern. Thus only when the second 4 is played has a pattern arisen, and it nets you 12 (3×4) points. The next patterned move, 1, nets you another 12 points, and then saying 5 gives you 60 points (as long as it is not undercut). As soon as you break the pattern, however, your cumulative product must start out again from scratch.

If you had played 3–4–1–5–3–4 and were worried about the obviousness of playing 1 next, you might choose to play 4, which, although it breaks one pattern, establishes another (namely 4–4). Now, in ordinary Flaunt this in itself would already be worth 16 points, but in Superflaunt only on your *next* 4 would you begin to reap the benefits of your patterned playing, since only then would you have made "the same choice" in "the same situation" twice in a row.

A limitation of Undercut and Flaunt is that both restrict your moves to a small range. I wanted a game in which numbers of arbitrary size were allowed. It was not too hard to come up with the following game, which I call Underwhelm. You and I both think of positive integers. If they are unequal and do not differ by 1, then whoever named the *smaller* one gets that number. (The other player gets, of course, nothing.) If the integers do differ by 1, then whoever named the *larger* one gets *both* numbers. In that respect Underwhelm is like a tipped-over version of Undercut. (Another name for it was Overcut.) If our two numbers are equal, then for this turn neither of us gets anything.

The goal can be a fixed number of points—any number. A good choice might be 1,000, although 100 or even a million will do just fine. Think about what this does to the game. Clearly it is not useful for you to name huge numbers, because I am likely to name a lower number and then you will get nothing and I will get something. There is pressure on both of us—it seems—to play fairly small numbers. If, however, we stick to very small numbers, then the likelihood of being overcut is fairly high. Furthermore, the scores will advance very, very slowly. If we are progressing toward the goal of 1,000 points at a snail's pace, someone will want to speed things up. And so someone will go out on a limb, naming a big number such as 81. Of course, doing so just once is not useful, because the other player will not have known in advance that the 81 was coming.

Suppose, however, I say 81 several

times in a row. (Pattern-flaunting is not rewarded in Underwhelm, by the way, at least not in this version.) You will soon catch on, and you may well be tempted to say 82 to overcut me. Or perhaps you will want to score points more conservatively off my foolishness, by simply choosing numbers close to 81 but below it, such as 70. Aha! Once I have lured you up into my vicinity, then of course I can start trying to jump *below you.* And maybe I can even anticipate when you will bite. If I can, then I can really take you to the cleaner's.

The interesting thing about Underwhelm is that by using obvious patterns as bait to lure an opponent either player can in essence establish one or more Undercut-like games at various positions along the number line. I can set up a game in the vicinity of 81, trying to coax you into saying 82 just when I anticipate it. Meanwhile you may be playing a baiting game down around 30, getting 30 points each time I extravagantly bait you with my 81, and you know that sooner or later I am bound to try to catch you there, either going below you or overcutting you.

What I find fascinating is how many parallel subgames can arise spontaneously. Particularly interesting is what happens toward the end, when one player has a significant lead. At that point the trailing player will tend to play very conservatively, naming very small numbers. This means the possibilities for overcutting are much enhanced. Moreover, there is an entirely psychological element to the game having to do with human impatience. No one wants to dawdle to victory by choosing smallish numbers several hundred times. Therefore the simple quest for variety will inevitably lead to some quirky, daring play every once in a while, and that will of course be exploitable.

Much of the spontaneous and creative teasing behavior that tends to arise in these games has its parallels in evolution. The most picturesque and vivid portrayal I know of the uncanny patterns and counterpatterns set up by living organisms competing against one another is given by Richard Dawkins in his book *The Selfish Gene.* The discussion centers on the notion of an "evolutionary stable strategy" (ESS), a term introduced by John Maynard Smith. An ESS is defined as "a strategy which, if most members of a population adopt it, cannot be bettered by an alternative strategy." Here, however, "adoption of a strategy by an individual" really means that that individual has genes for that behavioral policy.

Dawkins' first example of this concept involves rival genes for two types of aggressive behavior in a given species. The two strategies are named "hawk" and "dove," and they have the recent political connotation of those terms. If x positive points are assigned for winning a fight, y negative points for wasting time and z negative points for getting hurt, one can calculate as a function of x, y and z the eventual optimal balance of hawks and doves in the population. This may be an average over time, involving swings back and forth between the population's now consisting mostly of hawks and then consisting mostly of doves, or it may be an eventual equilibrium in which the ratio is stable.

Dawkins considers a wide variety of colorful everyday examples in human life, carefully comparing them with strategies in the world of nonhuman evolution. Such things as "gas wars," with their price-fixing and treacherous undercutting, fall neatly into line with the game-theoretical analysis Dawkins brings to bear. Some other strategies considered are "retaliator" (an individual that, when it is attacked by a hawk, behaves like a hawk, and when it is attacked by a dove, behaves like a dove), "bully" (an individual that goes around behaving like a hawk until another individual hits back, then immediately runs away), "prober-retaliator" (an individual that is like a retaliator but that occasionally tries a brief experimental escalation of the contest). These five strategies can all be activated at once in a computer simulation of a large population, just as the strategies in Undercut can fight each other on a computer. From such simulations one can learn about the optimal strategies without doing the game theory. In essence, Dawkins maintains, this is what nature has done over eons: vast numbers of strategies have competed with one another, nature's profligacy paying off in the long run in the development of species with optimal strategies, in some sense of the term.

Dawkins uses the concept to show how *group* selection can seem to be taking place in a population when in fact mere *gene* selection can account for what is observed. He writes: "Maynard Smith's concept of the ESS will enable us, for the first time, to see clearly how a collection of independent selfish entities can come to resemble a single organized whole.... Selection at the low level of the single gene can give the impression of selection at some higher level."

The book contains many other provocative examples of peculiar strategies that offer alarming parallels with situations in the world of human politics, some of which remind one of the perils of the current arms race. Indeed, the connection is made explicitly by Dawkins more than once. He refers to "evolutionary arms races" and the survival value of one species' being able to deceive another.

One of the funnier parts of Dawkins' book, although it is dead serious, concerns the evolution of sexuality. To explore how sexuality might have evolved he invents "sneaky" and "honest" gametes (fertilized eggs) and shows how, over many generations, sneaky gametes slowly evolve into males while the honest gametes evolve into females. Along the way he discusses such strategies as the "coy" and "fast" strategies (limited to females), the "faithful" and "philanderer" strategies (limited to males) and the "domestic bliss" strategy. He emphasizes that these are only metaphors and are not to be taken literally (and certainly not anthropomorphically). Taken with a grain of salt, though, they can greatly illuminate the mechanisms of evolution. And many of the strategies find counterparts in the kind of number games I have described above.

As I was preparing this article I had a long telephone conversation with Robert Boeninger in which we tried out various versions of our old and new number games. One idea that intrigued me was to play Underwhelm with no specific target number of points, such as 1,000, in mind. Instead play would be terminated by a convention of another kind. My candidate for the convention was "Stop when the two players' numbers differ by 2." Thus if I say 10 and you say 8, that is the end of the game (and neither of us gets any points on that turn).

Robert and I tried out this version, and we quickly discovered that whenever one of us started losing we would have no option but to go for a stalemate: a nonterminating game. One way for the losing player to do this is to name huge numbers at random, so that they cannot be anticipated and so that the condition for termination is never met. The player who is ahead, having nothing to lose, will cooperate by naming small numbers all the way, thereby gaining even more points and building up even more of a lead. Hence you get a kind of vicious circle in which both players wind up cooperating in a stalemate.

Robert suggested that one way to prevent the stalemate is to add the condition that if either player wins five turns in a row (that is, gets a positive number of points five times in a row), the game is over. This prevents the losing player from going for the stalemate, because such behavior would now ensure a loss. As Robert amusingly pointed out, even if you are behind, you can start to wind things up by trying to win five turns in a row, because by the time those five turns have passed you may be in the lead. My name for this game is Pounce, since it made me feel like a tiger hunting down a giraffe on the savanna, bringing down my prey in one swift, sudden move.

One day, several years after the Undercut episode, my sister Laura and our friend Michael Goldhaber and

Physician, did you miss any of these significant developments in medical science?

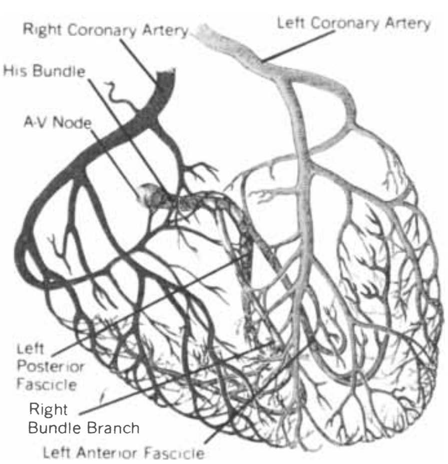
• *Campylobacter fetus* subsp. *jejuni* is associated with a colitis that can clinically and sigmoidoscopically resemble acute idiopathic ulcerative colitis. Stool cultures are in order for *C. fetus* before beginning nonspecific anti-inflammatory therapy.

• Coumarin derivatives cross the placenta. A recent study shows that the consequences for the fetus can be severe. These include embryopathy, stillbirth, and premature delivery.

• Nonsteroidal anti-inflammatory drugs may produce a marked reduction in glomerular filtration rate; with termination of the drug, GFR returns to normal.

• The first documented incident of indigenous transmission of dengue in the continental United States since 1945 has been reported in Brownsville, Texas.

IF THESE ITEMS are familiar you must be a prodigiously energetic or prodigiously lucky reader. With 2,000 or more journals published each year, information that significantly affects patient management all too easily slips by. Textbooks are out-of-date before they are published.



Branches of the right and left coronary arteries supply blood to the A-V node and intraventricular conduction system.

SCIENTIFIC AMERICAN *Medicine* is lucidly illustrated with drawings and photographs. Some examples are seen here and on the facing page.

SCIENTIFIC AMERICAN *Medicine* is the busy clinician's answer to this problem.

Because its authors update SCIENTIFIC AMERICAN *Medicine* every month, it is always current. Because the new information appears in a single source, it is there when you need it.

This 2,500-page, innovative union of publishing and electronic technology is the work of leading scholar-practitioners from Harvard and Stanford. The editors are Edward Rubenstein, M.D., F.A.C.P., and Daniel Federman, M.D., F.A.C.P.

Each month as authors update their contributions, revisions are entered on the magnetic tape on which the text and index are stored. The tape drives high-speed phototypesetting equipment so that subscribers receive about eight new chapters and a new index every four weeks; a bulletin highlights new developments.

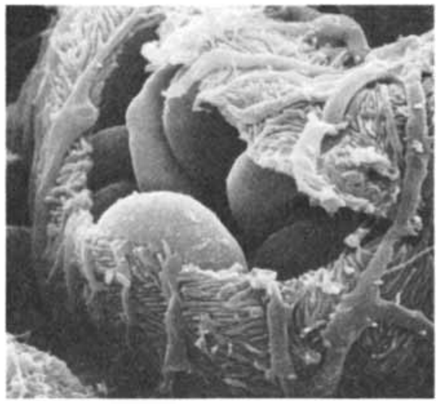
New material replaces old material in the living text, so that the information is there — up-to-date, at your fingertips.

A CME program of eight patient management problems offered over a 12-month period is available at no extra cost. As an organization accredited for continuing medical education, the Stanford University School of Medicine designates this continuing medical education activity as meeting the criteria for 32 credit hours in Category 1 for Educational Materials for the Physician's Recognition Award of the American Medical Association, provided it has been completed according to instructions. This program has been reviewed and is acceptable for 32 prescribed hours by the American Academy of Family Physicians. This program has been approved by the American College of Emergency Physicians for 32 hours of ACEP Category 1 credit.

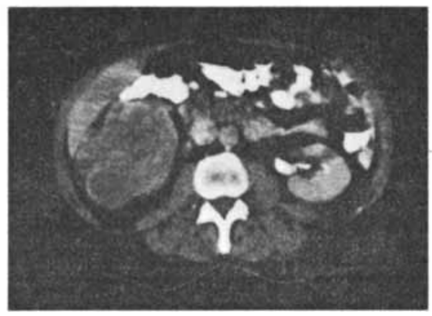
Trial Offer

We invite you to try SCIENTIFIC AMERICAN *Medicine* — for two months at no cost. Send us the coupon and you will receive the two-volume text and two monthly updates. You may also take a CME test for credit. At the end of 60 days, if you decide to continue the subscription, we will bill you for $220 for the full 12 months (renewal is currently $170); otherwise return the two volumes.

Please mail the coupon today and let us take the hassle out of keeping up.



Scanning electron micrograph of a glomerulus shows a portion of a capillary loop.



Abdominal computed tomogram reveals large renal carcinoma replacing part of right kidney.

THE DISTINGUISHED AUTHORS AND THE FIFTEEN SECTIONS OF SCIENTIFIC AMERICAN ***MEDICINE***

1. Cardiovascular Medicine

Edgar Haber, M.D., Harvard Medical School and Massachusetts General Hospital

E. William Hancock, M.D., F.A.C.P., Stanford University School of Medicine

Roman W. DeSanctis, M.D., Harvard Medical School and Massachusetts General Hospital

Adolph M. Hutter, Jr., M.D., F.A.C.P., Harvard Medical School and Massachusetts General Hospital

Eve Elizabeth Slater, M.D., Harvard Medical School and Massachusetts General Hospital

2. Dermatology

Eugene M. Farber, M.D., Stanford University School of Medicine

Elizabeth A. Abel, M.D., Stanford University School of Medicine

3. Endocrinology

Daniel D. Federman, M.D., F.A.C.P., Harvard Medical School

4. Gastroenterology

Gary M. Gray, M.D., Stanford University School of Medicine

Peter B. Gregory, M.D., Stanford University School of Medicine

John Austin Collins, M.D., Stanford University School of Medicine

Douglas Wilmore, M.D., Harvard Medical School and Brigham and Women's Hospital

5. Hematology

Stanley L. Schrier, M.D., Stanford University School of Medicine

6. Immunology

John David, M.D., Harvard Medical School and Harvard School of Public Health

Willy F. Piessens, M.D., Harvard Medical School and Harvard School of Public Health

7. Infectious Disease

Thomas C. Merigan, M.D., Stanford University School of Medicine

Morton N. Swartz, M.D., F.A.C.P., Harvard Medical School and Massachusetts General Hospital

Cyrus C. Hopkins, M.D., Harvard Medical School and Massachusetts General Hospital

Adolf W. Karchmer, M.D., F.A.C.P., Harvard Medical School and Massachusetts General Hospital

Robert H. Rubin, M.D., F.A.C.P., Harvard Medical School and Massachusetts General Hospital

Harvey B. Simon, M.D., F.A.C.P., Harvard Medical School and Massachusetts General Hospital

Peter F. Weller, M.D., F.A.C.P., Harvard Medical School; Beth Israel Hospital, Boston; and Brigham and Women's Hospital

8. Intensive and Emergency Care

Edward Rubenstein, M.D., F.A.C.P., Stanford University School of Medicine

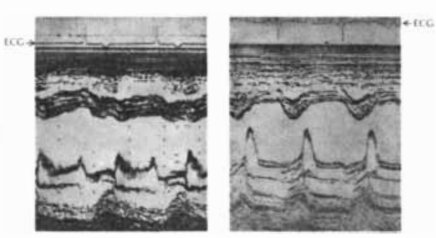
9. Metabolism

George F. Cahill, Jr., M.D., Harvard Medical School, Howard Hughes Medical Institute; and Brigham and Women's Hospital

10. Nephrology

Roy H. Maffly, M.D., Stanford University School of Medicine and Palo Alto Veterans Administration Medical Center

William Bennett, M.D., Health Sciences University, Portland, Oregon



Echocardiograms from patients with aortic regurgitation.

11. Neurology

Robert W. P. Cutler, M.D., Stanford University School of Medicine

12. Oncology

Saul A. Rosenberg, M.D., F.A.C.P., Stanford University School of Medicine

13. Psychiatry

Ned H. Cassem, M.D., Harvard Medical School and Massachusetts General Hospital

14. Respiratory Medicine

Eugene D. Robin, M.D., F.A.C.P., Stanford University School of Medicine

15. Rheumatology

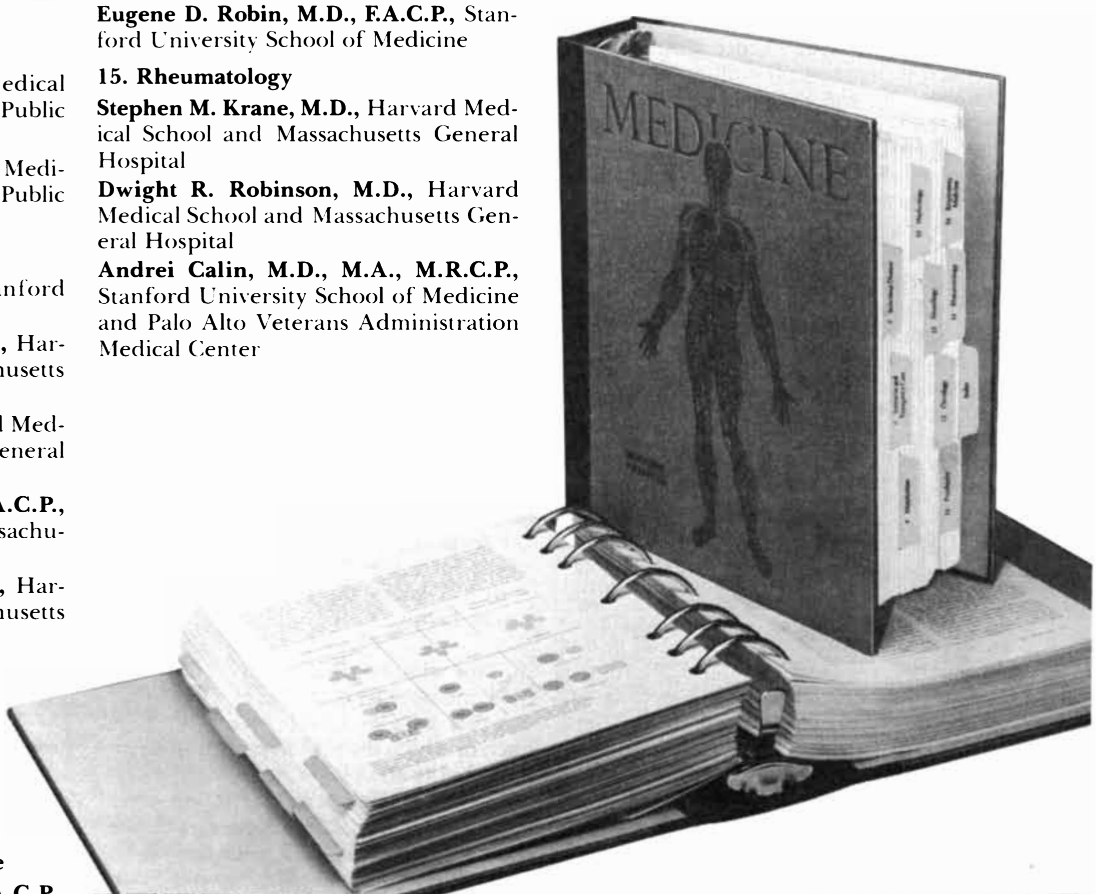
Stephen M. Krane, M.D., Harvard Medical School and Massachusetts General Hospital

Dwight R. Robinson, M.D., Harvard Medical School and Massachusetts General Hospital

Andrei Calin, M.D., M.A., M.R.C.P., Stanford University School of Medicine and Palo Alto Veterans Administration Medical Center

Order by Phone

You can order SCIENTIFIC AMERICAN *Medicine* by telephone. **Please call this toll-free number: 1-800-345-8112 (in Pennsylvania, call 1-800-662-2444);** you will be billed after your subscription begins. Toll-free calls are acceptable only for orders placed in the continental United States.



SCIENTIFIC AMERICAN MEDICINE

415 Madison Avenue, New York, N.Y. 10017

Please enroll me as a subscriber to SCIENTIFIC AMERICAN *Medicine.* On receipt of this coupon you will send me the advanced two-volume text described in your announcement and update it regularly by sending me new monthly subsections. I understand that the price of $220 for first year of service is tax deductible, as is current renewal at $170. If I am not entirely satisfied, I may cancel at any time during the first 60 days, returning all materials for a *complete refund.*

☐ Please enter my subscription for SCIENTIFIC AMERICAN *Medicine*
☐ I shall also enroll in the CME program
☐ I enclose a check made out to SCIENTIFIC AMERICAN *Medicine* for $220*
☐ Bill me ☐ VISA ☐ MasterCard

Expiration Date__________ Account Number__________

*Please add sales tax for California, Illinois, Massachusetts, Michigan and New York

Name__________

MD Specialty__________

Address__________

City__________State__________Zip__________

R1

Please allow six to eight weeks for delivery. All payments must be in U.S. dollars. CME is available outside the U.S. and its possessions for an additional $65. Surface routes used unless airmail delivery is prepaid.

I were having lunch in the Peninsula Creamery and jotting down various trivia on napkins, as was our wont, and somehow it came to us to play a number game involving *three* people. We decided that on each turn each of us would choose a number in a certain range, and since it seemed uninteresting simply to let the biggest number win, and scarcely any more interesting to let the smallest number win, the idea of having the middle number win suggested itself. We decided that on each turn only the "most mediocre" player's score would be allowed to increase. It would increase, of course, by the mediocre number; the other players' scores would stay fixed. (A bit of a problem came up when two players chose the same number, but we found a makeshift way out.)

Thus at the end of, say, five turns we would compare our scores, and the highest one—no, wait a minute. Why should we let the *highest* score win? To do that, after all, would be contrary to the spirit of each turn. We saw quite clearly that, if the spirit of the whole was to be consistent with the spirit of its parts, then the player whose score was in the *middle* should win. We called the game Mediocrity, but I have since renamed it Hruska.

This name is inspired by a famous remark by Senator Roman Hruska of Nebraska. At the time we invented the game President Nixon was trying to get Senate confirmation for his appointment of G. Harrold Carswell to the Supreme Court. In a radio interview defending Carswell against his critics, Senator Hruska came out with the following profundity: "Even if he [Carswell] were mediocre, there are a lot of mediocre judges and people and lawyers. They are entitled to a little representation, aren't they, and a little chance? We can't have all Brandeises and Frankfurters and Cardozos and stuff like that there." Alas for mediocrity, Carswell was not confirmed. But it worked out fine for Hruska, who shall forevermore be known as a champion of mediocrity—and stuff like that there.

Speaking of champs, after eating our sandwiches and consuming our milkshakes (served in the Peninsula Creamery's old-fashioned metal containers) the three of us sat in our booth and played a few rounds of this quirky game, and what came to our minds but the inspired idea of determining the world champion of *mediocrity.* Therefore we totaled our scores over several games to see who had come out highest. Highest? Again something seemed wrong. The pervasive spirit of mediocrity that had settled on us like a heavy smog urged us to deem champion not the player who had won the most games, not the player who had won the fewest games, but the player who had won the *middlemost* number of games. So it was, and I forget who had the honor. This may be appropriate.

At that point a general principle seemed to be emerging that created a hierarchy of levels of Hruska. To win at Level Two (that is, our "championship" level) it is best to be a mediocre player at Level One (the single-game level). This means that whereas before it was desirable to be *extremely* mediocre at choosing mediocre numbers, now it is desirable to be *mediocrely* mediocre at choosing mediocre numbers. How perverse! How wonderful! How wonderfully perverse! It fits in with a general principle of perversity, a Zenlike principle, that applies to many aspects of life: you wind up a loser if you try too hard.

After the session where the game of Hruska was born I worked on other versions of it, trying to polish it and make it elegant. Let me present the rules as they now stand.

The main problem is how to avoid ties, not only ties at Level Zero but also at all higher levels. My current best solution is the following: Let each player have a slightly shifted range with respect to the other two players. More concretely, let player *A* pick integers from, say, 1 to 5. Then players *B* and *C* will have staggered ranges: *B* picks numbers of the form $n + 1/3$ and *C* picks numbers of the form $n + 2/3$, where n runs from 1 to 5. Clearly there can be no ties at Level Zero.

Now what happens at Level One? Remember that a Level One game consists of five Level Zero games, in each of which the middlemost number is awarded to the player who chose it and the other two players should get zero. Well, the first part of this scoring scheme is fine, but the second part has to be modified slightly in order to avoid ties at higher levels. Suppose the numbers chosen are *A*, 3; *B*, $2\frac{1}{3}$, and *C*, $4\frac{2}{3}$. Having the middle number, *A* gets three points. *B* and *C*, however, do not get zero points each but the closest possible approximation to zero they can, given their staggered ranges. Thus 1/3 point goes to *B* and 2/3 point to *C*.

The reasoning behind this goes as follows. After five turns each player has had five numbers of the same *form*. Player *A*'s five pure integers will add up to a pure integer; player *B*'s five numbers of the form $n + 1/3$ will add up to a number of the form $n + 2/3$, and player *C*'s five numbers of the form $n + 2/3$ will add up to a number of the form $n + 1/3$. Thus at the next level up *B* and *C* have exchanged roles in terms of the form of their numbers. As a result the three total scores at the new level are all of different form and cannot tie. Hence there will always be a most mediocre Level One score: a winner.

If we now go on to consider a game at Level Two, we must award points to each Level One game. The winner of a Level One game gets, of course, that middling number of points, whereas the other two players once again receive the closest approximation to zero possible in their respective forms. For player *A* this means, as before, exactly zero points. For *B*, however, it now means 2/3, and for *C* it means 1/3. Five games of Level One constitute a game at Level Two. The heretofore tacit "principle of uniformity of levels" obliges us to add up the five Level One numbers to get Level Two scores. Needless to say, the same reasons as before will prevent tie scores, and so there will always be a Level Two winner.

The same general principle will of course allow us to extend the game of Hruska to *any* number of levels. One game of Level $n + 1$ Hruska consists of five Level n games. The winner of each Level n game is awarded the Level n score, and the other two players get the minimum amount (0, 1/3 or 2/3) of the form of their scores at that level. The five Level n numbers are added up to yield totals for the three players, and the middlemost player wins. Actually there is nothing sacred about always having *five* Level n turns in a Level $n + 1$ game; the "width" could just as well be four or even two. (Multiples of three must be avoided, because after three moves all three players have scores that are perfect integers, thus making ties possible.) With a width as narrow as two this allows a very deep (that is, many-level) game to be played much more easily. For instance, with width two a five-level game of Hruska requires only 32 Level Zero turns, whereas with the standard width of five merely three levels of depth will require 125 Level Zero turns.

Moreover, there is nothing sacred about the Level Zero choices' having n bounded by 5; n could run from 1 to infinity! This is just one of many possible variants of Hruska. I can testify that the strategy for playing even Level Two Hruska gets mighty confusing very quickly. I have played Level Three Hruska on a couple of occasions and found it completely beyond my reach. I find this both fascinating and frustrating. Think, however, what it implies about world politics if such simple games as the ones described in this article can be baffling. How much more complex are the "games" of international bargaining and bluffing and warmaking. All the tricky concepts I have discussed here have their counterparts in international politics—only "squared," so to speak. As one watches these huge themes being played out on the world stage one can hardly help but feel like a single cell in some vast organism that long ago became fixed in its evolutionary strategy, whose consequences one can only observe, hoping all will turn out for the best.

BOOKS

A composite physicist, the creole languages, marvelous mammals and tunnel technology

by Philip Morrison

NIGHT THOUGHTS OF A CLASSICAL PHYSICIST, by Russell McCormmach. Harvard University Press ($15). It was the 50th month of the First World War. The old professor of theoretical physics was losing weight seriously; official meatless weeks came again and again, and the compensatory ration was only a little extra bread, grayer than ever, to spread with turnip jam. The day's cup of milk went to Maxwell, the stray cat that had adopted Herr Professor Jakob and had in return been given that unpopular English name, perhaps because it scratched its back against the two volumes of James Clerk Maxwell's *Treatise on Electricity and Magnetism* on the lowest shelf. The Geheimrat director of the physical institute had tried for a while to dub the cat Röntgen, surely more fitting for a German cat in this time of war against the English and even their ideas. But Clerk Maxwell's photograph hung highest on the wall of the professor's study, just above the larger central depiction of noble Hermann von Helmholtz. With the precisionist experimenter August Kundt, those two had been the professor's revered teachers in Berlin and Strassburg after his return to his studies, a wounded veteran of the 1870 Franco-Prussian War.

Victor Jakob—no Jew in spite of his name, which had no doubt somewhat affected his career—is the man we come to know in this remarkably evocative book (whose period photographs, documents and drawings add visual depth to the intimate narrative). The tone is somber, even melancholy. We meet Jakob late in 1918 at a "patriotic evening" to sell war bonds in his university town, telling of his brave experiences of the year of Bismarck's great victory. We follow him internally: his thoughts of the past, his diary of 1870, a few dreams, a visit to the theater, a lecture to students on acoustics (the earnest young women of the wartime student body seem strangers to the lecturer), his memories of Gustav Hertz and Paul Drude, and the rise of a new atomic physics whose hypotheses seem as wild as the unending war. Beset by his sense of failure, by ill health, even by overt ridicule, the professor, patriot and internationalist at once, shocked again and again by the obituaries of fallen physicists each month in his *Physikalische Zeitschrift,* trudges high into the dark hills above the town to place the heavy old revolver taken out of his desk drawer to his open mouth.

There is no such person as Victor Jakob even in the careful biographical dictionary of Poggendorff. He is a composite physicist, constructed and explicitly documented "from real voices and events of the past." If Jakob prefers the mature Bach with its profound structure to the emotional chorales of Bach's youth, their passion so easily joined to a cult of German national fervor, it is because the real Göttingen theorist Woldemar Voight so wrote in 1911. Drude's real suicide is echoed by Jakob's own. The bright little interior painted by Menzel—each scratch in its place—becomes Jakob's example for the physicist's idea of nature, as indeed it was so discussed by Felix Auerbach. He muses on James Chadwick and Peter Pringsheim doing physics as prisoners of war; so they did.

Jakob thinks of the trenches where dead hands protrude from the walls, one layer of German bodies and then a layer of English ones. He concludes that even physics cannot always help, the very words from Einstein's own letters of the period. His entire consciousness, not to say the entire social and material milieu built around him in a town not far from Bonn, is elegantly assembled from the documents of 50 years of physics and physicists in Germany. The author, a historian of science at Johns Hopkins, has made a brilliant mosaic of authentic and glittering tiles, *echt Deutsch,* although untinted by Hegel.

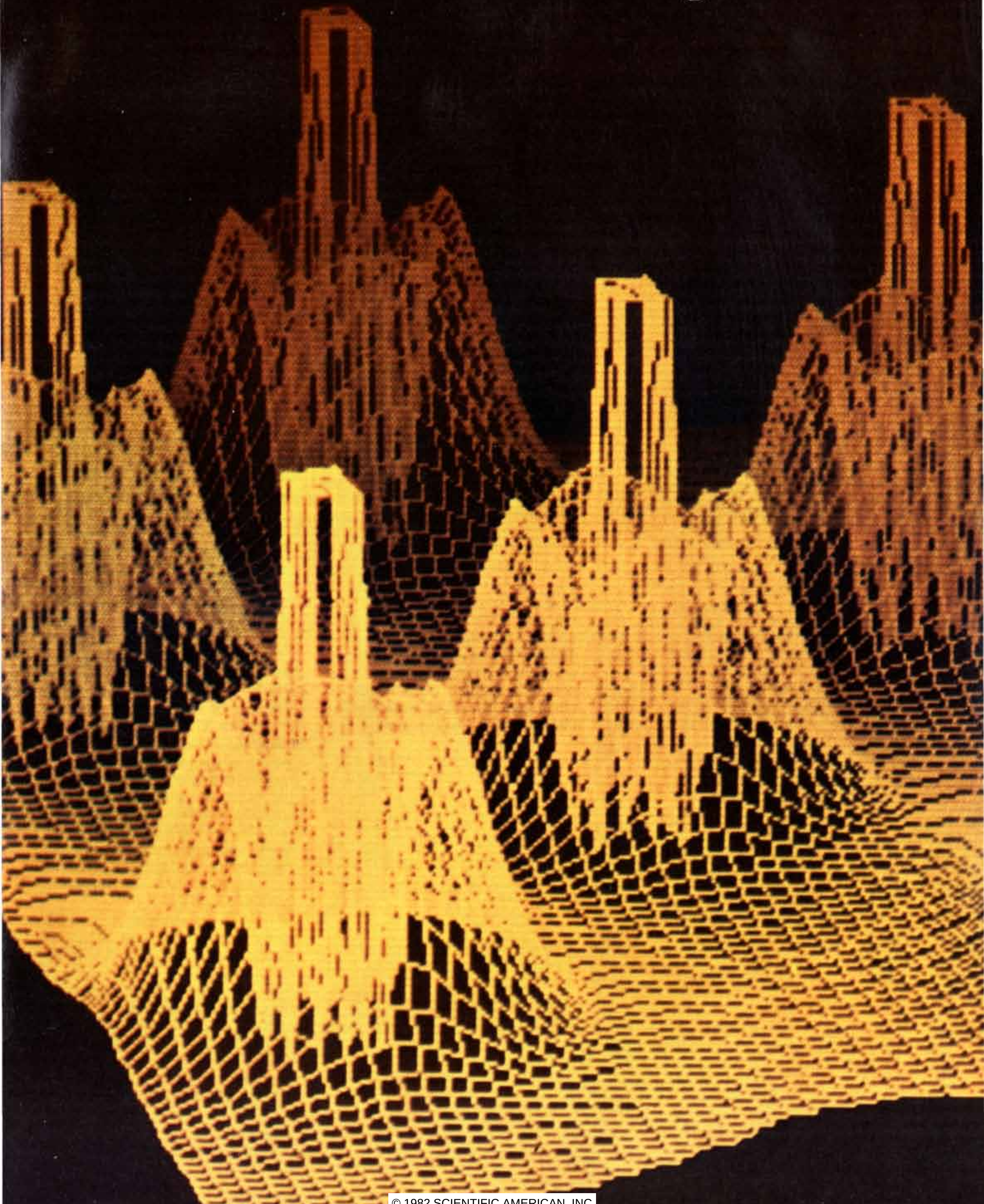
A reader, particularly one familiar with theoretical physics and attentive to our uneasy times, cannot fail to enter a few reservations to this compelling genre, a serious counterpart of myth-building dramatizations on television. *Night Thoughts* is more tendentious than history plain. One does not see any page, real or pastiche, of the *Annalen der Physik,* where the physics of Professor Jakob and his colleagues appeared, nor any piece of Jakob's blackboard notes so humiliatingly erased by Jakob's young atomic-physicist colleague. No doubt Jakob was not a forest giant among physicists but a mere ground creeper. Yet it is likely that his everyday work remains closer to the life of physics today than the tone of the book admits. The inner narrative amounts to a sunny time past, an era of romantic idealism and a slow approach to certainty, to be replaced by the implacable darkness of rampant nationalism, brandishing the weapons fashioned by the physicists of atomic uncertainty. But yesterday's history, like today's, is more ironic than that. Self-centered Europe was at peace, it may be, after 1870. The classical Maxim machine gun guaranteed that on other continents life was less assured.

It was the grand concept of the world ether, a complex pervasive medium of electromechanical nature, around which Jakob's deepest research hopes centered. His eulogist would remark that his ambitious concepts retain no modern value; only his devoted teaching and a few small technical triumphs would be remembered. Yet the world ether is still visible, in the strange form of a fluctuating, energy-rich vacuum from which all the world, so they speculate today, once formed. Perhaps people in the advanced countries of the north remain caught in the spell of a cold gray sky turning to a summer blue, with a sense of complete replacement in due season. Most of the rest of the world, physics too, enjoys a less consistent mental climate, a single day bringing both storm and sun.

Jakob's acoustic gun-ranging experiments in the First World War proved less useful to the military than the electromagnetic novelties of the radar magnetron at the start of the Second War, but the physics of both was classical. The issue is open; the consequences of the new power of a physics at once relativistic, atomic and probabilistic remain in dark recesses of the future.

ROOTS OF LANGUAGE, by Derek Bickerton. Karoma Publishers, Inc., 3400 Daleview, Ann Arbor, Mich. 48103 ($24.95). Faced with "a catastrophic break in linguistic tradition," human beings invent languages; they must have language to survive. What happens when people drawn from widely separated language groups are thrown into a life of single-crop agriculture under the rule of a small European minority? Soon there arises a new auxiliary language of contact, native to no one, known as a pidgin. It is followed in time by an expanded form known as a creole, now the native speech of a larger

The Surface Contingent



The Surface Contingent

A recent finding at the General Motors Research Laboratories has changed scientific thinking about the behavior of electrons in metal surfaces. This discovery provides a greater understanding of the fundamental physical processes involved in such surface events as adhesion, corrosion and catalysis.

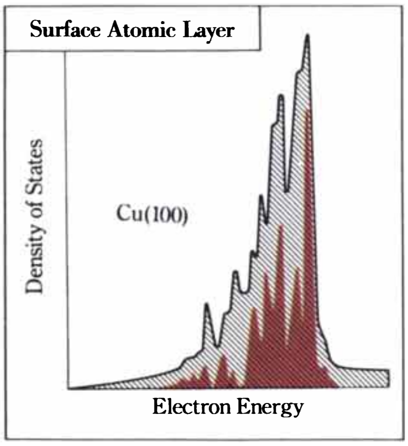
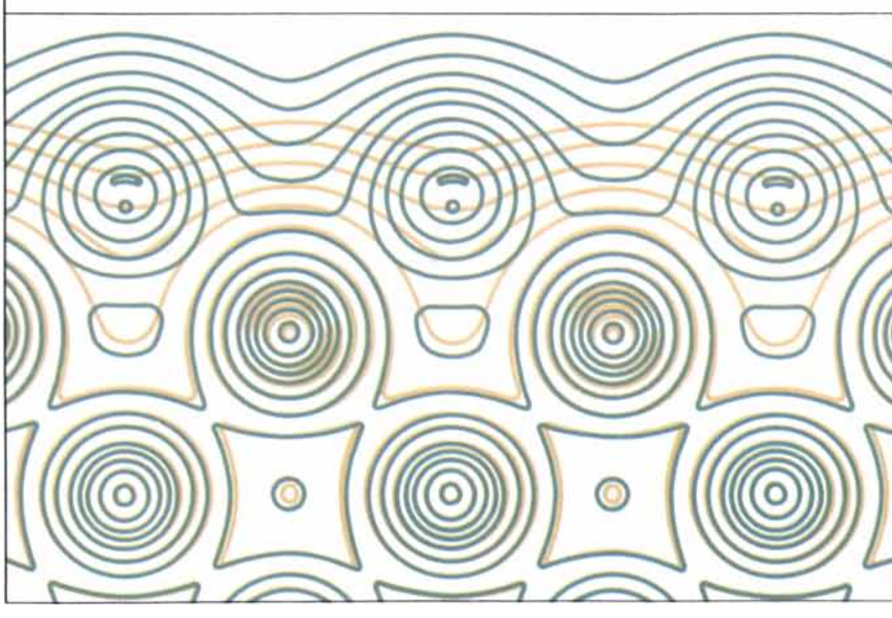


Figure 1: Energy distribution of electrons in outermost atomic layer. Red area indicates electrons in surface states.

Figure 2: Two electron density contour maps of the cross-section of a Cu(100) surface. One map shows a clean copper surface (tan); the other shows a nitrogen-covered copper surface (green).



CONVENTIONAL scientific thought treats virtually all of the valence electrons found in the surface atomic layer of a metal as if they are free to roam throughout the metal's interior. The work of three physicists at the General Motors Research Laboratories suggests otherwise. Through calculations confirmed by experimental data, the theorists have shown that more than a quarter of the valence electrons in the top atomic layer of some metals are effectively trapped in the surface. The presence of so many "surface state" electrons must be considered when analyzing physical and chemical surface phenomena, including such surface events as oxidation leading to corrosion.

Drs. John Smith, Jack Gay and Frank Arlinghaus applied their theoretical analysis to the (100) surface of five metals: copper, nickel, silver, rhodium and palladium. They made bold predictions concerning the percentage of electrons in the surface atomic layer to be found in surface states: Cu(36%), Ni(23%), Ag(23%), Rh(23%) and Pd(19%). The ratio of the red area to the hatched area of figure 1 gives the percentage for copper.

Electrons in surface states are not only abundant, but also highly localized on the surface. Chemisorption on a metal is also confined to the surface region. Figure 2 shows what happens in the case of nitrogen chemisorbed on copper. The two contour maps coincide except in the surface layer, where the interaction is largely exhibited. Localization of the interaction holds for the chemisorption of other gases, including oxygen in the initial stage of metal oxidation. These observations led the physicists to conclude that surface states are important in chemisorption.

One way to probe electrons in surfaces is to chemisorb atoms on a clean metal surface and look for changes in photoemission spectra. Such an experiment was performed at GM for fractional monolayers of nitrogen, oxygen and sulfur on Cu(100). The dominant change in the photoemission spectrum was the disappearance of a large peak whose shape and energy location was independent of the chemisorbed atom. It was of special interest that the shape and energy location of this peak was nearly identical to the envelope around the surface state peaks in figure 1. This suggests that surface state electrons play a major role in the chemisorption process.

THE THEORETICAL advance at the heart of the discovery is the "Self-Consistent Local Orbital (SCLO) Method" for solving the Schrödinger equation. This new mathematical method was devised by the GM theorists to handle the classic dilemma posed by the self-consistency requirement. The characterization of electron behavior used to complete the equation must be consistent with the behavior predicted by the equation. In other words, one almost needs to know the answer in order to make the calculation.

Self-consistent solution of the equation for a metal surface is made exceedingly difficult by the three-dimensional nature of the electron density distribution. The theorists dealt with this challenge successfully by dividing the electron density distribution into two parts—the first part due to overlapping atomic density distributions; the second part equaling the difference between this atomic contribution and the exact density distribution.

One of the more stringent tests of the accuracy of the SCLO method was an angular photoemission experiment conducted by Heimann et al. at the University of Munich, subsequent to publication of the GM research. The German research team confirmed a prominent surface state band predicted by the three GM physicists. This was the first time a surface state band on a solid had been calculated prior to its being seen experimentally. The SCLO method makes possible something that could not be done before—accurate prediction of the actual behavior of electrons whirling around nuclei at the surface of a metal.

"The large body of surface states we found on metal surfaces," says Dr. Smith, "may be a controlling factor in many physical and chemical surface phenomena. By replacing conjecture with calculation, the new surface theoretical methods give us the means to make major steps forward in the analysis of surface and interface properties."

THE MEN BEHIND THE WORK

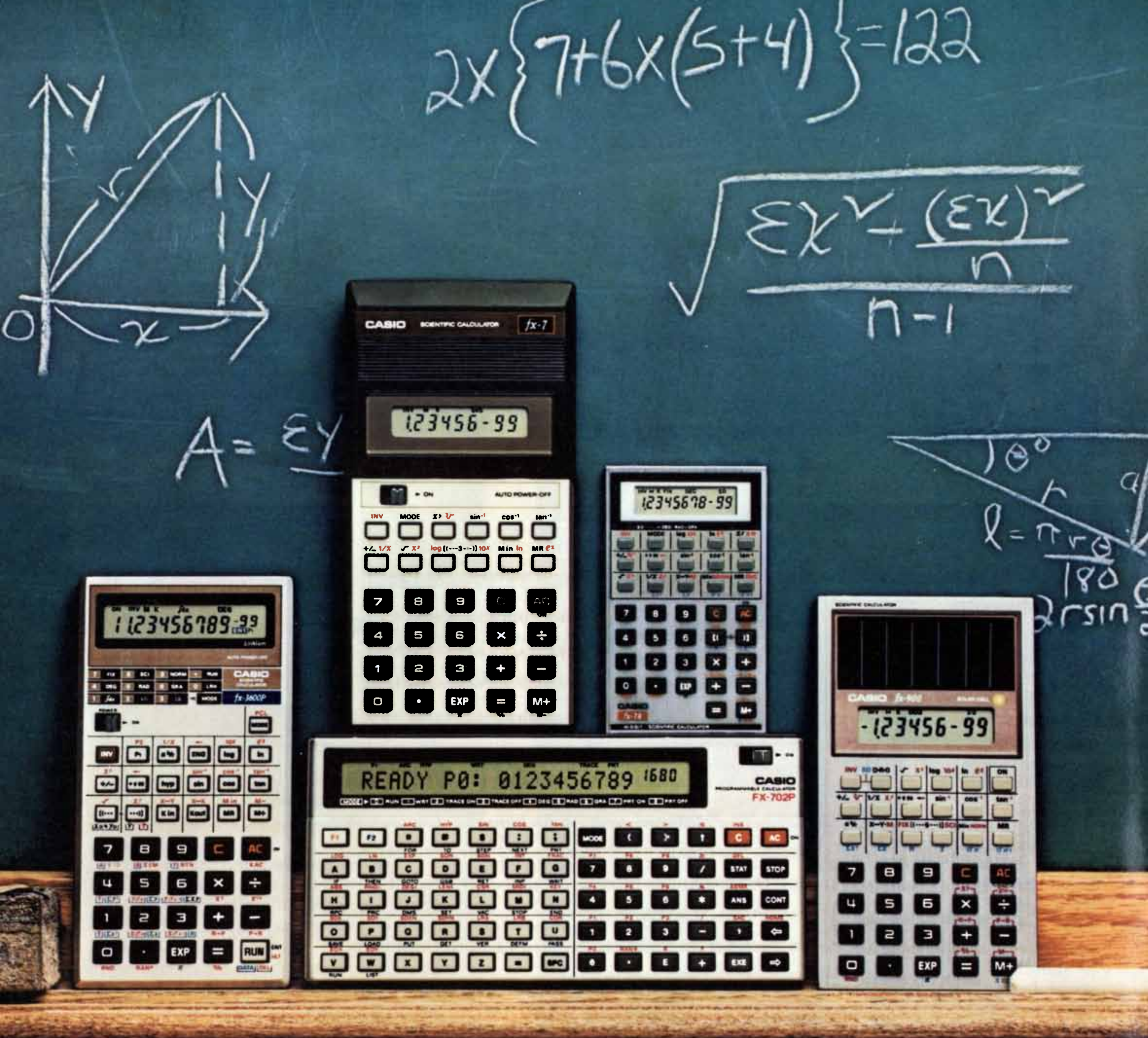
Drs. Smith, Gay and Arlinghaus are theorists in the Physics Department at the General Motors Research Laboratories.

John Smith (center) and Jack Gay (right) received doctorates in physics; Smith from Ohio State University and Gay from the University of Florida. Frank Arlinghaus received his Ph.D. in physical chemistry from the Massachusetts Institute of Technology.

John Smith, leader of the GM solid state physics group, did postdoctoral work at the University of California in La Jolla. He joined General Motors in 1972. Frank Arlinghaus and Jack Gay joined the corporation in 1964 and 1965, respectively.

Each member of the team brings to the project a different expertise: Smith in surface physics, Gay in solid state theory, and Arlinghaus in bulk band structure calculations.

GM General Motors
The future of transportation is here



Casio takes you from high school to high tech.

It's a fact. Nobody has a more complete line of state-of-the-art, scientific calculators than Casio.

Even our basic FX–7 model gives you 23 functions—everything you need to up-grade your grades.

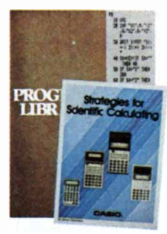
Our FX–900 model is the lowest priced solar scientific calculator under the sun. And not only will its 41 functions save you lots of time and energy, but because its solar cell operates in any normal light, it'll save you from ever having to buy batteries.

And while our credit card size model FX–78 is among the smallest calculators of its kind, with its 41 functions, you have to give it credit for thinking big.

If you want to get with a programmable, our FX–3600P gives you 61 functions and 38 programming steps at only a small step up in price.

And when you're ready for the step up to a computer, our FX–702P will make it easy. It uses BASIC programming language. And its 1680 programming steps and optional peripherals can take you to the highest tech possible using a hand held computer.

High tech at low prices. That's something you've always been able to count on from Casio.



A detailed applications manual comes with each Casio scientific calculator.

CASIO®

Where miracles never cease

Casio, Inc. Consumer Products Division: 15 Gardner Road, Fairfield, N.J. 07006 New Jersey (201) 575-7400, Los Angeles (213) 923-4564.

community. Generally a creole differs enough from any parent tongue to be a new language; superficial resemblance is usually found in the vocabulary, but the syntax may be strongly distinct.

This cultural disaster and this burst of human creativity has been seen again and again under colonial and slaving regimes. A couple of dozen examples worldwide are cited in this witty, powerfully argued, personal treatise by a linguist from the University of Hawaii. Here and there the technical prerequisites of Professor Bickerton's book go beyond the resources of a general reader, but the current of the argument is strong enough to carry the reader along with it.

We cannot say what year Italian came into existence; there was surely some continuum flowing from Latin along a river delta of dialects. We can be sure, however, that before 1880 there was no Hawaiian creole, because the date when new tariff laws allowed the duty-free importation of Hawaiian sugar to the U.S. was 1876. In a few years the planters brought into the islands a multilingual community of workers from China, Portugal, Korea, the Philippines and Puerto Rico, people who soon greatly outnumbered the Hawaiian and English speakers alike. Careful field study has shown that a complex and quite homogeneous new language, Hawaiian Creole English, had originated between perhaps 1910 and 1920. That first creole-speaking generation had invented a language. They did not speak the antecedent pidgin; theirs was the task of producing rules of syntax for which their parents had no counterpart.

The first half of this book examines in detail the evidence for distinct rules of grammar in the new island tongue. The evidence is hard to follow, and not just because of its intricacy; any nonlinguist knows a good deal of grammar through using it, once the general nature of the rules is pointed out. The difficulty is that the items are small details, the statistics are few and far between, and the generalizations, although plausible enough, do not always seem demanded by the data. It is here that linguistic experience is needed.

An example is provided by articles. One Japanese speaker in pidgin used only a tenth of the indefinite articles English would have required. A Filipino used in his pidgin sentences many more. Pidgin is variable. In the Hawaiian creole, however, the uses of the definite article (*da boi*), the indefinite article (*wan blaek buk*) and zero marking (*dag smat*) are strictly governed by the nature of the references. Yet the rules do not resemble the rules of English or of any other language in the substrate. A dozen other such features of Hawaiian Creole English are presented.

Now enters the argument. In a long summary of creoles worldwide, admittedly only a first careful look, Bickerton concludes that their syntaxes show deep resemblances, and the sooner the samples come after that first inventive generation the closer the resemblances are. Creoles everywhere are built on one and the same set of rules from the basic constituents necessary to a still unfixed pidgin. To arrive at this conclusion many apparent exceptions have been argued away on a variety of knowledgeable grounds. If we accept the inference, however, it appears that an incredible play of chance has led people forced to construct a new tongue out of elements as disparate as Yoruban and Dutch, without common model or interchange, across continents and centuries, to follow the same structural path.

There is a way out of this scarcely believable convergence. Here at last we are in touch with an innate human bioprogram that determines the form of all human language, one that unfolds normally whenever it has a chance. All children need to acquire a language and are marvelously able to do so. In a fascinating chapter Bickerton proceeds to show that the child's language learning too can be interpreted along such lines; the examples chosen include the markers for the specific and the nonspecific, the relation between single events and regular events, a feature related to tense, and a couple of others. The data are recent. Not very rich, they include ingenious studies of the acquisition by children of English articles, of French, Turkish and Italian verbs, of questions in English and so on. No child learns language in a vacuum; the heart of this inquiry is the steps children take toward acquiring language. The chapter concludes with a tentative confirmation: what the creole inventors did the children can learn effortlessly. Once that simple program has evolved to novelty, the children more gradually conform to the model they hear.

The last chapter, at once a polemic and a direct line of argument, enters on the speculative. How did language begin? The key is a rejection of that simplistic question. "Anything as complex as language ... must consist of a number of interacting systems." The question is how far along the road to useful language has the would-be speaker come? Language is not the mere elaboration of an animal call system, a set of alarms, the squeaks and grimaces so well distributed among other animals. It is first of all abstraction. There is some system of long-term storage of perceptions, packed away but recallable because they are stored not as whole images but as particular features.

Next comes the ability to recombine such stored features; they are by now concepts, somehow defined neurologically. We do not see our world anew from second to second; rather we interpret it as we go from previous categorization. A dot surrounded by 10 rays, one of them longer than the others, is universally understood as a flower; it is reproduced in the book. Now that cognitive map must be coded, taking a gambler's—or a scientist's—chance of correctness. No more is needed. Approximation is often success enough, and speed is plainly its fruit.

Then the concepts must be spread socially. Language is the means to that end; the species must go beyond the here and now to "achieve not just predictability but the dissemination of predictability." Grammatical marks for tense, continuation and the counterfactual are the most natural and necessary of distinctions. These structures are subject to natural selection. The creole builders had such programs straight, as have our children. Long ago our hominid ancestors went beyond the simple concepts clearly held by the unspeaking apes. What came just before language was not a cruder form of communication; that skill was old indeed. What came first was the mind's power to compute, to judge the probable consequences of events not at hand. The language bioprogram arose out of selection for such powers. When it was that our forebears mastered the communication of abstractions we cannot now date; probably the rise of our own species, *H. sapiens,* and its proliferation over the past 500 centuries hold the answer.

Altogether bold, fresh and incisive, this by no means easy book is a find, if its argument is still less than validated. Bickerton is no detached scholar; he sounds and re-sounds one ironic theme out of the cruel history of creolization. Just as in the Bible honey came out of the rent carcass of the lion, so human self-knowledge may emerge from insight into the creativity of the victims of "the Cartesian savagery" of the colonialists.

EAST AFRICAN MAMMALS: AN ATLAS OF EVOLUTION IN AFRICA, by Jonathan Kingdon. Volume III, Part C and Part D (Bovids). Academic Press (each part $99.50). "An initial explosion of excited impala," the big gregarious gazelles leap and jink before you in sudden magnificent spectacle. Once again the large pages of his handsome pair of books put the pencil and keen eye of this artist-mammalogist before any fortunate reader.

It is a full decade since the first volume of Kingdon's ambitious work appeared. That told of primates and a few other oddities. Two books on the smaller fry followed. Then came Volume III, its first book on carnivores, its second on the larger mammals of that world. The seven-book set closes now with the horned ungulates, 15 years well spent in "the contemplation of physical beauty in mammals;... a reward in itself."

Our reward is more tangible; first is the apt text, a profile of each species among nearly 50 of the bovid family, creatures almost heraldic for our visions of East Africa. The exciting drawings are here too: poised skeletons, carefully drawn full-page life portraits species by species, with pages and columns and odd corners turned by pencil into informal evocative field views or anatomical highlights. A dozen forms of red duikers parade across a color plate. Perhaps the eye is most enchanted by one plate of genuine heraldry: pair by pair confront each other, their horns locked, the fighting males of 14 species, from waterbuck to sheep. Such interplay of information and design is the delight of Kingdon's work.

How did such hoof-horn variety arise? Nowadays the ruminants are mostly cellulose eaters, thanks to their active microbial symbionts, but the trick is not universal among them. The beginning of their rise was probably a "security strategy": fill up when you can and then go somewhere in the tall grass to digest in peace—good advice for the avoidance of predation and for the economical utilization of animal energy. (The hippopotamus follows a submarine version of that strategy.) In this niche the bovids grew, first little seen in the fossil record because of predation. Then they radiated; in Africa in particular they diversified, some as solitaries to the forest glades, others to the watercourses and the sumplands. In good time they included the high-density grazers of the open green grass, wildebeest (1.3 million now in Serengeti alone), impala, hartebeest and the little gazelles, all the horned herds that crowd our images of those great reserves.

Tsavo and Serengeti are known by name everywhere; they are the largest of 145 mapped conservation areas. Kingdon has himself visited 105 of those areas. He returns with his trove of image and understanding, and with a clear plea. The cattle, sheep and goats on which so many human beings depend, from the Masailand to Amazonia and Texas, are only three West Asian bovid species. From all their worldwide lucrative exploitation no support has flowed to the few scientists who have tried to understand in detail the origins and nature of the bovids of Africa, where 20 million years have perfected an intricate system to unite herd and sun through grass. War and poverty and misgovernment now threaten the elegant adaptations of the grasslands; the local people cannot bear the burden of the world's future alone. Cropping the wildlife offers a direct benefit; both long-term conservation and shorter-term yield might be served by timely investment in land and in research. The light airplane has made wildlife stocks a quantifiable resource; radio tracking and immobilizing drugs are powerful tools for rational management.

This final book includes a set of useful appendixes. The reserve maps constitute one appendix; others include an essay on conservation, one on animal management in captivity (by the London Zoo's chief veterinary officer, who treats all the orders of mammals) and a summary of diseases and parasites.

The heaviest antelope is the eland, which can weigh a ton. "In the prehistoric paintings of Kondoa and in the rock art of southern Africa, the eland is the most frequently represented of animals, which is perhaps more a measure of the bushmen's preoccupation with the animal than an indication of its abundance. As with the horses, cattle and bison represented in European cave paintings, still lives of seventeenth-century Holland and advertisements in the modern world, beauty lies not in the eye but in the stomach of the beholder." A certain inconsistency may be perceived in this comment by our delighted artist, but every reader will sense a rueful bond of understanding.

TUNNEL ENGINEERING HANDBOOK, edited by John O. Bickel and T. R. Kuesel. Van Nostrand Reinhold Company, Inc. ($52.50). The crimson laser beam is a slim, unbendingly solid shaft down the dusty axis of the dark tunnel. A couple of hundred feet ahead it strikes the transparent bull's-eye of its first target, to pass another 10 feet to a terminal red spot. The specs are plain: the mile-long tunnel is expected to hold centerline well enough to hit a three-inch bull's-eye. (A prudent heading engineer will bolt in place for each setup an intermediate control target, a small metal aperture the beam can clear. After all, someone might bump the laser brackets in the night, and it is best to be sure whether it is the survey line or the tunnel wall that has shifted.)

The direction is set by a high-quality primary survey run on the surface; a big project such as a tunnel can well afford it. That line is brought underground by sighting at two opposite ends of an access-shaft diameter; the line segment below grade then fixes the laser's path. A long program of computed offsets will next steer the tunnel, an inch at a time, around the curves and up and down the grades of the planners. In normal conditions the tunnel-machine crew can hold line and grade within a tenth of a foot; taken together, the errors of the primary survey and its transfer underground are about the same.

The tunnel-boring machine is a modern wonder. The Hoosac railway tunnel was bored in hard western Massachusetts rock by the powerful likes of John Henry, steel-driving men with a hammer in their hands. It took them about five years per mile. Just 10 feet were made there in 1856 by the first tunneling machine before it was abandoned; inventor John Wilson was ahead of his time. This year is the centenary of the first successful boring machine. Colonel Beaumont's compressed-air-powered borer made 50 feet a day through the Dover chalk, but work on the famous Channel Tunnel was stopped for political reasons.

Perhaps a dozen tunnel borers were built and tried in all the years gone by; none of them saw lasting service. The rock face attacked by power drills held on big steel-armed frames—"jumbos"—and then shattered by hand-set dynamite or ammonium nitrate blasts for conveyor removal is still entirely competitive, almost standard. (A full account is given in the book.) Then in the 1950's James Robbins of Seattle brought out his tunneling machines. Their success (his first machine with rotary disk cutters pushed a 10-foot tunnel through medium-hard Toronto rocks at 10 feet per hour) stimulated worldwide development of the technology. A dozen companies in America, Europe and Japan now build such machines, both for hard-rock work and for soft ground. They are here carefully listed and compared.

The biggest rotary machines for full-face rock tunneling run to some 10 meters in diameter. The body may be 100 meters long, packed with its motors, jacks, buckets and big conveyor belt. Its thrust against the rock face may rise to 1,000 tons or even more. That thrust by the big jacks presses the whirling cutters slowly ahead a few feet at a time, shove by shove as the tunnel proceeds. Big, well-padded retractable hydraulic legs carry the thrust to the tunnel wall; a series of jacks allows systematic advance and steering. The cutting head will use 1,000 kilowatts of electric-power drive; a couple of dozen men serve the machine on each shift, usually around the clock. The monster is at work about half of the hours of the week and is down for maintenance for the rest. It eats rock at the rate of a mile per year, speaking broadly; that speed may double under good conditions but may drop to half under poor ones.

Two dozen chapters of this hefty book, replete with working drawings and performance sheets, cover an ancient and lively branch of civil engineering with detail adequate for engineers entering the work or for their administrative colleagues. The chapters are unusually appealing for the general reader because much of the complex task is illustrated by example and even narrative. There are some of the tables and charts that make up most handbooks, but this diverse activity is less codified by its nature. The consulting engineers of long experience who wrote these pieces have worldwide interests; tunnels hold the subways of Moscow, Washington and São Paulo, take the Japanese bullet trains under the sea floor from Honshu north to Hokkaido, carry the waters of the Jhelum before it reaches the Indus, and drain a flooded mine in Peru (where the hot underground waters reached a temperature of 156 degrees Fahrenheit, so that only refrigeration made the work possible).

Tunnel construction is widely viewed here: there are chapters on cut-and-cover and on boring tunnels in soft ground with the hazardous (for physiological reasons) aid of compressed-air work space. The method of choice for underwater tunnels is to fashion great tubes ashore, launch them into the water and sink them in place, trenched permanently into the bottom. Steel shells, concrete boxes and even big ribbed fiberglass tubes make good underwater tunnels.

Tunnels once opened must be put to use. They often need surface finish, drainage, air, light, power, fire protection, signals and a good deal more, even washing. All are treated in some detail, often with quantitative support. Those who daily drive automobiles in tunnels should appreciate the newer estimates for traffic density: allow a 1.8 second headway and add a small term quadratic in the speed. Ventilation is the design task that most calls for good traffic-flow data. A well-lighted and well-ventilated tunnel can carry 2,000 vehicles per hour along one standard 12-foot lane, but don't ignore the truck correction.

TRACES OF THE PAST: A FIELD GUIDE TO INDUSTRIAL ARCHAEOLOGY, by David Weitzman. Charles Scribner's Sons ($17.95). In 1883 the price of steel in the U.S. fell below that of iron. Any abandoned railway track, its rusty rails incongruous in the quiet countryside, will declare its period from that fact alone. Steel rail has flatter surfaces joined by sharper curves in its section profile; iron rail is rightly called "pear rail." Our standard railway gauge is not from time immemorial; the changeover from a regional variety of gauges was not complete until about 1900. The mighty Pennsylvania and its tributaries maintained an alternate standard, half an inch wider than Stephenson's four feet 8½ inches that had been standard until then.

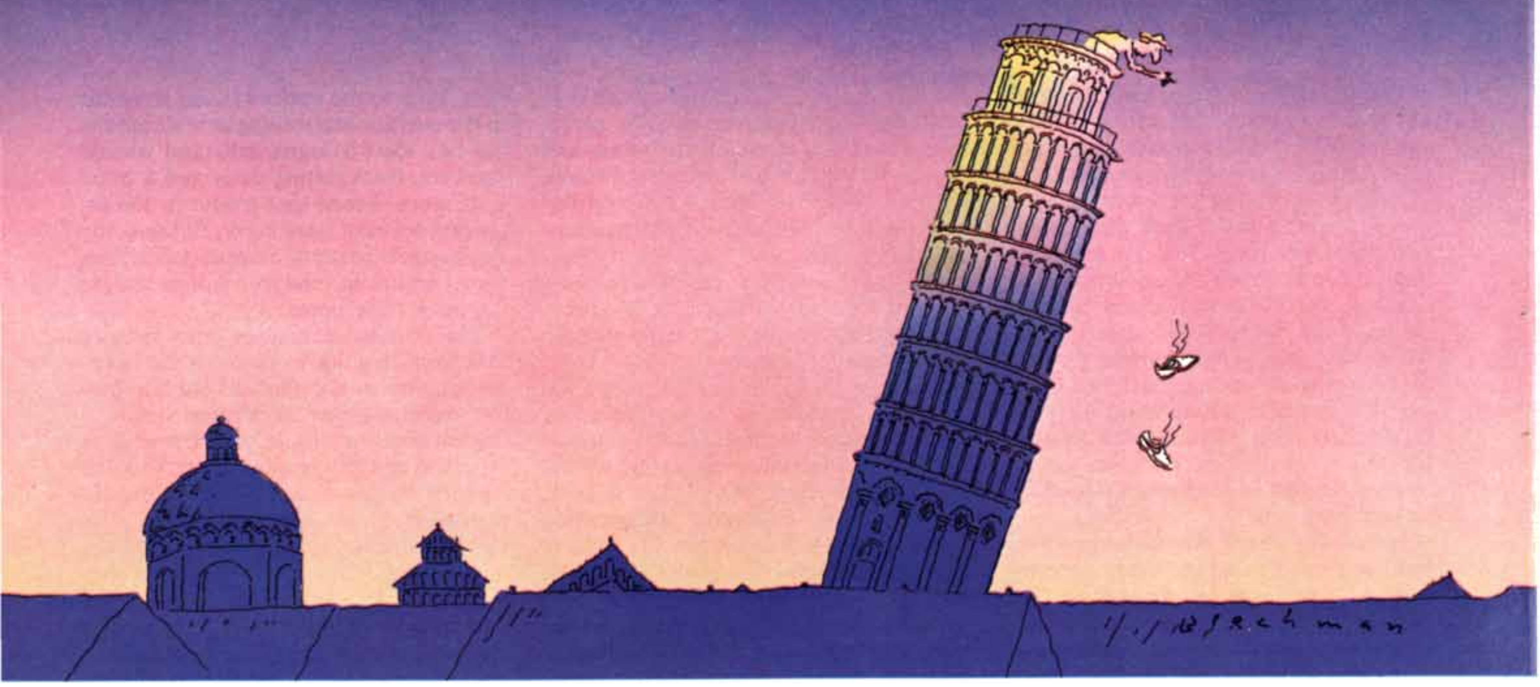
Equipment could tolerate that small difference. Real narrow-gauge "light iron" is of course still more distinct. Any old line of track is a pathway into the past, back to the entire system it recalls to the surface archaeologist walking the old ties, alert to signs, mile and whistle markers, track-dating nails and a good deal more. From that evidence the inquiring traveler may come to know the smokestack pattern, elegant paint jobs, wheel arrays and the locomotive weight the track once bore.

This modest excursion into railway lore from tangible evidence is the opening chapter of a personal field trip into the industrial past. Weitzman is now a Californian, but he is well traveled in time and space and can report how the visitor who stands on any line of the old Boston & Maine can look down rails that "run straight and true to Boston (a map of the entire system forms a perfect arrow pointed at the city)," not idly known as the Hub.

Even more engaging is the account of trusses, the articulated bridges in iron or steel or timber that nowadays often sleep in countryside, carrying a few strollers or vehicles a day, with only the memory of the rails they once bore still present. The book offers a good capsule history of rational bridge design in America, since Squire Whipple wrote it all down when he built bridges in the days of the Erie Canal. Those spans have by no means all become scrap; many are hidden in a gentler service, moved long ago to new sites as valuable structures. This particular quest is governed by the author's Rule of the Commonplace; get out of your car the next time you cross that deck bridge and walk down the embankment to see what is underneath. You may have a surprise. It may not be a Whipple, but trusses were everywhere until very recently, and the book offers a dozen-odd carefully drawn spreads to help you name and date that bridge you cross. Roof trusses too are not to be forgotten. The bridges and their builders deserve and repay attention.

That is the burden of this gentle study. The book goes on to treat iron furnaces and old oil fields, much less common artifacts but surely worth the trip if you hear of one. There is an attractive essay on bringing schoolchildren to recognize this past of ours; one particularly admires the aphorism that "learning to see means *sketching*." Of course, the printed word too is an essential resource for the aspiring archaeologist with or without a spade; good brief leads to the map literature and to the weighty books are included.

This book is not well described as a field guide; the term has been given solidity by a flood of admirable works that attempt a more or less complete summary of some broad class of experiences afield. Rather it is a sampling of the future, a pleasing evocation of a guide that does not yet exist.



NEVER JUDGE A SHOE BY ITS WEIGHT.

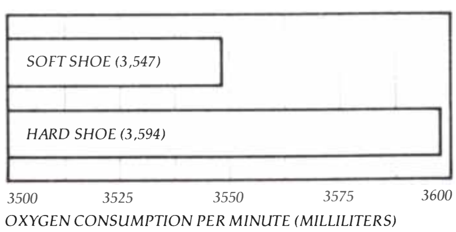
If you're looking for the fastest shoe, you've got to do more than count grams.

We know. We've done extensive research on racing flats, and there's one thing we know for sure. Light doesn't always make right.

If it did, the best shoe for race day would be no shoe at all. But things don't work that way. The fact is, you can run faster in just about any training shoe on the market than you can barefoot.

It appears to have something to do with the amount of shock created with every footstrike. Somehow that shock has to be dissipated. If your shoe won't do the work, your body will.

And work, as we all know, takes energy.

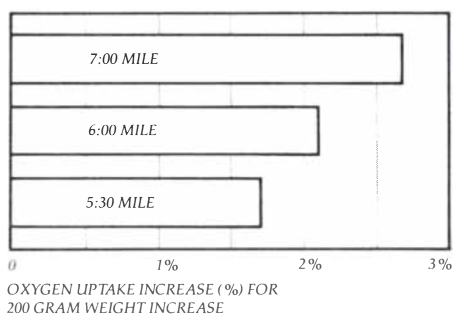


It takes more energy to run in a hard shoe than a soft one. Our lab tests were based on two models, one registering 8 peak G's in impact tests, the other 18 G's. Both were same weight and profile.

But how much? To find out, we ran a series of tests that compared two shoes of identical weight. The only significant difference was that one model was nearly twice as hard as the other. As it turned out, runners saved 1.3 percent more energy when they ran in the model with more cushioning.

Interesting. And it became even more so when we investigated our air shoes. The Mariah, for example, can deliver the same 1.3 energy savings. Even when compared to a pair of shoes weighing 100 grams less.

Then why, you ask, has everyone been so fanatical about reducing shoe weight? It goes back to some earlier research which says that for every 200 grams you knock off a pair of shoes, you gain a 2 percent energy rebate.

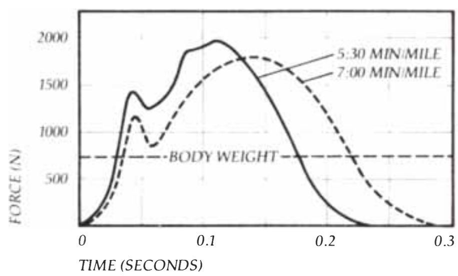


Shoe weight makes less difference the faster you run. Our study of oxygen consumption rates was based on comparison of two racing flats, one pair being 200 grams heavier than the other.

Well, we found it wasn't quite so simple. Our tests show that the saving in energy is less the faster you run. At a 7:00 mile pace, that 200 gram difference is actually worth about 2.7 percent added energy. But at a 5:30 pace, the savings—while still noteworthy—drops to 1.7 percent.

Then, you conclude, it's the slower runner who stands to benefit most from a lighter shoe. Well, yes and no. In the longer races, mainly no. Because a 3:00:00 marathoner is going to hit the ground 6,500 times more often than a 2:24:00 marathoner. And that shock accumulates and leads to fatigue.

Another thing to keep in mind. Regardless how fast you train, when the gun goes off and the adrenalin starts pumping, you'll run even faster. That's good. But that's no time to forget cushioning. As our research points out, when speed goes up, so do the impact forces.



The faster you run, the greater the impact. Vertical force record based on same runner, same shoe, at two different speeds.

It all boils down to this. Whatever your speed, whatever the conditions, don't throw away cushion to get rid of a few extra grams.

Now, we're not doing all this research just to amuse ourselves. We take all the data and turn it into shoes that are faster, safer. Not according to theory. But according to fact.

With innovations like the Air-Sole® in our Mariah. Or Phylon, in our Terra T/C. And since the right balance of weight and cushioning is going to vary from runner to runner, we offer a choice. From the super lightweight American Eagle, to the Magnum, Elite, and the Boston and Atlanta.

We know it sounds like heresy, but you're better off never trying to pick a racing flat by weight alone.

Even if it's a tough habit to drop.

NIKE

Beaverton, Oregon

Established 1845 SCIENTIFIC AMERICAN August 1982 Volume 247 Number 2

Carbon Dioxide and World Climate

That the atmospheric content of carbon dioxide is increasing now seems well established. As a result some latitudes could become warmer and wetter, but the effects may not all be bad

by Roger Revelle

The "carbon dioxide question," which has become the subject of extensive concern in recent years, is actually three questions. The first is how much carbon dioxide will be added to the atmosphere in future years, and at what rate, from the burning of fossil fuels and the clearing of forests. These human activities have already increased atmospheric carbon dioxide by about 15 percent in the past century. The second question is whether the increase in carbon dioxide will cause an important global rise in average temperature and other changes in the climate of various regions. Mathematical models of the world's climate indicate that the answer is probably yes, but an unambiguous climatic signal has not yet been detected. The third question is whether possible climatic changes and other effects of the growing concentration of carbon dioxide in the atmosphere would have major consequences for human societies. Here too the answer is probably yes, with the qualification that the consequences would be complex: some regions and some human activities would benefit, whereas others would be harmed.

Carbon dioxide alters the heat balance of the earth by acting as a one-way screen. It is transparent to radiation at visible wavelengths, where most of the energy of sunlight is concentrated, so that the sun is able to warm the oceans and the land. On the other hand, molecules of carbon dioxide in the atmosphere absorb and reradiate some of the longer-wavelength, infrared radiation that otherwise would be transmitted back into space from the surface of the earth. This is the "greenhouse effect." If there were no carbon dioxide at all in the atmosphere, heat would escape from the earth much more easily. The surface temperature required for a balance between incoming and outgoing radiation would be lower and the oceans might be a solid mass of ice.

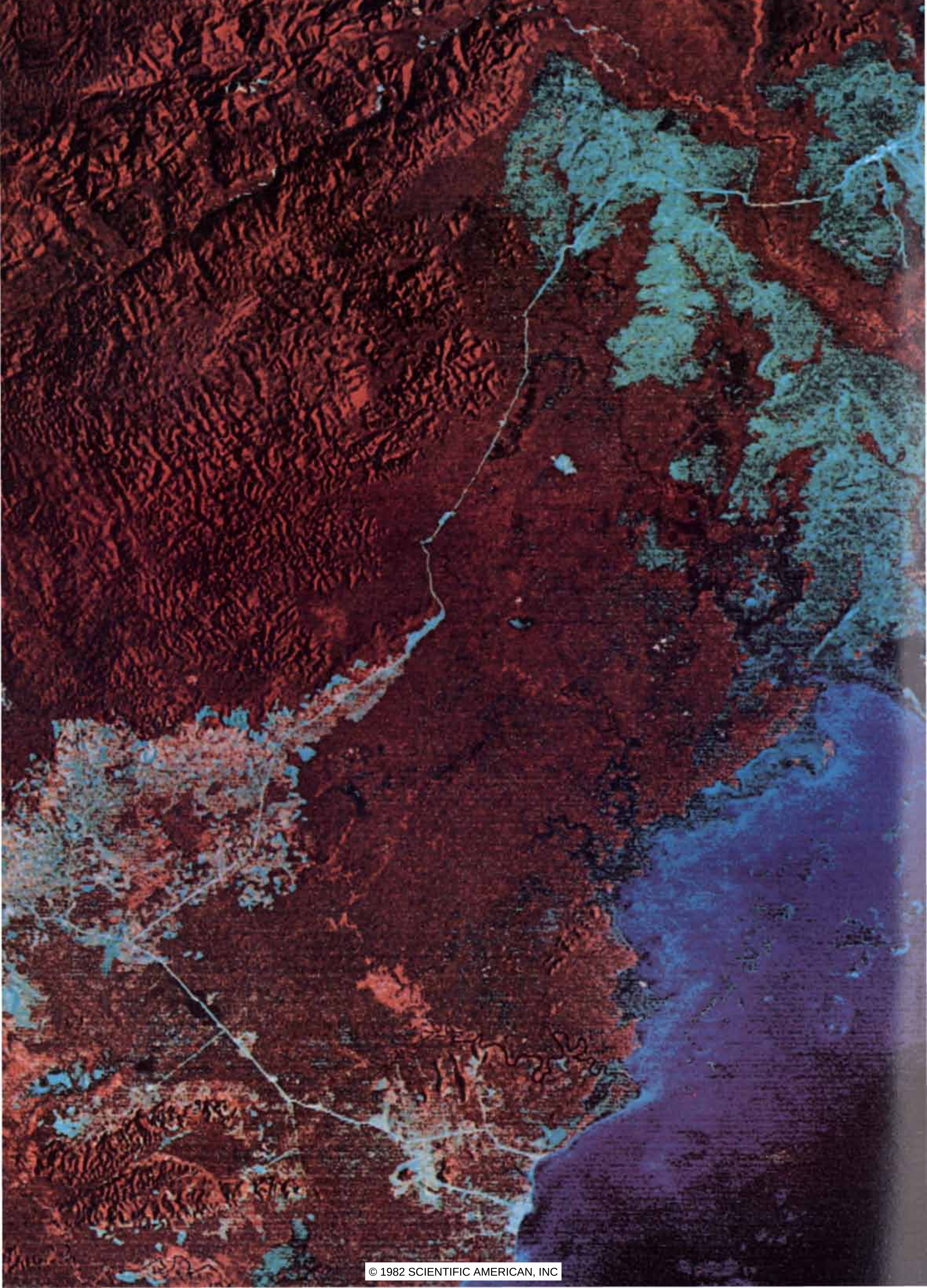
Over geologic time vast quantities of carbon dioxide have been emitted by volcanoes. Almost all of it has been chemically transformed into calcium carbonate and magnesium carbonate or into organic matter and has been buried in marine sediments. The amount of carbon in sedimentary carbonates is estimated to be about 50 million gigatons (a gigaton is a billion tons) and the amount in sedimentary organic matter is estimated to be about 20 million gigatons. The total of some 70 million gigatons is almost 2,000 times larger than all the carbon in the atmosphere, the oceans and the biosphere, namely about 42,000 gigatons. The atmospheric component is itself a small fraction of this latter amount: roughly 700 gigatons. The fact that most of the carbon in the earth's surface layers has passed through the atmosphere and has been buried in sediments accounts for the moderate strength of the greenhouse effect in the earth's atmosphere. On Venus, which has no oceans, the atmosphere consists mainly of carbon dioxide, so that the greenhouse effect is much severer and the surface temperature is 400 degrees Celsius; on Mars, where the atmosphere is very thin, the effect is weaker and the surface temperature is −50 degrees C.

Nearly 100 years ago the Swedish chemist Svante Arrhenius and the American geologist Thomas C. Chamberlin independently advanced the hypothesis that changes in the abundance of carbon dioxide in the atmosphere would affect the surface temperature of the earth. Arrhenius estimated that a doubling of the concentration would cause a global warming of about nine degrees C. In 1939 G. S. Callendar suggested that the global warming observed over the previous 60 years might have been caused by an increase in atmospheric carbon dioxide from the burning of fossil fuels. Gilbert N. Plass argued along similar lines in the early 1950's.

Neither suggestion was taken very seriously by other scientists. One reason was that at about the same time the average global temperature stopped rising; over the next two decades it even decreased slightly. A second reason was that many scientists assumed that almost all of the industrially produced carbon dioxide would be taken up by the oceans and thereby withdrawn from the atmosphere. The sea holds some 60 times as much carbon dioxide as the atmosphere, and it was thought this proportion would be maintained. Third, measurements of the carbon dioxide in samples of air taken at different times and places varied so much that it was impossible to determine whether the total amount in the atmosphere was increasing or decreasing.

Most of the carbon dioxide in the oceans is in the form of carbonate and bicarbonate ions, and only a little is dissolved "free" carbon dioxide. In 1957 Hans Seuss and I showed that a rather small change in the amount of free carbon dioxide dissolved in seawater corresponds to a relatively large change in the pressure of carbon dioxide at which the oceans and the atmosphere are in equilibrium. From this finding it followed that the rate at which the oceans can absorb the gas is much lower than had been thought.

Gases are readily exchanged between the oceans and the atmosphere only in a well-mixed surface layer some 80 meters deep on the average. If absorbed carbon dioxide is confined to this layer, a given change in the amount of carbon dioxide in the atmosphere would change the content of the surface water by



only about a ninth as much. We called this 9 : 1 ratio the "buffer factor" and showed that it increases with increasing atmospheric carbon dioxide. Downward diffusion, advection or convection between the well-mixed layer and the deeper water would increase the amount of carbon dioxide taken up by the ocean and thereby reduce the amount remaining in the atmosphere, but we believed these processes are relatively slow. Hence it seemed likely that much of the carbon dioxide released by the combustion of coal, oil and natural gas would stay in the atmosphere for a considerable time, possibly with profound effects on the world's climate.

Man was inadvertently conducting a great geophysical experiment. If it were adequately documented, it could lead to a new level of understanding of the ocean and the atmosphere. It might also result, however, in climatic changes that could be seriously disruptive.

Projects for the measurement of atmospheric carbon dioxide were incorporated into the program of the International Geophysical Year in 1957–58. Part of the work was assigned to the Scripps Institution of Oceanography, which appointed Charles David Keeling to make the measurements. He set up highly accurate gas analyzers for the continuous measurement of carbon dioxide at two sites: near the summit of Mauna Loa in Hawaii and at the South Pole station of the U.S. Antarctic Program. The sites were chosen because they are free from local contamination and presumably would offer well-mixed air for sampling. The data from these sites were supplemented by measurements of air samples collected over the oceans by airplanes.

The accompanying graphs [*page 39*] show the results obtained by Keeling and his collaborators in the U.S. National Oceanic and Atmospheric Administration in two decades of measurements. At both stations an annual cycle of carbon dioxide concentration is evident; the cycle reflects the changing seasonal activities of green plants. During the growing season there is a net removal of carbon from the air by photosynthesis; at other times carbon dioxide is returned to the air by the oxidation of plant tissues. With the seasonal variation removed the records show a roughly exponential increase at both stations. When the measurements began, the concentration was rising at a rate of about .7 part per million per year; by the late 1970's the rate had reached 1.4 parts per million per year.

Keeling and his colleague Robert Bacastow concluded that the average concentration at Mauna Loa increased from 315.8 parts per million in early 1959 to 334.6 in January, 1978—a rise of 6 percent. The South Pole station lags behind Mauna Loa because of the time required for mixing across the Equator, and so the observed increase was slightly smaller at the South Pole. The increase of about 19 parts per million at Mauna Loa corresponds to the addition of 40 gigatons of carbon to the atmosphere in those 19 years.

Ralph M. Rotty of the Institute for Energy Analysis in Oak Ridge, Tenn., has compiled and evaluated United Nations data on the production of coal, oil and natural gas over this period. He has concluded that the total carbon dioxide produced (including a small contribution from cement manufacture) corresponds to about 70 gigatons of carbon, rising from 2.4 gigatons per year in 1959 to five gigatons per year in 1977. This quantity is 1.75 times the measured amount of carbon added to the atmosphere.

Rotty has also calculated that the fossil-fuel carbon dioxide produced between 1860 and 1958 corresponds to 76.5 gigatons of carbon. Applying the same ratio of production to atmospheric increase, the carbon added to the atmosphere between 1860 and 1977 was 84 gigatons, or 40 parts per million of carbon dioxide. The estimate is uncertain, however, because there must have been a major input of carbon dioxide during the century after 1860 from the clearing of hundreds of millions of hectares of forest to increase agricultural land areas. Until recent decades such an expansion of farmland was necessary to feed the world's growing population; crop yields per hectare did not increase much until after World War II.

The convincing evidence that the concentration of atmospheric carbon dioxide is rising exponentially has resulted in an even more rapid growth of scientific measurement and theorizing about the carbon dioxide problem. There has been a plethora of conferences and symposiums, and measurements of atmospheric carbon dioxide have been made in many places, ranging in latitude from Barrow in Alaska to Halley in Antarctica, in environmental conditions from western Europe to weather station PAPA in the center of the North Pacific and Amsterdam Island in the south Indian Ocean and in time from the present to the coldest part of the last glacial period about 18,000 years ago. The measurements make possible the construction of a cross section showing the variation with latitude of the mean annual concentration of carbon dioxide in the atmosphere. The highest values turn out to be north of 40 degrees latitude in the Northern Hemisphere, where the combustion of fossil fuels and the seasonal changes in plant productivity are greatest. The lowest values are found in the Southern Hemisphere, where fossil-fuel combustion and land-plant activities are relatively small.

For the earth to remain at a constant temperature the infrared emissions from the planet must balance the absorbed solar radiation. Under present conditions the balance is reached at a radiating temperature in the earth's atmosphere of −18 degrees C. and at an altitude of from five to six kilometers. Below this level the temperature increases by about six degrees per kilometer approaching the earth's surface, where the average temperature is 15 degrees. If the opacity of the atmosphere to infrared radiation increases, the radiating level and the temperature of the surface and the lower atmosphere must rise until the emissions from the planet and the absorbed solar radiation are again equal. An increase in the amount of carbon dioxide in the atmosphere tends to close the atmospheric infrared "window" because there are more molecules of carbon dioxide to absorb the radiation.

The mean increase in surface temperature can be calculated with the aid of a mathematical model that predicts the temperature as a function of altitude above the surface. James E. Hansen and his colleagues at the National Aeronautics and Space Administration have employed such a model to examine the main processes that could influence the temperature as carbon dioxide is added to the atmosphere. They assumed that both the relative humidity and the temperature at the cloud tops would remain constant and that the temperature would decrease with altitude at a rate of 6.5 degrees C. per kilometer. Under these conditions they found that doubling the atmospheric carbon dioxide would raise the global mean surface temperature by 2.8 degrees, with an estimated uncertainty factor of two.

Water vapor in the atmosphere is also an efficient absorber of radiation at infrared wavelengths. Part of the calculated rise in temperature results from the

CLEARING OF FORESTS is evident in the Landsat image of the Toledo District of Belize (formerly British Honduras) on the opposite page. In the infrared-color system of Landsat imagery the dark red areas are forests and the blue-green and blue areas (other than the Gulf of Honduras at the bottom right) represent recently cleared forest. The bright pink areas indicate regenerating vegetation, either converted cropland or minor forest growth. The denuded area is approximately 218 square kilometers, or nearly 1 percent of the area of the country. Similar clearing, in most cases to extend agricultural regions, has been done in Bangladesh, Brazil, the Philippines, Thailand, Venezuela and other countries. The removal of forests adds large amounts of carbon dioxide to the atmosphere because the trees incorporate from 10 to 20 times as much carbon per unit area as land in crops or pasture does. Countervailing factors are the regrowth of forests on abandoned farmland and deliberate reforestation in some areas.

assumption of constant relative humidity; because warm air can hold more moisture than cool air, keeping the relative humidity constant requires the amount of water vapor in the atmosphere to increase with the temperature. Therefore more infrared radiation would be absorbed and reradiated back to the earth's surface, creating positive feedback.

If the average temperature of the cloud tops does not change, the amount of infrared radiation being emitted into space from them remains constant. The clouds move to a higher level, however, because of the warming of the lower atmosphere. If the rate of temperature change between the surface and the cloud tops remains constant, the result is still more warming, giving rise to another positive feedback effect.

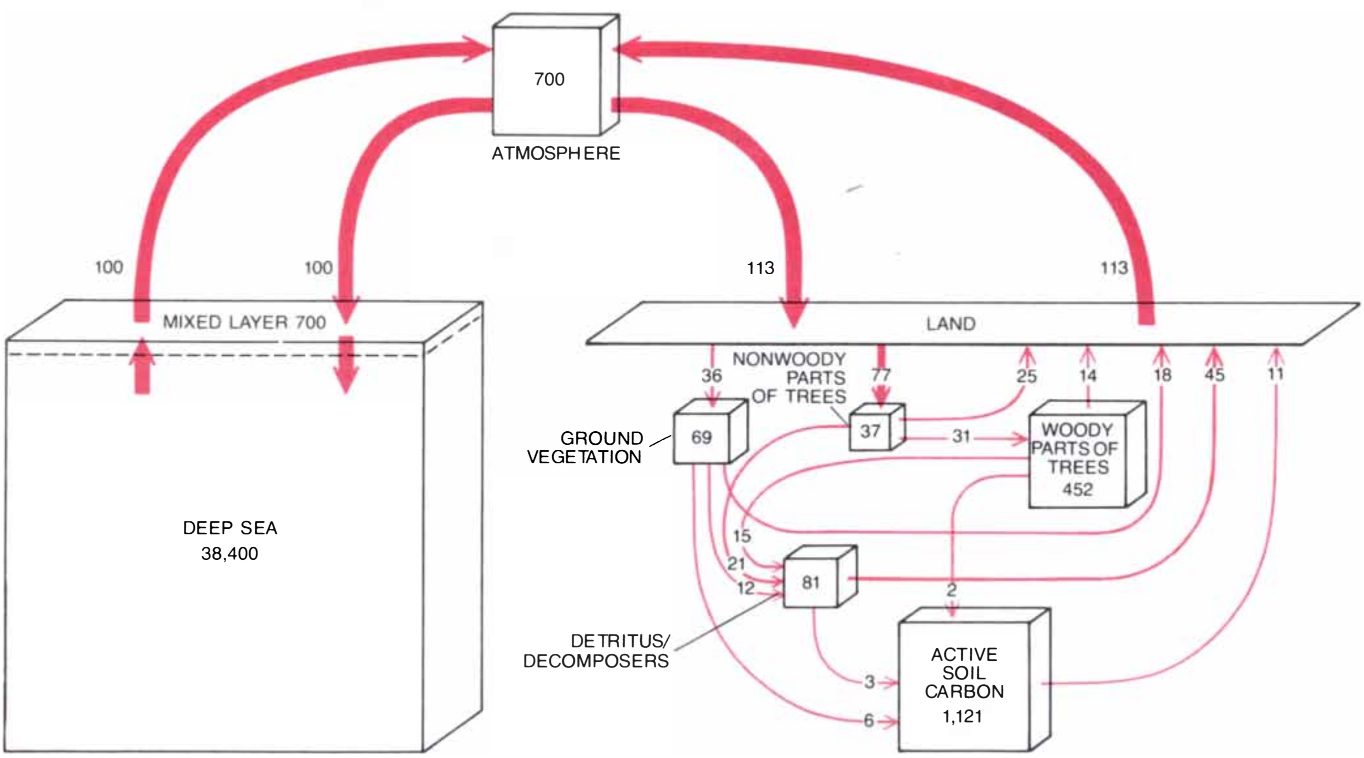
The warming at the surface could be expected to melt snow and ice, reducing the earth's albedo, or reflectivity. More solar radiation would then be absorbed, leading to a further increase in temperature. The result would be still another positive feedback effect about equal in magnitude to the one implied by the assumption of fixed temperature at cloud tops. Future warming near the earth's surface from increased carbon dioxide would also be significantly enhanced, perhaps by as much as 50 percent, if the concentration of other infrared-absorbing gases, including nitrous oxide, methane and chlorofluorocarbons, increases as it is expected to in the next 50 to 100 years.

Hansen and his colleagues also examined records of the mean annual temperature worldwide between 1880 and 1980. They found that the temperature rose about .5 degree C. between 1885 and 1940, with a slight cooling thereafter. In records based entirely on data collected in the Northern Hemisphere north of the Tropics the recent cooling is a prominent trend, but Hansen's group found that the average global temperature in 1980 was almost as high as it was in 1940. During the period from 1880 to 1980 the mean global temperature rose about .4 degree, a change roughly consistent with their calculation that the increase in the carbon dioxide in the atmosphere was 43 parts per million over the same period.

The correlation between carbon dioxide level and temperature is not exact, however, indicating that other factors must have affected the global temperature. When Hansen's group took into account the absorption of heat by the oceans and variations in the quantity of volcanic aerosols, the fit improved considerably. An even better fit is obtained when a hypothetical variability of .2 percent in the sun's luminosity is also included.

The general agreement between modeled and observed trends in temperature strongly suggests that carbon dioxide and volcanic aerosols are responsible for much of the global variation in temperature in the past 100 years. It must be said, however, that so far the warming trend has not risen above the "noise level" of yearly fluctuations in temperature, which amount to plus or minus .2 degree C. Confidence in the carbon dioxide hypothesis will be much firmer if a warming trend exceeding the noise level becomes evident. Almost any reasonable estimate of how much fossil fuel will be burned in the coming years suggests that if carbon dioxide is indeed altering climate, an unmistakable warming trend should appear in the 1990's.

Syukuro Manabe, Richard T. Wetherald and Ronald Stouffer of the Geophysical Fluid Dynamics Laboratory at Princeton University have devised a mathematical model to explore the geographic distribution of probable climatic changes resulting from a doubling or a quadrupling of atmospheric carbon dioxide. The model shows a markedly different climatic response at different latitudes. With a doubling of carbon dioxide the rise in the average temperature at the surface is less than two degrees C. in the Tropics, but it is three degrees C. at a latitude of 35 degrees North (roughly the latitude of Los Angeles and Tokyo). At 50 degrees North (near Paris and Vancouver) the temper-



CARBON CYCLE is shown in terms of the amount of carbon stored in various reservoirs and the annual flows between reservoirs. The flows are indicated by the arrows. The numbers are in gigatons, or billions of tons. In the cubes representing stored carbon the numbers show the present carbon content of the reservoir. The numbers associated with the arrows give the equilibrium flow of carbon in gigatons per year, which would take place without any human intervention. Human activities such as the burning of fossil fuels and the clearing of forests are increasing the amount of carbon (as carbon dioxide) in the atmosphere. The increase could lead to a global warming because molecules of carbon dioxide in the atmosphere absorb and reradiate some of the infrared radiation from the earth's surface. The estimates of stored carbon and carbon flows from the land biosphere were made by Jerry Olson of the Oak Ridge National Laboratory.

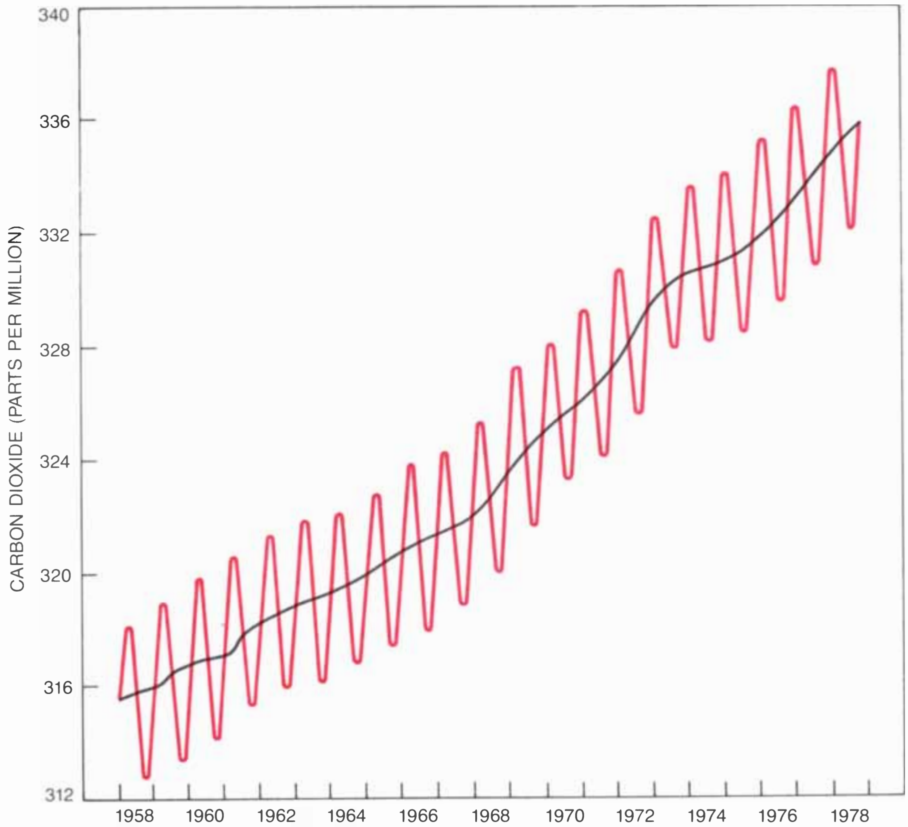
ature is higher by four degrees C., and at 76 degrees North (well above the Arctic Circle) the increase is seven degrees C. The projected increase at high northern latitudes is much larger between October and May than it is during the summer, thereby reducing the amplitude of seasonal temperature variations over northern lands.

The model also shows a large increase in the rates of precipitation and runoff at high northern latitudes. Between 35 and 45 degrees North, however, there is a significant decrease in the difference between precipitation and evaporation, which could result in a marked decline in runoff. For a doubling of atmospheric carbon dioxide the climatic changes predicted by the model are larger than any since the end of the last ice age about 12,000 years ago. Global temperatures might be warmer than civilized man has ever experienced. The high temperatures would persist for hundreds of years until slow absorption by the oceans removed excess carbon dioxide from the atmosphere.

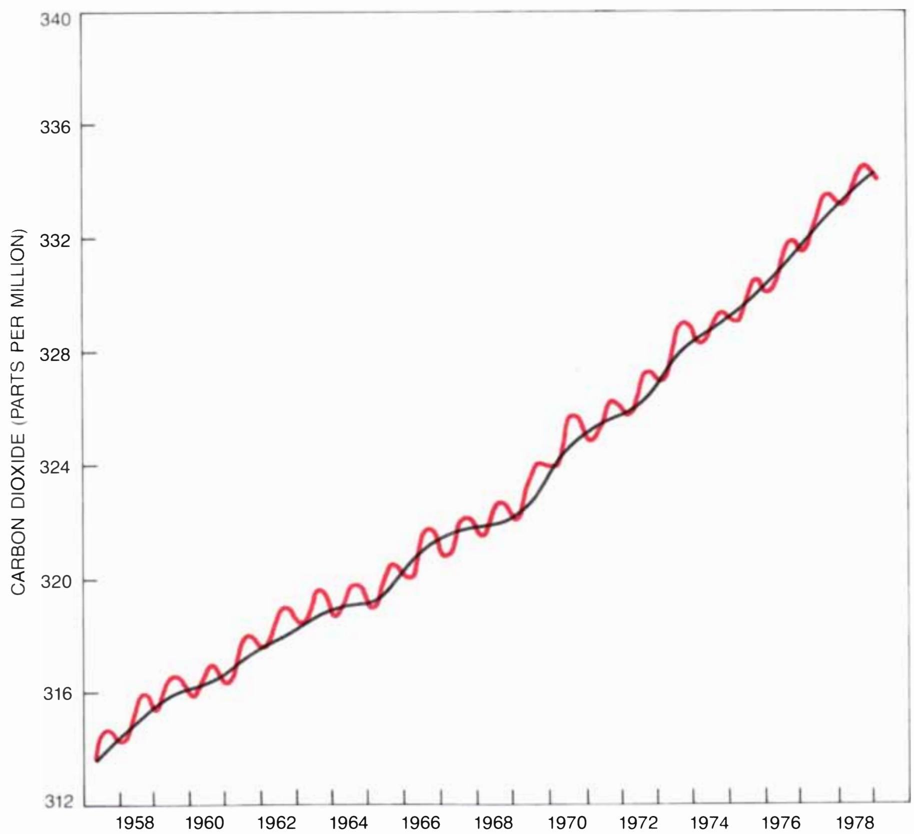
Models of the type employed by Manabe and his colleagues entail many uncertainties. For example, the model has only a partially realistic coupling between the atmosphere and the ocean, and so it cannot represent the transport of heat from low to high latitudes by ocean currents or the exchange of heat between the mixed surface layer of the ocean and deeper water. Another uncertainty stems from the difficulty of predicting the response of clouds to changes in climatic conditions. Similarly, it is difficult to predict the effect of variations in the geographic extent of sea ice on the exchange of heat and water vapor between the sea and the air. The coarse spatial resolution of the model obscures regional variations in climate.

All these uncertainties suggest that the mathematical modeling of changes in climate should be evaluated and controlled by studies of past climates, and particularly of past warm periods. A study of this kind has recently been done by Hermann Flohn of the University of Bonn. From his findings he has estimated the changes in average temperature and precipitation that are likely in several latitude belts if the concentration of atmospheric carbon dioxide rises into the range from 560 to 680 parts per million (about twice the level of a century ago) and other infrared-absorbing gases also become more abundant [*see bottom illustration on page 41*].

Flohn's analysis of the historical record suggests there could be appreciable decreases in precipitation and increases in temperature (and hence in evaporation) in bands of latitude centered on 40 degrees North and 10 degrees South. Precipitation would be greater, on the other hand, between 10 and 20 degrees North and in the regions north of 50 degrees North and south of 30 degrees



RISING CONCENTRATION of carbon dioxide in the atmosphere at Mauna Loa in Hawaii is evident in continuous measurements made by Charles David Keeling of the Scripps Institution of Oceanography and by the National Oceanic and Atmospheric Administration. The seasonal variation (*color*) arises because carbon dioxide is removed from the air by the photosynthetic activity of plants during the growing season and is returned later by the oxidation of plant tissues. Smoother curve (*black*) shows change with seasonal effect artificially suppressed.



RISE IN CARBON DIOXIDE at the South Pole is shown with seasonal variations (*color*) and without them (*black*). The continuous measurements were made under Keeling's direction.

South. These changes could have profound effects on the distribution of the world's water resources. For example, in the Colorado River system of the U.S. the major drainage basin is at about 40 degrees North. Not only would rainfall diminish in this area but also the higher temperature would augment evaporation. At present about 85 percent of the precipitation evaporates and only 15 percent is carried by the river. With a rise of several degrees C. in air temperature and a 10 to 15 percent decline in precipitation the average flow of the Colorado could diminish by 50 percent or more. Even today's flow, backed up by large volumes of water stored in reservoirs, is barely enough to meet the demands of irrigated agriculture.

Major changes in surface and underground water supplies would be likely in other parts of the world. In northern Africa the average flow of the Niger, Chari, Senegal, Volta and Blue Nile rivers could increase substantially because their basins would receive from 10 to 20 percent more precipitation while temperatures at the surface would rise only slightly. In many other rivers the flow could greatly decrease: the Hwang Ho in China, the Amu Darya and Syr Darya in one of the prime agricultural areas of the U.S.S.R., the Tigris-Euphrates system in Turkey, Syria and Iraq, the Zambezi in Zimbabwe and Zambia and the São Francisco in Brazil. Somewhat smaller runoff and underground storage could be expected in the Congo River in Africa, the Rhone and the Po in western Europe, the Danube in eastern Europe, the Yangtze in China and the Rio Grande in the U.S. Many of these rivers form the basis of extensive and highly productive irrigated-agriculture systems, and the projected slackening of their flow could have grave consequences. At the same time large increases in the flow of the Mekong and the Brahmaputra rivers could lead to frequent destructive floods over wide areas of Thailand, Laos, Cambodia, Vietnam, India and Bangladesh.

The large-scale effects of increased atmospheric carbon dioxide on rain-fed agriculture would be complex and are more difficult to estimate than the effects on irrigated agriculture. Carbon dioxide is an essential plant nutrient; it is one of the raw materials from which organic matter is formed in photosynthesis. As Sylvan H. Wittwer of Michigan State University and Norman J. Rosenberg of the University of Nebraska at Lincoln have emphasized, experiments both in the greenhouse and in the field show that a higher concentration of carbon dioxide promotes photosynthesis and gives rise to faster growth. Other things being equal, a higher atmospheric content of carbon dioxide should lead to increased production of such crops as rice, wheat, alfalfa and soybeans.

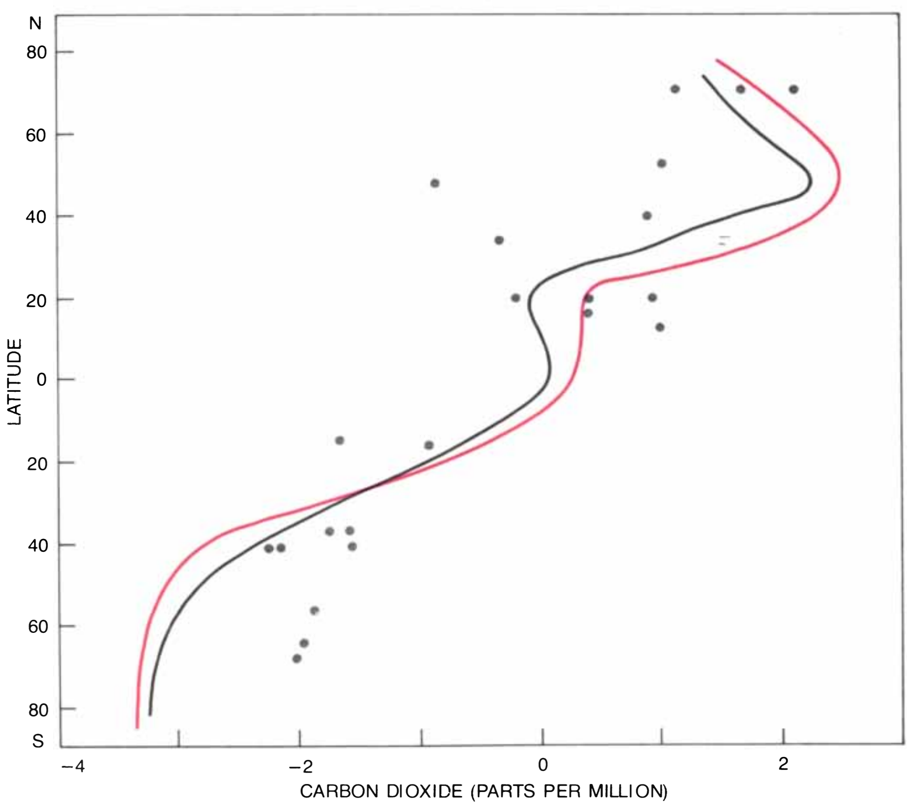
Of perhaps equal importance is the fact that many plants tend to partly close their stomata, or leaf pores, in an environment high in carbon dioxide, with a resulting reduction in the transpiration of water. Hence with marginal rainfall maize, sugarcane and sorghum, as well as the crops already named, are likely to be less affected by water stress as the atmospheric content of carbon dioxide increases. At high latitudes the higher temperatures induced by carbon dioxide may lengthen the growing season, making it possible to expand rain-fed agricultural areas.

A possible consequence of climatic warming that has been widely discussed is the disintegration of the West Antarctic ice sheet, which many glaciologists think is unstable because much of it is below sea level. The volume of ice above sea level is about two million cubic kilometers. If it were all carried into the ocean, the sea would rise by five to six meters, inundating many coastal cities and much farmland in the Netherlands, Bangladesh, the coastal lowlands of the southern U.S. and populated river deltas throughout the world. Half of the state of Florida would be covered by seawater.

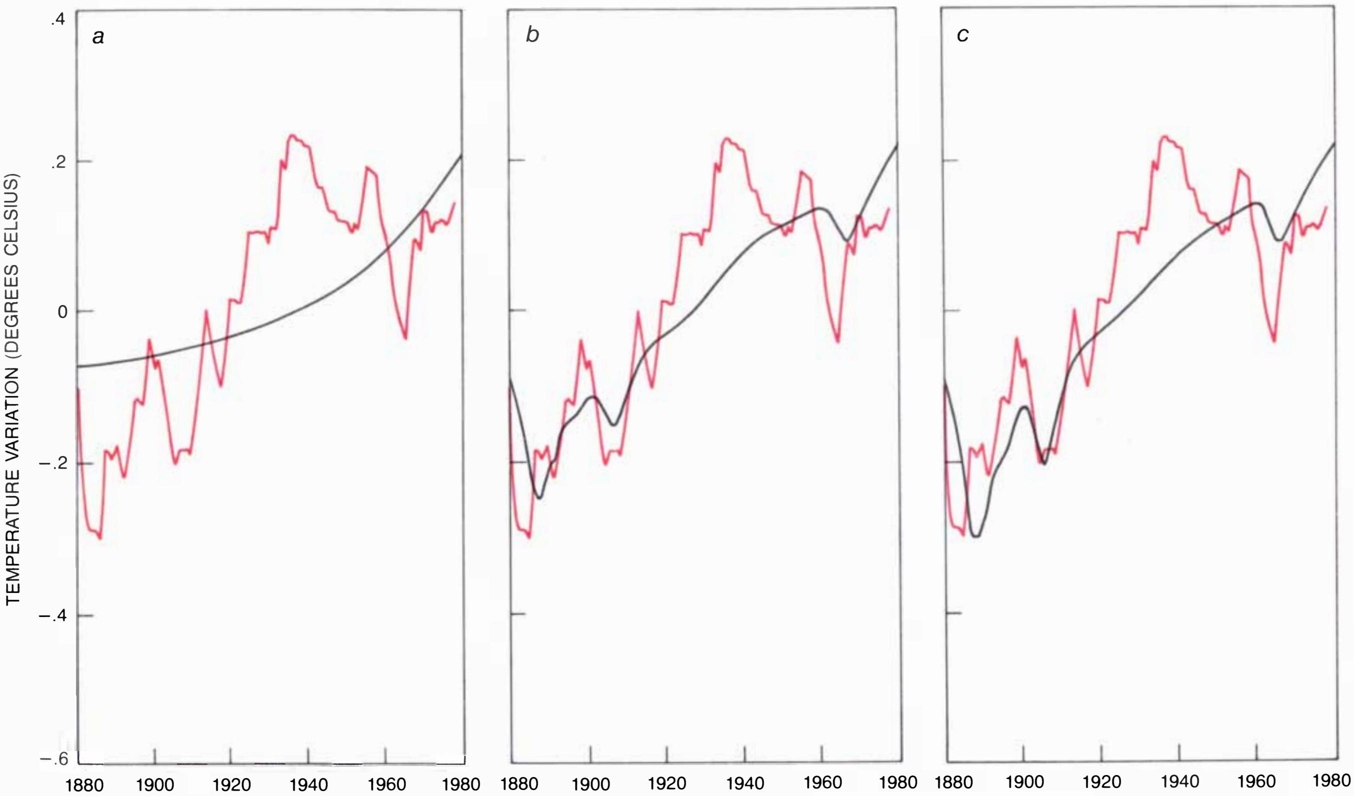
Although such changes in the map of the world may seem to be the most dramatic potential consequence of climatic warming, social and economic adaptation might not be difficult, at least for the cities, if the change were sufficiently slow. Except for historical monuments, cities are constantly being rebuilt even without the impetus of changing coastlines. The "half-life" of city buildings—the period after which half of the buildings erected in a given year have been destroyed or replaced—seems to be between 50 and 100 years. Thus the disruptiveness of climatic change will depend strongly on the rate of change.

It has been argued above that important consequences can be expected when the atmospheric content of carbon dioxide reaches a level twice what it was in the middle of the 19th century. When might that happen? The range of probable answers depends on two other matters: the future rate of combustion of fossil fuels and the future proportion of carbon dioxide emissions from all human activities that remains in the air.

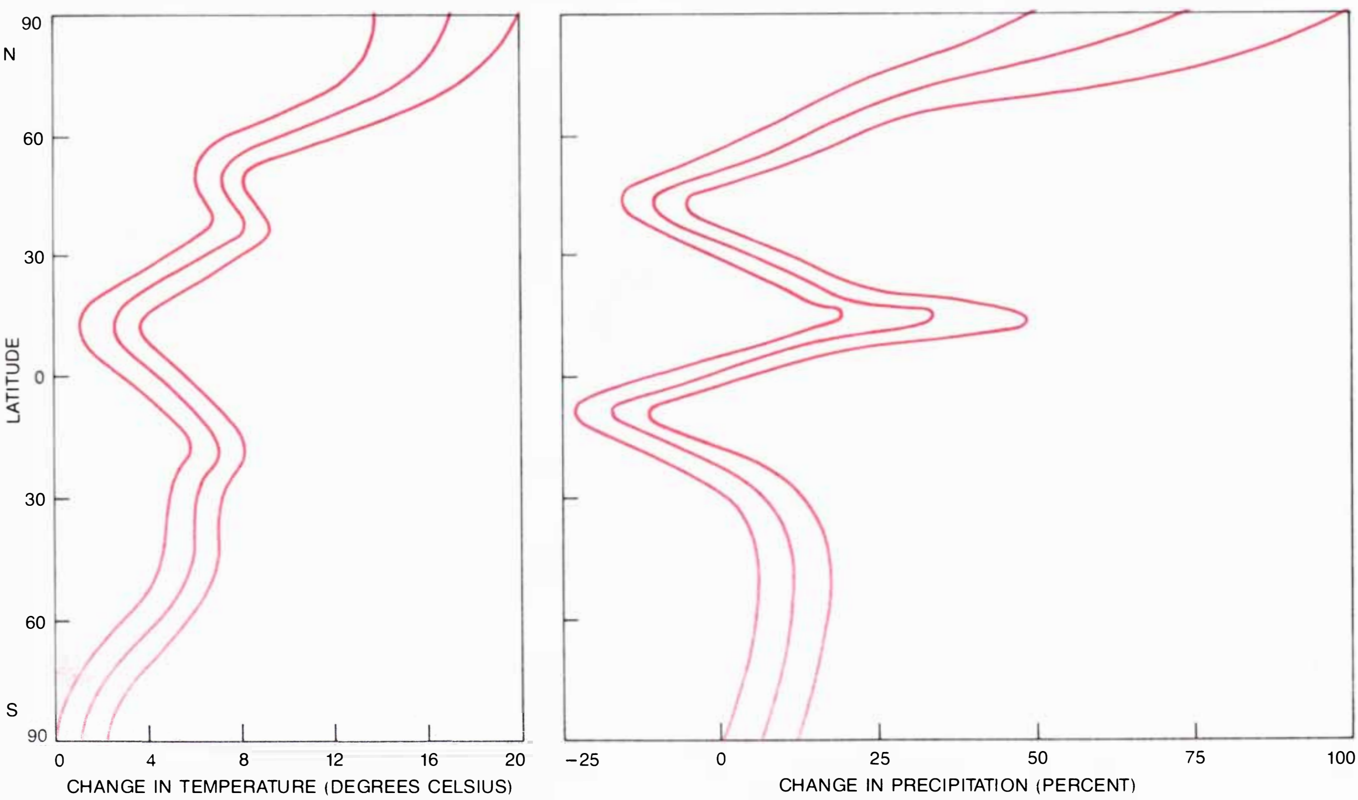
This proportion is called the airborne fraction. A future doubling of carbon dioxide over the probable 19th-century level would require the addition of about 500 gigatons of carbon to the atmosphere beyond the amount there in 1980. If the airborne fraction remains the same as it has been in the past two decades, the required input of carbon to the air would be between 1,000 and 1,250 gigatons, some 200 times the annual emissions during the past few



CALCULATED LATITUDINAL VARIATION in the atmospheric content of carbon dioxide is given for altitudes where the barometric pressure is 700 millibars (*black*) and 900 millibars (*color*). The gray dots show measurements over various periods from 1958 to 1979 at latitudes from Barrow in Alaska to the South Pole. The data are in parts per million of carbon dioxide above and below the annual mean concentration at Mauna Loa. The highest readings are north of 40 degrees north latitude, where the combustion of fossil fuels is also the highest.



WARMING TREND that may have arisen at least partly from the increase in the amount of carbon dioxide in the atmosphere is evident in recorded temperature variations from 1880 to 1980 (*color*). James E. Hansen and his colleagues in the National Aeronautics and Space Administration compared the observations with a climatic model (*black*) that assumes a rise of 2.8 degrees Celsius in the global mean temperature if the atmospheric content of carbon dioxide doubles. Versions of the model took into account the effects of carbon dioxide only (*a*), the effects of both carbon dioxide and volcanic aerosols (*b*) and the effects of these factors and a hypothetical variability of .2 percent in the luminosity of the sun (*c*). Also considered was the ability of the top several hundred meters of the ocean to absorb heat.

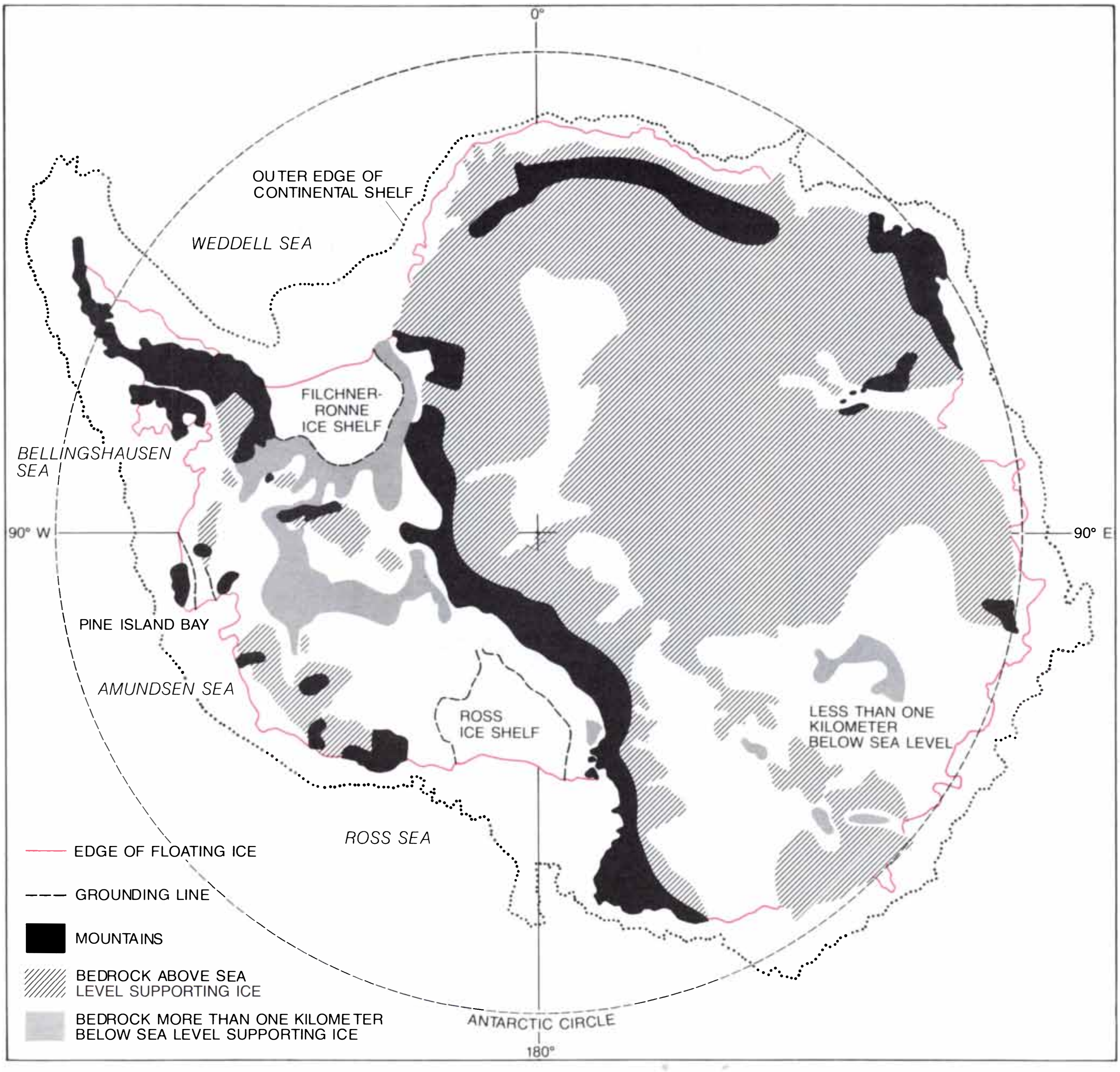


CLIMATIC EFFECTS of a rise in atmospheric carbon dioxide into the range from 560 to 680 parts per million (about double the level of 100 years ago) were projected by Hermann Flohn of the University of Bonn. The graphs show the projected variation by latitude in temperature (*left*) and precipitation (*right*). Flohn assumed that a global warming would melt the ice in the Arctic Ocean during the summer but would leave a significant amount of ice on the continent of Antarctica. The three curves in each group reflect the range of possibilities.

years. Because of continuing population growth, however, the world's rate of energy use will need to be greater in the future than it is now in order to maintain a constant supply of energy to each person. Populations are growing slowly in the developed countries of Europe and North America, the U.S.S.R. and Japan, and the consumption of energy per person there could be significantly reduced by several feasible conservation measures. But the 70 percent of human beings in the developing nations, where populations are still growing rapidly, will need much more energy per person than they are now using if they are to lift themselves out of poverty.

For even moderate economic growth in the developing countries the world's energy use should probably increase by 2 to 3 percent per year. This would amount to a continuation of present trends. Since 1973 the combined consumption of coal, oil and natural gas has been growing at 2.25 percent per year. Before the oil shock of 1973 the rate of growth was more than 4 percent per year. If most energy continues to come from fossil fuels and net contributions from the biosphere are small, carbon emissions into the atmosphere in 2050 could be from 20 to 40 gigatons per year. The atmospheric carbon dioxide would be twice the 19th-century value by about the middle of the next century.

Because other energy sources may be partly (or someday completely) substituted for fossil fuels, one can make only plausible guesses about the quantity of fossil fuels that will be consumed in times to come. It is certain, however, that the present pattern of fossil-fuel consumption, in which nearly two-thirds of the carbon dioxide production comes from petroleum and natural gas, cannot continue for long because the recoverable supplies of those fuels are limited. With the present fuel proportions and rates of growth in fuel consumption most of the petroleum and natural gas will be gone by 2050, where-



WESTERN ANTARCTICA, meaning the land and ice west of the zero meridian to the 180-degree meridian, is a focus of concern because of the possibility that a global warming resulting from a rise in the amount of carbon dioxide in the atmosphere might melt the ice sheet there. Part of the continental ice is supported by bedrock that is above sea level and part of it by rock below sea level. Most of the ice in the Ross and Filchner-Ronne ice shelves is floating, but it is "pinned" here and there by submarine features that rise high enough to touch the base of the ice. The melting of any ice above sea level could raise the level of the world's oceans by as much as six meters.

as from 90 to 95 percent of the recoverable coal will remain. The carbon dioxide problem is thus fundamentally related to the future consumption of coal.

Almost 90 percent of the recoverable coal is in the U.S.S.R., China and the U.S. Therefore only these three countries can have a decisive role in any large future increase in atmospheric carbon dioxide. The U.S. has more advanced mining and transportation systems than the other two countries have, and so for the next few decades the primary responsibility for what will be done about coal will probably rest with the U.S.

As for the fraction of carbon dioxide that remains in the atmosphere, it should eventually be possible to make useful estimates through analytic modeling of the carbon cycle and of oceanic circulation. It is generally agreed that the clearing of forests has contributed fairly large quantities of carbon dioxide to the air. The reason is that forests incorporate from 10 to 20 times more carbon per unit area than cropland or land in pasture does. On the other hand, the regrowth of forests on abandoned agricultural land and the deliberate increase of forests by tree planting and improved silviculture in China, Europe and North America must compensate to some degree for the loss of wooded land elsewhere.

Direct worldwide monitoring of the living biomass and the organic content of soils by means of satellite imagery combined with ground surveys should in principle yield data on changes in these quantities caused by human activities. Studies of this kind are being done by George M. Woodwell and his associates at the Marine Biological Laboratory in Woods Hole, but evaluating the data has turned out to be extremely difficult. The amount of carbon liberated annually by forest clearing is probably between one gigaton and two gigatons.

It can be concluded from the limited evidence available that in the past two decades between 40 and 50 percent of the carbon dioxide produced by human activities has remained in the air. It is not certain, however, that the airborne fraction will remain the same in the future. Several factors have an influence on the size of the airborne fraction. The buffer factor should become larger as atmospheric carbon dioxide increases, and this will reduce the rate at which carbon is transferred from the air to the ocean. On the other hand, if the rate of carbon emissions diminishes, as it has in the past few years, there will be more time for the oceans to absorb a given quantity of carbon dioxide and consequently more of the gas can be dispersed in the deeper ocean layers.

It is likely that the rate at which forests are cleared for agriculture will decrease in coming decades as the limits of cultivable land are approached and as improved technology results in larger crop yields from land already under cultivation. At the same time the increase in atmospheric carbon dioxide may bring about an increase in photosynthesis and hence in the quantity of biomass, at least in some regions. The combination of factors could result in a significant net flow of carbon from the atmosphere to the biosphere. In 1977 Walter H. Munk and I constructed a rather primitive model that attempts to take account of these various factors. Our calculations suggest that the fraction of carbon dioxide from human activities remaining in the air will be almost constant, with a value ranging from 40 to 50 percent, until the peak rate of fossil-fuel consumption is passed. This peak should be reached near the beginning of the 22nd century.

In sum, the carbon dioxide question is obscured by many unknowns and uncertainties. Indeed, about the only facts available are the actual measurements of atmospheric carbon dioxide, particularly the two-decade series at Mauna Loa and the South Pole, and some fairly reliable data from the UN on the annual consumption of fossil fuels in industrial countries. These facts are sufficient to suggest, however, that steps should be taken to obtain more evidence and to consider the consequences of a continuing increase in atmospheric carbon dioxide.

Experience has shown that the planning and construction of water-resource developments in major river basins can take several decades. It is not too soon to begin to think of ways in which the planned utilization of water could ameliorate the potential effects of climatic changes or even take advantage of them. Several measures come to mind: (1) changes in legislation that would allow water to be transferred from one river basin to another, for example from the Columbia to the Colorado; (2) improved efficiency in the use of water for irrigation; (3) conservation of waste water and of municipal water supplies; (4) limitations on the size of irrigated areas; (5) increases in crop yields per unit volume of applied water, and (6) enhancement of the recharge of aquifers.

Modern research-intensive agriculture is highly adaptable to different climatic conditions. Rosenberg has cited the example of hard red winter wheat, which is grown from North Dakota to northern Texas in areas with a wide range of temperature and precipitation. A few decades ago it was grown only in eastern Nebraska. Comparable efforts could be made to ascertain which crops might grow best in places where climatic change is possible as a result of increased atmospheric carbon dioxide.

Evidence from fossil corals reveals that the sea level was approximately six meters higher during the last interglacial period (some 120,000 years ago) than it is now. Coral reefs established during the period of high water were stranded by the subsequent retreat, leaving fossil terraces. The transient rise in sea level may have been caused by the breakup of the West Antarctic ice sheet. Drilling through the present sheet could give useful information. One might expect to find a discontinuity in the annual ice layers at a depth corresponding to a time about 120,000 years ago. The nature of the discontinuity could give an indication of the sequence of events.

If the ice sheet did disintegrate in the last interglacial, an attempt to determine the rate of its destruction would seem to be a research problem of great importance. If the disappearance was rapid, a surface layer of low-salinity water might have persisted long enough to leave a measurable trace. Fresh water and seawater differ in their content of certain isotopes of oxygen; the difference might be detected in the fossilized coral terraces.

The dynamics of ice flow also present an important research problem. At present the West Antarctic ice shelves are calving off huge icebergs at a rate that corresponds to a seaward motion of about a kilometer per year. Along 1,200 kilometers of ice front some 500 cubic kilometers of ice are being deposited in the sea each year. Other things being equal the result should be a rise of 1.5 millimeters per year in sea level, which is in fact observed in worldwide annual averages of the sea level obtained from tide gauges.

Terrence J. Hughes, George H. Denton and James L. Fastook of the University of Maine at Orono and John H. Mercer of Ohio State University have constructed models indicating that the flow of the West Antarctic ice shelves could conceivably accelerate to as much as 20 kilometers per year, raising the sea level by about three centimeters per year or three meters per century. Such a rise would have serious repercussions for coastal areas. Many glaciologists think such a high velocity is unlikely, but they acknowledge it is possible. The movement of the ice sheet could be monitored by satellite observations.

Societies have had much experience in responding to short-term environmental catastrophes: events such as hurricanes, floods, droughts, volcanic eruptions, earthquakes and forest fires. The changes that may come as a result of an increased amount of carbon dioxide in the atmosphere, however, will not be events. They will be slow, pervasive environmental shifts. They will be imperceptible to most people from year to year because of the annual range of climatic variation. It would be prudent to begin thinking now about what the changes might be and how humankind might best avoid or ameliorate the unfavorable effects and gain the most benefit from the favorable ones.

Microbiological Mining

The central role of bacteria in the leaching of copper from low-grade ore long went unrecognized. The minerals industry now stands to gain from the application of novel methods of microbiological technology

by Corale L. Brierley

The recovery of copper from the drainage water of mines was probably a widespread practice in the Mediterranean basin as early as 1000 B.C. Although such mining operations are difficult to document, it is known that the leaching of copper on a large scale was well established at the Río Tinto mines in Spain by the 18th century. What none of the miners engaged in this traditional method of mineral extraction realized until about 25 years ago is that bacteria take an active part in the leaching process. They help to convert the copper into a water-soluble form that can be carried off by the leach water. Today bacteria are being deliberately exploited to recover millions of pounds of copper from billions of tons of low-grade ore. Copper obtained in this way accounts for more than 10 percent of the total U.S. production. In recent years bacterial leaching has also been applied to the recovery of another nonferrous metal: uranium.

Recent progress in the genetic manipulation of microorganisms for industrial purposes promises to revitalize not only the bacterial leaching of metal-bearing ores but also the microbiological treatment of metal-contaminated waste water. The enthusiasm of the microbiologists working on the development of the new "biomining" techniques is matched by a need in the minerals industry to find alternatives to conventional methods of mining, ore processing and waste-water treatment. The need arises from recent trends in the industry: the continued depletion of high-grade mineral resources, the resulting tendency for mining to be extended deeper underground, the growing awareness of environmental problems associated with the smelting of sulfide minerals and the burning of sulfur-rich fossil fuels and the rising cost of the prodigious amounts of energy required in the conventional recovery methods. The current methods will surely prevail for many years to come, but biological processes are generally less energy-intensive and less polluting than most nonbiological ones, and so the role of biological technology in mining, ore processing and waste-water treatment is likely to become increasingly important.

To understand the applications of microbiology to the minerals industry several questions must be answered. What are the microorganisms that extract metals from rocks and from metal-bearing solutions, and where are they found? What biochemical functions do the microorganisms perform, and what do they need in the way of nutrients and environmental conditions to maintain their activity? What are the constraints on the commercial exploitation of such biological techniques? What impact will the new tools of genetic engineering have on the future of biomining? I shall address each of these questions in turn.

The bacteria involved in the leaching of metals from ores are among the most remarkable life forms known. The microorganisms are said to be chemolithotrophic ("rock-eating"); they obtain energy from the oxidation of inorganic substances. Many of them are also autotrophic, that is, they capture carbon for the synthesis of cellular components not from organic nutrients but from carbon dioxide in the atmosphere. The leaching bacteria live in environments that would be quite inhospitable to other organisms; for example, the concentration of sulfuric acid and of soluble metals is often very high. Some thermophilic, or heat-loving, species require temperatures above 50 degrees Celsius (122 degrees Fahrenheit), and a few strains have been found at temperatures near the boiling point of water.

For many years the only microorganism thought to be important in the leaching of metals from ores was the rod-shaped bacterium *Thiobacillus ferrooxidans.* This microorganism was discovered in the acidic water draining coal mines; it was not until 1957 that a correlation was recognized between the presence of the bacterium and the dissolution of metals in copper-leaching operations. Since then a great deal of information has been amassed on *T. ferrooxidans* and on its vital role in the leaching of metals.

T. ferrooxidans is acidophilic, or acid-loving; it tends to live in environments such as hot springs, volcanic fissures and sulfide ore deposits that have a high concentration of sulfuric acid. It is also moderately thermophilic, thriving in the temperature range between 20 and 35 degrees C. The bacterium gets energy for growth from the oxidation of either iron or sulfur. The iron must be in the ferrous, or bivalent, form (Fe^{++}), and it is converted by the action of the bacterium into the ferric, or trivalent, form (Fe^{+++}). Several forms of sulfur can be attacked. They include both soluble and insoluble sulfides (compounds containing the bivalent sulfur ion S^{--}), elemental sulfur and soluble compounds that incorporate either the thiosulfate ion ($S_2O_3^{--}$) or the tetrathionate ion ($S_4O_6^{--}$). In each case the product of the transformation is a substance in which the sulfur atom has fewer valence electrons, culminating in the formation of the sulfate ion (SO_4^{--}). *T. ferrooxidans* obtains carbon autotrophically from atmospheric carbon dioxide.

Although *T. ferrooxidans* is essential to the bacterial leaching of metals, it is by no means the only microorganism with an important role in the process. Among the other microorganisms taking part is *T. thiooxidans,* a rod-shaped bacterium not unlike *T. ferrooxidans* that grows on elemental sulfur and some soluble sulfur compounds. Studies by Donovan P. Kelly and his associates at the University of Warwick have confirmed the importance of mixed cultures of bacteria in the extraction of metals from ores. *T. ferrooxidans* and *T. thiooxidans* combined, for example, are more effective in leaching certain ores than either organism is alone. Similarly, the combination of *Leptospirillium ferrooxidans* and *T. organoparus* can degrade pyrite (FeS_2) and chalcopyrite ($CuFeS_2$), a feat neither species can accomplish alone.

In acidic environments supporting leaching bacteria one can often isolate a number of heterotrophic microorganisms: bacteria and fungi that scavenge the small amounts of organic matter present in these environments or that survive on the organic by-products of other organisms' autotrophic metabolism. The role of the heterotrophic microorganisms in the leaching process is largely undetermined. Thiobacilli that attack some sulfide minerals and certain soluble sulfur compounds under neutral conditions (that is, neither acidic nor alkaline) are often found in sulfide ore deposits and in other habitats where sulfur is available. Thiobacilli of this type may be responsible for the initial increase in acidity that establishes an environment conducive to the growth of the more acidophilic leaching bacteria.

At temperatures between 60 and 75 degrees C. and under neutral conditions the filamentous bacterium *Thermothrix thiopara* oxidizes sulfhydryl ions (HS^-), sulfite ions (SO_3^-), thiosulfate ions and elemental sulfur to form sulfate ions. There is increasing evidence of the widespread existence of *Thermothrix* species and similar filamentous, sulfur-oxidizing bacteria in thermal springs and near volcanic fissures. Few leaching sites have been tested for the presence of these bacteria. Their existence in sulfur-bearing springs, however, suggests they could colonize sulfidic ores and thereby prepare such environments for the more acidophilic species.

Among the most interesting of the leaching microorganisms are the moderately thermophilic and acidophilic bacteria designated TH (for thermophilic and *Thiobacillus*-like). Studies by James A. Brierley of the New Mexico Institute of Mining and Technology and Norman W. Le Roux of the Warren Spring Laboratory in Britain have established that strains of TH bacteria grow on ferrous iron or on minerals such as pyrite, chalcopyrite, covellite (CuS) and pentlandite [$(Fe,Ni)_9S_8$]. An organic supplement to the mineral diet is apparently required by these organisms for growth. Although the TH strains are similar in form to the rod-shaped thiobacilli, their distinctive temperature domain near 50 degrees C., their very different biochemical activities and their inability to metabolize carbon dioxide rule out a close relation with *T. ferrooxidans*. TH strains have been isolated from acidic hot springs and from leaching environments where copper sulfide is present. The role of the organisms in extractive processes has not been fully

LEACHING OF COPPER from low-grade ore is done on a huge scale in the western U.S., accounting for more than 10 percent of the country's production of the metal. The photograph is an aerial view of a large dump-leaching operation adjacent to an open-pit copper mine near Salt Lake City. The trucks backed up to the rim of the dump at the left haul as much as 170 tons of waste rock each from the mine to the dump. To recover the copper, which is dispersed at low concentration throughout the millions of tons of material in the dump, acidified water is applied to the top surface. As the water percolates through the pile, bacteria of the genus *Thiobacillus,* which are ubiquitous in such sulfidic rocks, proliferate. The bacteria greatly accelerate the leaching process by breaking down the sulfide minerals, converting ferrous iron into ferric iron and generating sulfuric acid. The ferric iron serves to oxidize the degraded sulfide minerals, yielding soluble copper sulfate, which is washed from the rock with the leaching solution and collected in catch basins for further processing. In some parts of the dump heat-releasing chemical reactions may raise the temperature to more than 80 degrees Celsius, making the environment too hot even for the moderately thermophilic, or heat-loving, bacilli. Such "hot spots" may harbor more robust bacteria.

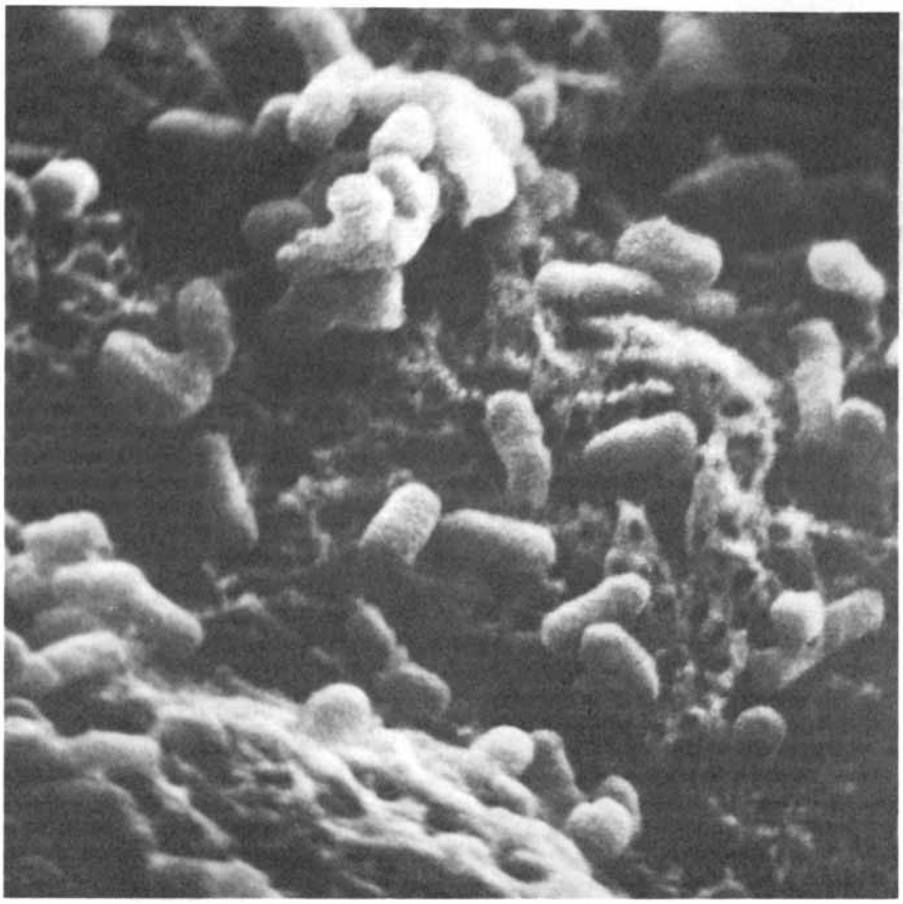
LEACHING SOLUTION is sprayed by a sprinkler system on top of a copper-leaching dump similar to the ones shown on the preceding page. The recycled waste water, from which the copper sulfate has been removed, retains much ferrous iron and some sulfuric acid.

PRECIPITATION PLANT removes copper from the leaching solution by a cementation reaction, in which ferrous iron replaces the copper in solution. Here scrap iron is used to precipitate the copper in two modules; each module consists of 13 units called precipitation cones. The leaching solution is concentrated in the two large, round "thickener" tanks. In some leaching operations solvent-extraction systems have replaced the precipitation plant, thereby eliminating the need for smelting, which is a source of sulfur dioxide pollution.

elucidated; from an industrial point of view, however, they must be considered potentially important to the development of any high-temperature biomining process.

The most robust of the leaching microorganisms are the extremely thermophilic and acidophilic species of the genus *Sulfolobus*. These bacteria flourish in acidic hot springs and volcanic fissures at temperatures that can exceed 60 degrees C. Some strains of *Sulfolobus* have been observed in springs at temperatures near the boiling point of water. The cell wall of the *Sulfolobus* bacteria has a different structure from that of most bacteria. Microorganisms of this type are thought to belong to the Archaebacteria, a group of unusual bacteria proposed as a separate kingdom of life forms [see "Archaebacteria," by Carl R. Woese; SCIENTIFIC AMERICAN, June, 1981].

Sulfolobus acidocaldarius and *S. brierleyi* oxidize sulfur and iron for energy, relying on either carbon dioxide or simple organic compounds for carbon. Ordinarily oxygen is required by *Sulfolobus;* as in other aerobic organisms, the oxygen serves as the ultimate acceptor of the electrons removed in the process of chemical oxidation. *Sulfolobus* bacteria can also grow anaerobically, however. It has been demonstrated that molybdenum (Mo^{6+}) and ferric iron can serve as electron acceptors in the absence of air. Minerals that resist most microorganisms, such as chalcopyrite and molybdenite (MoS_2), are readily attacked by *Sulfolobus,* and the resulting soluble metals are not toxic to the organism. Molybdenum, which is extremely toxic even to the metal-tolerant thiobacilli, is readily endured by *S. brierleyi* in concentrations as high as 750 milligrams per liter. *Sulfolobus* has not been isolated from commercial leaching operations, but laboratory studies confirm the ability of the organism to proliferate in such environments. The potential of *Sulfolobus* species to leach metals from ores is only now being recognized; because of the extraordinary ability of these organisms to attack resistant mineral structures, however, they are certain to be among the leaching bacteria that will be successfully exploited in the future.



ROD-SHAPED BACTERIA of the genus *Thiobacillus* are essential to the leaching of metals from sulfide minerals. Microorganisms of this type get their energy from the oxidation of inorganic substances; in addition many of them obtain carbon from carbon dioxide in the atmosphere. The species in this scanning electron micrograph, *T. thiooxidans,* is seen growing on elemental sulfur. The micrograph, which enlarges the bacteria by approximately 15,000 diameters, was made by Karen A. Couch of the New Mexico Institute of Mining and Technology.

How exactly does the growth of the leaching bacteria result in the extraction of metals from rocks? By convention bacterial leaching has been divided into two approaches: direct and indirect. Direct bacterial leaching entails an enzymatic attack by the bacteria on components of the mineral that are susceptible to oxidation. In the process of obtaining energy from the inorganic material the bacteria cause electrons to be transferred from iron or sulfur to oxygen. In many cases the more oxidized product is more soluble. It should be noted that the inorganic ions never enter the bacterial cell; the electrons released by the oxidation reaction are transported through a protein system in the cell membrane and thence (in aerobic organisms) to oxygen atoms, forming water. The transferred electrons give up energy, which is coupled to the formation of adenosine triphosphate (ATP), the energy currency of the cell.

Indirect leaching, in contrast, does not proceed through a frontal attack by the bacteria on the atomic structure of the mineral. Instead the bacteria generate ferric iron by oxidizing soluble, ferrous iron; ferric iron in turn is a powerful oxidizing agent that reacts with other metals, transforming them into the soluble, oxidized form in a sulfuric acid solution. In this reaction ferrous iron is again produced and is rapidly reoxidized by the bacteria. Indirect leaching is usually referred to as bacterially assisted leaching. In an acidic solution without the bacteria ferrous iron is stable, and leaching mediated by ferric iron would be slow. *T. ferrooxidans* can accelerate such an oxidation reaction by a factor of more than a million.

Direct and indirect leaching by bacteria are difficult to differentiate quantitatively because most minerals include some iron. Even if leaching were to begin with the direct process exclusively, the iron would be released from the mineral and would establish an indirect leaching cycle. Direct leaching by thiobacilli has been demonstrated in the laboratory with iron-free synthetic metal sulfides.

In practice the leaching of metals is far more complex than the above analysis might suggest; there are numerous processes in addition to direct enzymatic oxidation and bacterial generation of ferric iron. Some chemical reactions between ferric iron and metal-sulfide minerals result in the formation of secondary minerals and elemental sulfur, which can "blind," or inactivate, the reactive surfaces. When sulfur is formed, *T. thiooxidans* plays an indispensable role in oxidizing the sulfur to sulfuric acid, thus exposing the metal for further leaching.

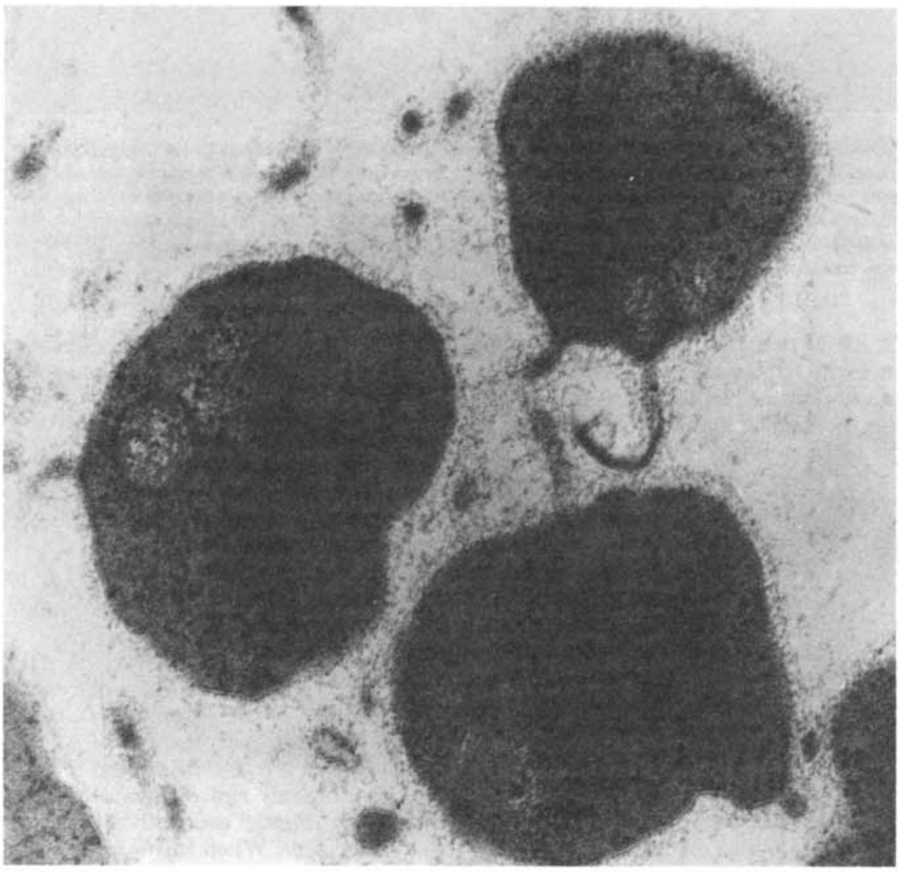
The control of acidity is of utmost importance in leaching, because an acidic environment must be maintained in order to keep ferric iron and other metals in solution. Acidity is controlled by the oxidation of iron, sulfur and metal sulfides, by the dissolution of carbonate ions and by the decomposition of ferric iron through reaction with water. The last reaction promotes leaching by generating hydrogen ions (which make the solution more acidic), but it may also be detrimental because precipitates of basic ferric sulfates may inactivate the surfaces of metal-sulfide minerals and in some cases may even prevent the flow of the leaching solution. The chemical and biological processes are part of a complex system whose functioning depends on elements of hydrology, geology, physics and engineering.

In commercial dump-leaching operations millions of tons of sulfidic overburden and waste rock, containing small but valuable quantities of copper and other metals, are transported by truck or train from open-pit copper mines to the dump site. The dump is often built in a valley to take advantage of natural slopes for maintaining the stability of the pile and for facilitating the recovery of the applied solutions. Such a dump can be immense, with slopes up to 1,200 feet high holding four billion tons of material. To these impressive formations thousands of gallons of acidified water are applied by flooding or sprinkling the top surface. Sprinkling introduces air, a vital component of both the chemical and the biological oxidation reactions, into the solution.

The dumps are not inoculated with the leaching bacteria. The organisms are ubiquitous, and when conditions in the rock pile become suitable for their growth, they proliferate. Rock samples collected near the top of a leach pile typically harbor more than a million bacteria of the species *T. ferrooxidans* per gram; *T. thiooxidans* bacteria are present in somewhat smaller numbers. The leach solution percolates through the leach dump, and the "pregnant," or metal-laden, solution is collected in catch basins or reservoirs at the foot of the dump. The copper is removed from solution either by a cementation reaction, in which ferrous iron replaces the copper in solution, or by solvent extraction, in which the copper is concentrated by transferring it from the aqueous leaching solution to an organic solution. The "barren," or copper-free, solution is then recycled to the top of the dump.

In a large dump-leaching operation the ore is processed for many years to recover as much of the copper as possible. Because of the construction methods employed and the volume of the solid material treated, dump leaching is a crude operation. The placement of the dump in a natural valley can impede the flow of air to the interior of the pile. The large size of some of the rocks limits contact among the metal-sulfide minerals, the oxidizing solution and the bacteria. During the dump's construction large ore haulers compact the surface, creating impermeable zones in the pile. Gypsum ($CaSO_4$), ferric hydroxide [$Fe(OH)_3$] and basic ferric sulfate precipitates also decrease permeability, further reducing the contact of the solution with the sulfide rocks.

Although studies of bacteria in the dump environment have barely begun, some factors that may adversely affect the populations are known. They include high metal concentrations, particularly of ions such as silver and mercury that are known to be toxic to the organisms, lack of air, and temperatures higher than those tolerated by the organisms. Because the limitations of dump leaching are more clearly defined today than they were in the past, considerable forethought now goes into the construction of the leaching piles. Special "finger" dumps allow greater air circulation, and haulage is controlled to minimize compaction. From a biological viewpoint, however, dump leaching remains an essentially uncontrolled process.



SPHERICAL BACTERIA of the genus *Sulfolobus* oxidize sulfur and iron under extreme conditions of temperature and acidity. The single-cell microorganisms are classified as Archaebacteria, a distinct grouping of primitive life forms. Species of *Sulfolobus* thrive in environments such as hot springs and volcanic fissures, where they get energy from the degradation of minerals in solution and carbon from carbon dioxide in the atmosphere. They can readily leach metals from minerals such as chalcopyrite and molybdenite, which resist attack by most other microorganisms. The large difference in acidity between the inside of the cell, which is almost neutral, and the outside, which in most cases is highly acidic, may be important to the survival of the microorganisms in extreme environments. The electron micrograph, made by Joseph V. Scaletti of the University of New Mexico, shows cells of the species *S. brierleyi*, enlarged by some 80,000 diameters. The unusual cell wall is made of hexagonally packed protein subunits.

Extractive methods other than dump leaching offer somewhat more regulation of biological, chemical and engineering factors. Heap leaching, for example, is used to extract metals from sulfide and oxide minerals in ores of a somewhat higher grade than those subjected to dump leaching. In heap leaching the rocks are often crushed to avoid the solution-contact problems encountered when leaching large boulders, and the heaps are built up on impermeable pads to prevent loss of the solution into the underlying soil. Aeration systems have been installed to increase the flow of air in the piles.

In-place leaching is a promising technique for the recovery of metals from low-grade ores in inaccessible sites. This technology, which has minimal impact on the environment, is currently employed to extract residual minerals from abandoned mine workings and to recover uranium from low-grade deposits. To leach metals from depleted mine workings the leaching solution is applied directly to the walls and the roof of an intact stope (an underground excavation from which ore has been removed) or to the rubble of fractured workings. In-place leaching techniques have been successful in the recovery of both copper and uranium.

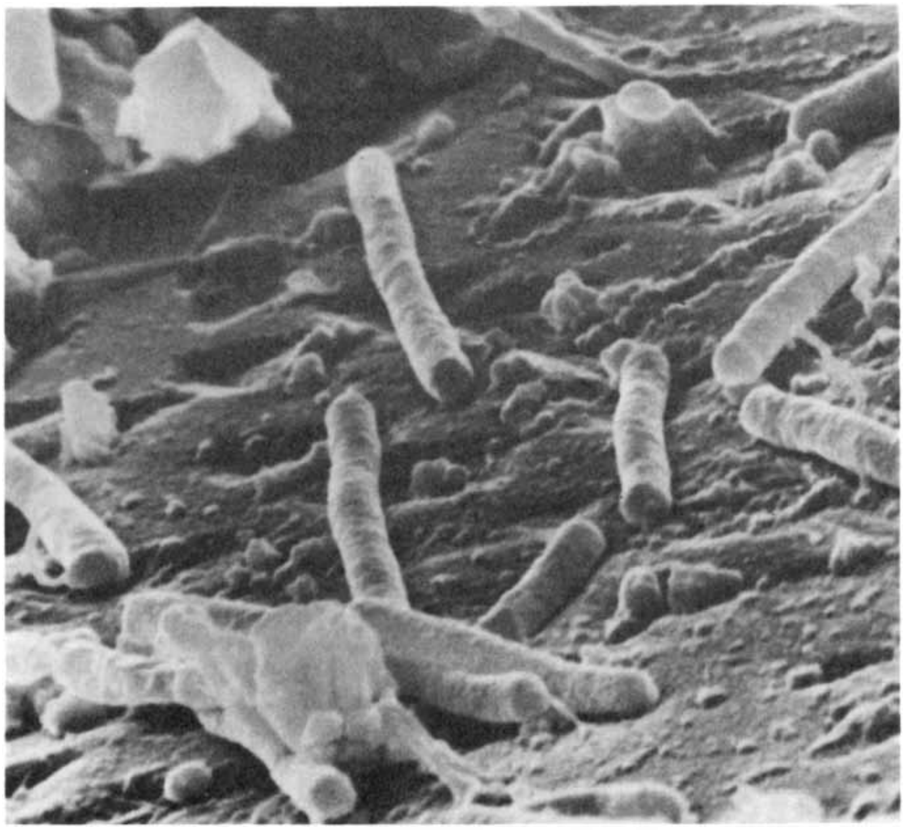
In the extraction of uranium the bacteria do not directly attack the uranium mineral; instead they generate ferric iron from pyrite and soluble ferrous iron. Ferric iron readily attacks minerals incorporating quadrivalent uranium (U^{4+}), converting this ion into hexavalent uranium (U^{6+}), which is soluble in dilute sulfuric acid.

The bacterially assisted leaching of uranium could in principle be applied not only to the recovery of residual metals but also to the in-place leaching of low-grade uranium ore bodies. There are a number of low-grade deposits in the western and southwestern U.S., but they are low in pyritic material and include rocks that tend to neutralize acids. Such mineralogical conditions are not conducive to bacterial leaching, and an exclusively chemical method of leaching has been adopted instead. Wells are drilled into the isolated ore bodies at depths ranging from tens to hundreds of feet. A carbonate solution containing an oxidant is then injected into the mineralized zone, where the uranium is solubilized. The uranium-bearing solution is withdrawn from the formation through a precisely engineered pattern of recovery wells.

The same practices might be applied in formations suitable for in-place bacterial leaching. Further study is needed, however, to assess such factors as the effect of hydrostatic pressure on the leaching bacteria and the loss of permeability in the formation resulting from bacterial growth.

Vat leaching is employed mainly for the chemical extraction of copper from oxide ores. The technology is based on the controlled agitation of concentrated ore particles with precisely determined amounts of acid. Vat leaching is seldom used for sulfide ores owing to the need for an oxidant and to the long leaching times; these requirements, however, do not preclude the vat leaching of sulfide ores under certain circumstances. Studies by A. Bruynesteyn of British Columbia Research and by Arpad Torma of the New Mexico Institute of Mining and Technology indicate that the bacterial leaching of sulfidic concentrates can be competitive with existing methods of extraction.

Although bacterial leaching is currently exploited only for the recovery of copper and uranium, the appetite of the leaching bacteria is fairly nonspecific. The organisms readily degrade other sulfide minerals, yielding zinc from sphalerite (ZnS) and lead from galena (PbS). The leaching bacteria can therefore be considered for the extraction of many other metals. The bacteria readily catalyze the dissolution of inorganic sulfur from coal, and recent advances indicate that organic sulfur may also be vulnerable to microbiological attack. Progress in this area suggests that the precombustion desulfurization of sulfur-laden coal is on the horizon.



SUBSIDIARY ROLE in leaching may be played by bacteria of the species *Thermothrix thiopara,* shown here growing on pyrite. These microorganisms, which inhabit moderately warm springs under neutral conditions (neither acidic nor alkaline), are capable of oxidizing sulfur aerobically in the presence of air or anaerobically in the presence of nitrates. *Thermothrix* species have not been assigned a specific function in the extraction of metals by leaching, but by generating sulfuric acid they may create an environment favorable for more thermophilic and acidophilic species. The scanning electron micrograph, made by Thomas L. Kieft and Gregory P. Minion of the University of New Mexico, enlarges the cells by about 130,000 diameters.

Given the seemingly limitless capability of bacterial leaching for the recovery of metals, why has such a potentially powerful tool not been more widely exploited? Probably the main reason is that the technology was not much needed until now, when energy is costly and the supply of accessible, high-grade metal ores is limited. Now that a genuine need exists for biomining, the technology lacks refinement. The study of bacterial leaching has largely been in the hands of a few investigators who have been systematically examining the organisms and the reactions they carry out. There has been no coordinated engineering effort to harness the organisms and their beneficial reactions in controlled fermentation tanks. To do so will require a carefully directed effort and large sums of money. The effort will probably be financed by private industry. The difficulties in development are likely to include those that can befall any engineering project as well as a number of other problems heretofore unknown. Unlike the carefully nurtured microorganisms that yield chemical and health-care products, the leaching bacteria will be subject to adversities such as severe weather, erratic fluctuations in acidity, a continual onslaught of competitive "wild type" bacteria and variations in the concentration and type of the mineral feedstock.

It seems probable that at some time in the future genetic manipulation may produce leaching microorganisms with highly desirable characteristics. Two of the qualities sought are an enhanced tolerance of toxic metals and a greater ability to generate the oxidant, ferric iron. The genetic modification of the organisms may prove difficult, however, because little is known about the genetics of the leaching bacteria or the exact mechanisms they rely on for the degradation of minerals. The latter problem is particularly troublesome, since it is uncertain which bacterial characteristics should be genetically altered to enhance the organisms' leaching capabilities.

Increasing the leaching rate of the bacteria is certainly one of the most sought-after improvements. The major drawback of bacterial leaching has been the slowness of the biological process compared with some chemical methods of extraction. The disadvantage is most striking when the microbiological leach-

SWIRLING FILAMENTS of organic matter may represent a source of new microbial strains better suited to the requirements of the minerals industry. The filaments, which are seen in a sulfurous hot spring in Yellowstone National Park, are made up of the fused cells of millions of microorganisms of various species. The genetic manipulation of certain species found in such natural habitats may result in new strains with an enhanced capability for recovering nonferrous metals from sulfide ores or from metal-contaminated waste water. The photograph was made by James A. Brierley of the New Mexico Institute of Mining and Technology.

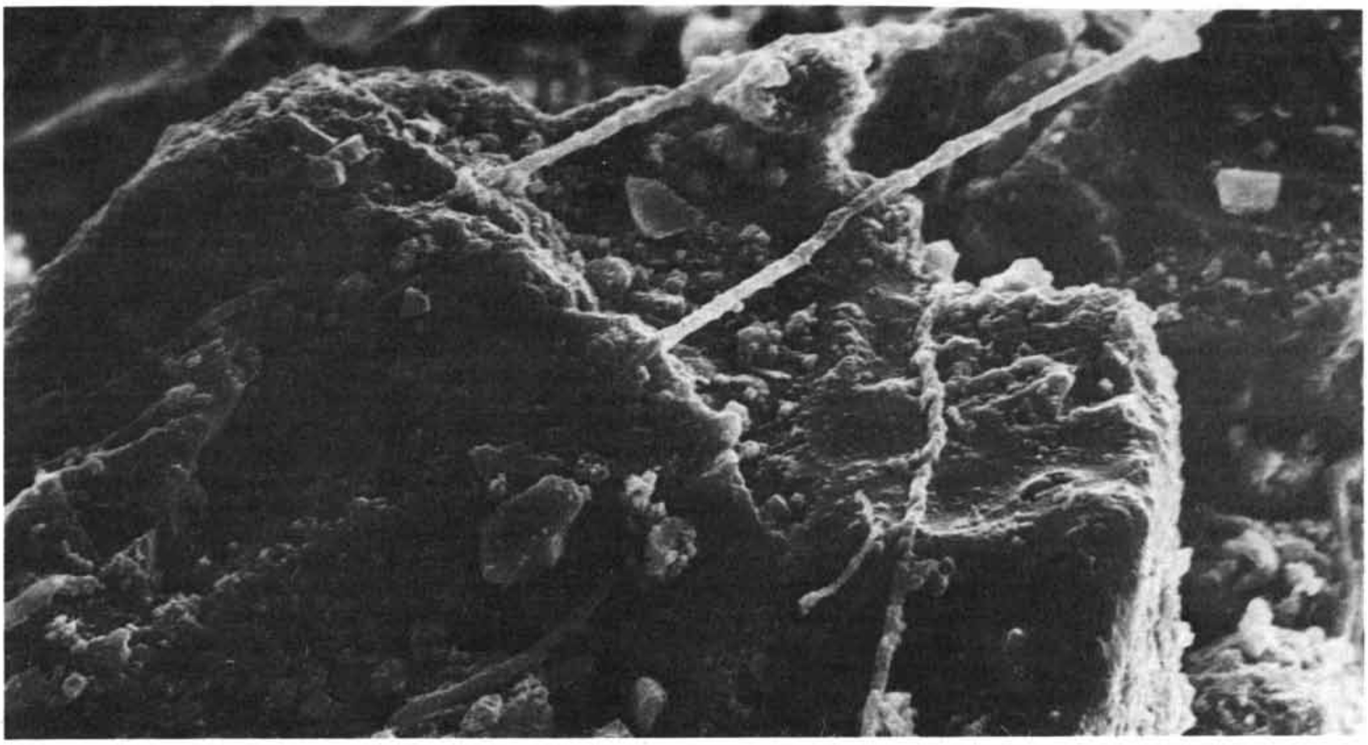
ing rate is compared with some high-temperature, high-pressure extraction processes in which strong oxidants act on finely ground particles. Valid comparisons between biological and chemical processes, however, must be based on economic factors.

Bacterial leaching is likely to find its greatest application in the controlled treatment of vast quantities of solid materials, such as discarded waste rock, overburden and tailings, in which small quantities of valuable metals are widely dispersed. For example, the U.S. has vast sulfide deposits, including more than seven billion tons of material with an average nickel content of .2 percent. At current prices the nickel is worth approximately $60 billion. The nickel is not mined because of inefficiencies in current mining and extractive technologies and because of possible detrimental effects on the environment. Advances in biomining technology may make it possible to recover not only some of the nickel but also some of the 400 million pounds of cobalt (worth $10 billion at current prices) found in the same sulfidic material. In this situation the emphasis is not so much on rapid reaction rates as it is on lower capital investment, greater recovery of metal and reduced environmental damage.

In spite of the many uncertainties associated with the genetic engineering of leaching bacteria and the scale-up of microbiological leaching processes, there is little doubt that such technologies will be developed. Interest in industrial microbiology, a need to develop more cost-effective methods for the recovery of metals from low-grade mineral deposits and the desire to utilize processes that are environmentally benign all favor this development.

The controlled use of microorganisms for the extraction of metals from ores and from solid waste materials is paralleled by the application of biological technology to the restoration of particulate-laden and metal-contaminated industrial waste water. Such technology not only would aid in clarifying the waste water but also would allow the recovery of valuable metals. Now under investigation are bacteria, algae and fungi that readily accumulate inorganic ions present in dilute concentrations in waste streams. The microbiological processes for the removal of metals from solution can be divided into three categories: the adsorption of metal ions onto the surface of a microorganism, the intracellular uptake of metals and the chemical transformation of metals by biological agents.

Most microorganisms have a negative electric charge owing to the presence of negatively charged groups of atoms on the cell membrane and the cell wall. The charged groups, or ligands, include phosphoryl (PO_4^{---}), carboxyl (COO^-), sulf-



EXTRAORDINARY ADAPTABILITY displayed by certain microorganisms in harsh volcanic environments is demonstrated by the metal-encrusted, filamentous bacteria seen in this scanning electron micrograph, made by John A. Baross of Oregon State University. The bacteria were found growing at a temperature of 100 degrees C. on rocks collected in May, 1981, from within the newly formed crater of Mount St. Helens. According to Baross and his colleagues, who have made a study of the microbial colonization of the site, the bacteria are present in great numbers in the cooling debris of the volcano's eruption. The magnification is roughly 25,000 diameters.

hydryl (HS^-) and hydroxyl (OH^-) groups and are responsible for the adsorption of positively charged metal ions from solution. The adsorption is typically rapid, reversible and independent of temperature and energy metabolism. The common beer yeast *Saccharomyces cerevisiae* and the fungus *Rhizopus arrhizus* have recently been shown to adsorb uranium from waste water. For *S. cerevisiae* the acidity level at which the surface binding of uranium is optimal suggests that the positively charged uranium ions are attracted to the negatively charged ligands on the cell. Uranium concentrations representing between 10 and 15 percent of the dry weight of the cell have been recorded for this yeast. *R. arrhizus* adsorbs uranium to the extent that the metal amounts to 18.5 percent of the dry weight of the cell. This is more than twice the uptake of a commercially available ion-exchange resin.

The deposition of insoluble metals has been observed at the surface of some microorganisms. The sheathed, filamentous bacteria of the *Sphaerotilus-Leptothrix* group and the polymorphic *Hyphomicrobium* can become encrusted with oxides of manganese. *Sphaerotilus-Leptothrix* and *Gallionella* are called iron bacteria because they deposit a sheath of iron on their twisted stalks.

Microorganisms ordinarily take up some ions that are necessary for cellular activity. The transport systems for the ions are dependent on both temperature and energy. Examples of trace substances that are readily transported into the cells of microorganisms are magnesium (Mg^{++}), calcium (Ca^{++}), potassium (K^+), sodium (Na^+) and sulfate (SO_4^{--}). Although the mechanisms by which the cells assimilate the ions are highly selective, substitutions are possible. For example, the negatively charged chromate (CrO_4^{--}), selenate (SeO_4^{--}), vanadate (VO_4^{--}), tungstate (WO_4^{--}) and molybdate (MoO_4^{--}) ions can be transported into the cells of some microorganisms by the system that ordinarily carries sulfate ions.

The most bizarre phenomenon of this kind is the reported intracellular accumulation of very high concentrations of toxic metals. The common soil and water bacterium *Pseudomonas aeruginosa* accumulates nearly 100 milligrams of uranium per liter of solution in less than 10 seconds. With the aid of an electron microscope Starling E. Shumate II and Gerald W. Strandberg of the Oak Ridge National Laboratory found that only 44 percent of the cells accumulated the uranium. This observation implies that in the active cells uranium accounts for as much as 56 percent of the dry weight of the cell. The mechanism that drives the cells to virtually commit suicide by the accumulation of toxic metals is not known; nevertheless, the phenomenon may someday be exploited for the restoration of metal-contaminated waste water.

Other microbiological mechanisms with a potential for waste-water treatment eliminate metals by precipitating them in chelates, or cagelike compounds, and by incorporating them into volatile compounds that can later be evaporated. Many microorganisms synthesize specific chelation compounds that immobilize heavy metals. More commonly in natural systems, however, metals adsorbed or absorbed by plants and microorganisms become sequestered in sediments following the death of the organisms.

Precipitation results when hydrogen sulfide (H_2S), which is generated by sulfate-reducing bacteria, combines with a metal to form an insoluble metal sulfide. The exploitation of microorganisms such as *Desulfovibrio* for the generation of hydrogen sulfide has considerable potential for application in the treatment of metal-contaminated effluents. The generation of ferric iron by *T. ferrooxidans* serves to remove soluble iron from acidic mine wastes. The organisms, attached to plastic disks, are immersed in acidic waters containing ferrous iron. Through oxidation the bacteria convert the iron into the less soluble ferric form. This pretreatment is expected to reduce the cost of neutralizing the acids in the drainage water.

The methylation of metals (that is, the substitution of a metal atom for the hydrogen atom of the hydroxyl group of a methyl alcohol molecule) can result in the metal's becoming volatile. Metals and metalloids known to be subject to biomethylation include mercury, selenium, tellurium, arsenic, tin, lead and cadmium. It has been predicted that platinum, palladium, gold and thallium can also be transformed in this way. The commercial applicability of biomethylation is questionable, however, owing to the difficulty of trapping the volatile compounds and the extreme toxicity of some of them.

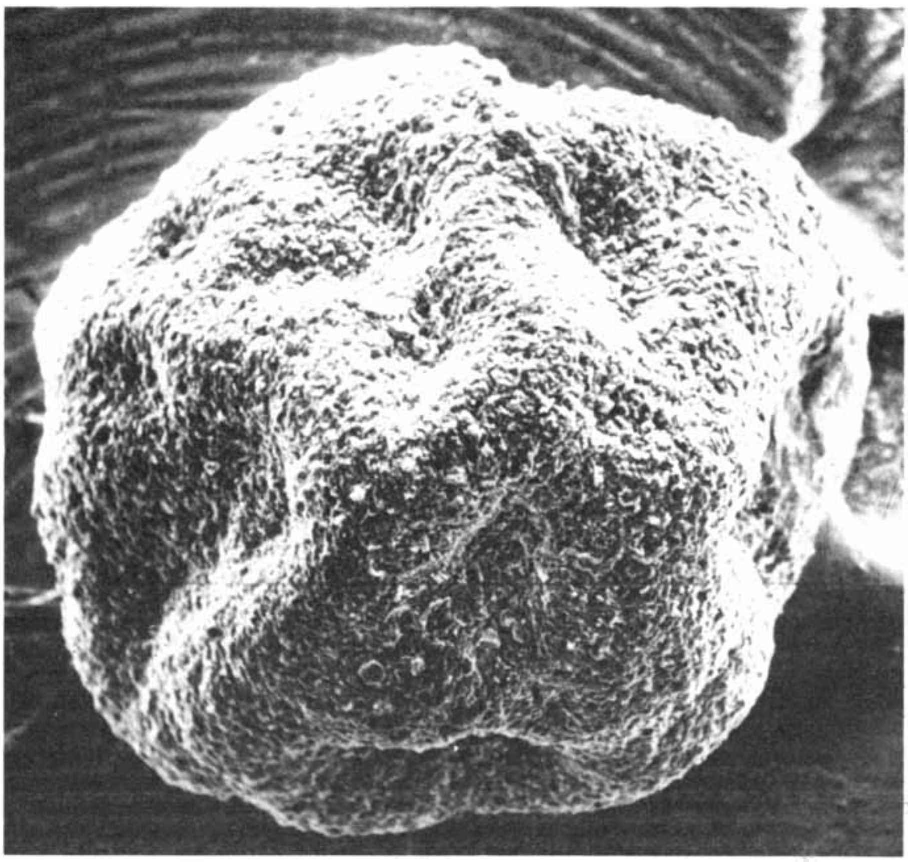
The microbiological processes mentioned above for the recovery of metals and metalloids from solution have been observed in the laboratory and in natural environments where conditions are suitable for specific types of biological activity. An example is Schist Lake in Manitoba, which receives metal-contaminated tailings from a mining and smelting operation and sewage effluent from a small town. Nutrients in the sewage promote algal blooms, and the algae in turn accumulate metals. The decay of the metal-laden algae is mediated by microorganisms that generate hydrogen sulfide, which precipitates the metals as sulfides.

Observations of such natural cleansing systems have encouraged workers in the mining industry to try to imitate them. An artificial meandering stream draining a tailing pond in the New Lead Belt of Missouri is inhabited by the algae *Spirogyra, Rhizoclonium, Hydrodictyon* and *Cladophora.* The microorganisms are credited with the removal of nutrients and soluble heavy metals and the entrapment of suspended mineral particulates. A sedimentation pond and a baffled outlet prevent the algae from escaping into the receiving stream. At a uranium mine in the Grants Uranium District of New Mexico the mine water is treated by means of an interconnected system of impoundments in which the algae *Spirogyra, Oscillatoria, Rhizoclonium* and *Chara* accumulate soluble ions of molybdenum, selenium, uranium and radium. The sediments of the impoundments are enriched with the metals and metalloids, suggesting an ion-removal scheme analogous to the natural process observed at Schist Lake. Large populations of sulfate-reducing bacteria in the sediments indicate that hydrogen sulfide may have a role in the sedimentation of the ions. Uranyl carbonates [$UO_2(CO_3)^{--}$ and $UO_2(CO_3)_3^{4-}$] probably interact with *Chara,* which is encrusted with crystalline hydrous carbonates, to form the mineral liebigite [$Ca_2UO_2(CO_3)_3 \cdot 10H_2O$].

The microbiological processes currently exploited by the minerals industry are fairly simple in engineering design, and their effectiveness is sensitive to seasonal changes and sudden alterations in the chemistry of the system. Recent developments in the study of the uptake of metals by *S. cerevisiae, R. arrhizus* and *P. aeruginosa* make it probable that these microorganisms can be utilized in precisely engineered processes for the recovery of metals from waste-water streams. The new tools of genetic engineering may well lead to the creation of modified organisms with greater effectiveness in metal removal. It was noted above that only 44 percent of the *P. aeruginosa* cells take part in the uranium-uptake process. Shumate and Strandberg speculate that if the factor or factors controlling the uptake can be identified, the bacteria may be genetically altered to increase the population that accumulates the metal.

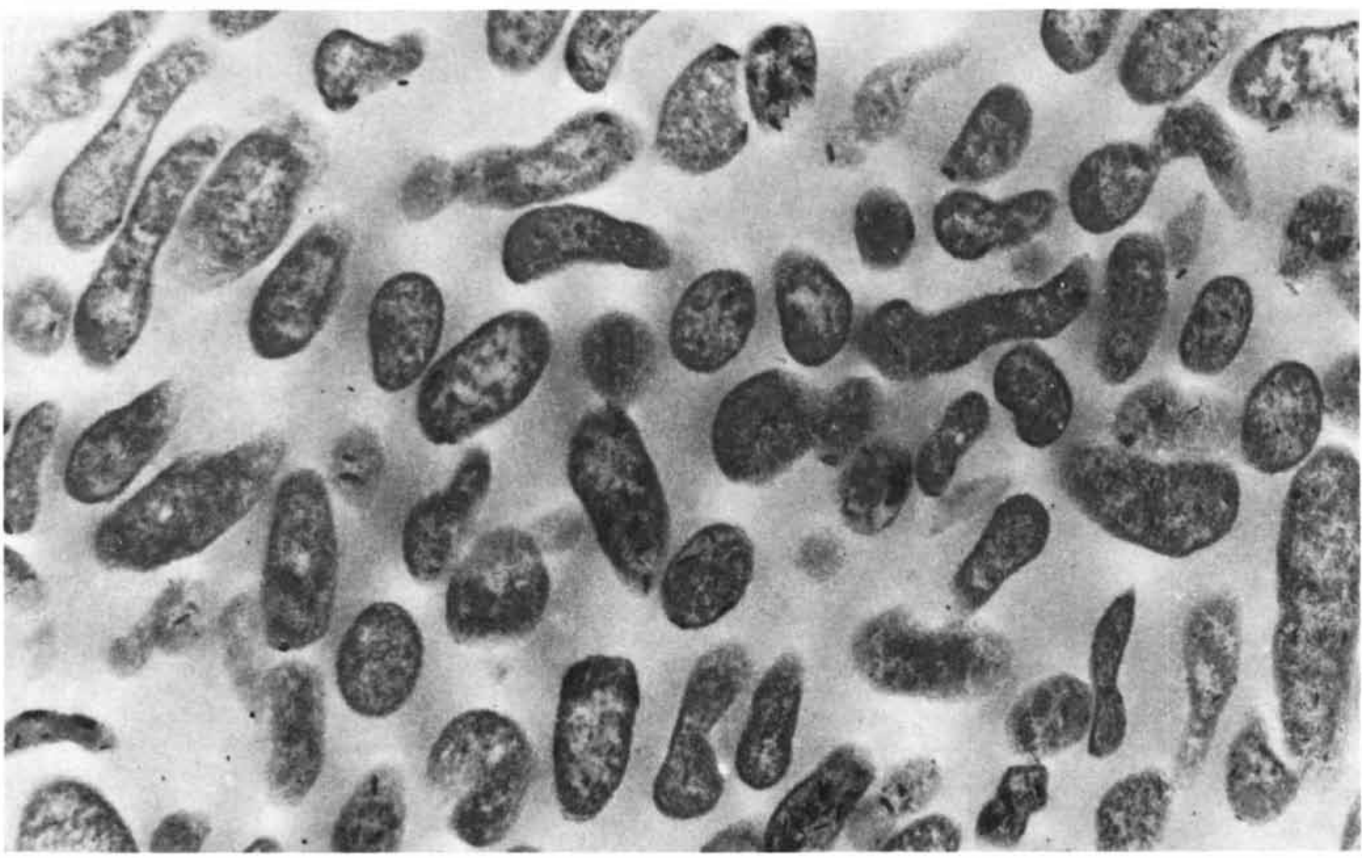
The accumulation of metals by microorganisms, whether the process is intracellular uptake or surface accumulation, is fairly nonspecific: negatively charged groups of atoms on the surface of microorganisms attract any positively charged ions in the solution. Many organisms have cellular components that are highly metal-specific. One of the best-understood metal-binding agents is the protein metallothionein. Structural studies of metallothionein indicate there is a high concentration of sulfur-containing amino acid units, which, when they are brought into juxtaposition by the folding of the protein chain, form a sulfhydryl (HS^-) chelation site. In the marine blue-green alga *Synechococcus* a comparatively small cadmium-binding metallothionein can bind an average of 1.28 atoms of cadmium per molecule of protein. The identification of the gene or the genes that specify the structure of the metallothionein of this organism or any other may enable geneticists to isolate and clone the genes in selected microorganisms. Cells carrying the cloned genes could be directed to synthesize massive quantities of metallothionein with a specific metal-binding capacity. The small protein could be immobilized on an inert carrier and waste water contaminated with metals could be passed over the fixed protein. Further studies of metallothioneins with the ability to bind specific metals may provide clues for the laboratory synthesis of simple compounds with an increased metal-binding capacity.

Based on nature's own workings, on controlled laboratory studies done by many workers and on current applications in the field, it is clear that microorganisms and their versatile activities will help man to lay claim to mineral wealth buried deep in the ground or available in amounts not economically feasible to recover at present. These small servants of man promise to help in cleaning the air and water while retrieving valuable metal resources.



MICROSCOPIC PELLET formed by the filamentous fungus *Cladosporium cladosporoides* encases suspended particles of clay in this demonstration of the application of microbiology to waste-water treatment. The pelletization of suspended particulates may be well suited to areas such as central Florida, where millions of gallons of water must be treated to settle suspended clay before the water can be recycled for the mining and beneficiation of phosphates. The particle is enlarged by about 54 diameters in this scanning electron micrograph, made by Couch.

EXTRACELLULAR ACCUMULATION of uranium ions by cells of the common beer yeast *Saccharomyces cerevisiae* is recorded in an electron micrograph made by Gerald W. Strandberg of the Oak Ridge National Laboratory. The positively charged uranium ions in the test solution are attracted to negatively charged groups of atoms associated with the membrane and the wall of the yeast cells, forming a layer of needlelike uranium crystals around each cell. The electron micrograph magnifies the yeast cells by approximately 30,000 diameters.



INTRACELLULAR ACCUMULATION of uranium ions by the common soil and water bacterium *Pseudomonas aeruginosa* is another potentially attractive approach to the microbiological restoration of metal-contaminated waste water. According to workers at Oak Ridge, the bacteria are capable of accumulating as much as 100 milligrams of uranium per liter of solution in less than 10 seconds. In the active cells the concentration of uranium can reach 56 percent of the dry weight of the cell. The magnification is about 20,000 diameters.

The Coronas of Galaxies

Satellite observations indicate that the entire disk of our galaxy has an extended envelope of hot gas. The same appears to be true of other spiral galaxies

by Klaas S. de Boer and Blair D. Savage

The familiar picture of a spiral galaxy is a thin disk of stars, gas and dust with a somewhat thicker central bulge. Our own galaxy is such a spiral system, with a disk some 100,000 light-years in diameter. Surrounding the disk is a region called the halo, thinly populated by old stars and globular star clusters. The disk of the galaxy has been patiently mapped by means of radiation emitted by the gas in it, and the halo has been mapped by means of radiation emitted by the stars in it. There is a third component of the galaxy that is difficult to detect from the earth. Called the galactic corona, it is an envelope of hot gas that extends tens of thousands of light-years on each side of the central plane of the galactic disk.

Given the difficulty of detecting the corona from the earth, how was it discovered and how is it studied? The answer is that the hot gas of the corona absorbs ultraviolet radiation. Therefore if the gas lies between an ultraviolet-emitting star and an ultraviolet-detecting instrument, the gas can be detected by its absorption of the star's ultraviolet radiation. Most ultraviolet wavelengths, however, cannot penetrate the earth's atmosphere, so that such observations can be made only from above the atmosphere. This was one of the reasons the satellite *International Ultraviolet Explorer* (*IUE*) was launched in 1978. Since that time the instruments of the satellite have begun to yield a rich picture of the galactic corona and its dynamic physical processes.

Until about 20 years ago it was believed that much of the volume on each side of the central plane of the galaxy held no gas. It is now thought that in this volume are gigantic streams of gas ascending from the galactic plane and descending toward it. The force that drives the streams is the explosion of supernovas in the galactic plane. Such an explosion heats the surrounding gas to a temperature of a million degrees Kelvin (degrees Celsius above absolute zero). The hot gas moves outward from the galactic plane, cooling as it does so. It then begins to condense and move back toward the plane of the galaxy, where the process may begin again. The entire cycle has been called the galactic fountain.

In addition to securing the first direct evidence for the corona of our galaxy the instruments of the *International Ultraviolet Explorer* have probably detected coronas around two neighboring galaxies. Indeed, coronas of hot gas may be common in the universe. If they are, they may provide a simple explanation for a long-standing problem about the light of the enigmatic quasars.

The galactic corona has been named by analogy with certain other phenomena. "Corona" is from the Latin for "crown" or "garland." In astronomy the word has been employed to describe the extended envelope of gas that surrounds some celestial objects and also to describe the light emitted or scattered by such gas. The most familiar corona is of course the corona of the sun. The solar corona cannot be seen with the unaided eye except during a total eclipse of the sun, when it appears as a luminous halo around the dark eclipsing disk of the moon. The light that appears to originate in the solar corona is mainly the light of the sun's luminous disk, scattered by free electrons in the corona's diffuse gas.

The spectrum of the radiation emitted by a celestial object is a signature of the object's temperature. The spectrum of the solar corona shows that the corona is extraordinarily hot: between one and two million degrees K. (The temperature of the sun's visible surface is 6,000 degrees.) Observations made with satellite instruments capable of detecting X rays indicate that many stars have hot coronas much like the corona of the sun. It has also been shown that planets and comets have coronas. The coronas of planets and comets, however, are much cooler than those of stars, having a temperature of at most a few thousand degrees K.

The evidence that has been gathered for the existence of a corona around the galaxy has created a problem of terminology for astronomers. The trouble arises from the fact that the meaning of the word corona overlaps that of the word halo. "Halo" is from the Greek *halos,* which originally meant the disk of the sun or the moon. Aristotle used the word in this sense; it was only later that it came to mean the nimbus around the head of a saint. In meteorology "halo" refers to the ring of scattered light around the sun or the moon when they are seen through cirrus clouds.

In astronomy the meaning of the word halo has gradually become limited to the region around the galactic disk populated by old stars and globular clusters. When it was concluded that this spherical region might contain gas, the envelope of gas was called the gaseous halo by some astronomers and the corona by others. Here we shall reserve "halo" for the population of old stars and "corona" for the gaseous envelope of the galaxy.

The disk of a galaxy such as ours has spiral arms that appear much brighter than the regions between them. The concentration of stars in the spiral arms is not, however, as high as it might seem; the reason is that the arms contain the hottest and most luminous stars. Such stars are short-lived, with a life span of from one to 10 million years. They make up only a small minority of the stars in the galaxy. Most stars, the sun among them, are less brilliant, have lifetimes of billions of years and are uniformly distributed in the galaxy.

Much of the gas in a spiral galaxy is concentrated near the central plane of the galactic disk. The gravitational forces exerted by the stars in the central plane tend to hold the gas in it. The thickness of the layer of gas is less than 1 percent of the diameter of the disk.

Although the disk of interstellar gas occupies a small volume compared with the volume of the galaxy as a whole, it is the scene of much activity of astronomical significance. One current hypothesis is that the gas of the disk is itself compressed into spiral arms by the gravitational forces exerted by all the rotating material in the disk. If the density of the gas in a particular region of the disk gets high enough, the mutual gravitational attraction of its own atoms or molecules may cause it to begin to coalesce. The process can result in the formation of the dense dark clouds in which stars originate. The gas can also be compressed by the blast wave from the explosion of a supernova: a massive star at the end of its evolutionary development. Furthermore, the gas of the disk has a magnetic field that may serve to maintain the spatial relations among some of the structures of the galaxy. The magnetic field may have loops that extend away from the central plane and into the galactic corona.

It was not until the early 1950's that the extent and significance of the gas in the central disk began to be appreciated. The best evidence for its existence came from observations made with radio telescopes. Most of the disk gas is hydrogen, roughly half in the atomic form (H) and half in the molecular (H_2).

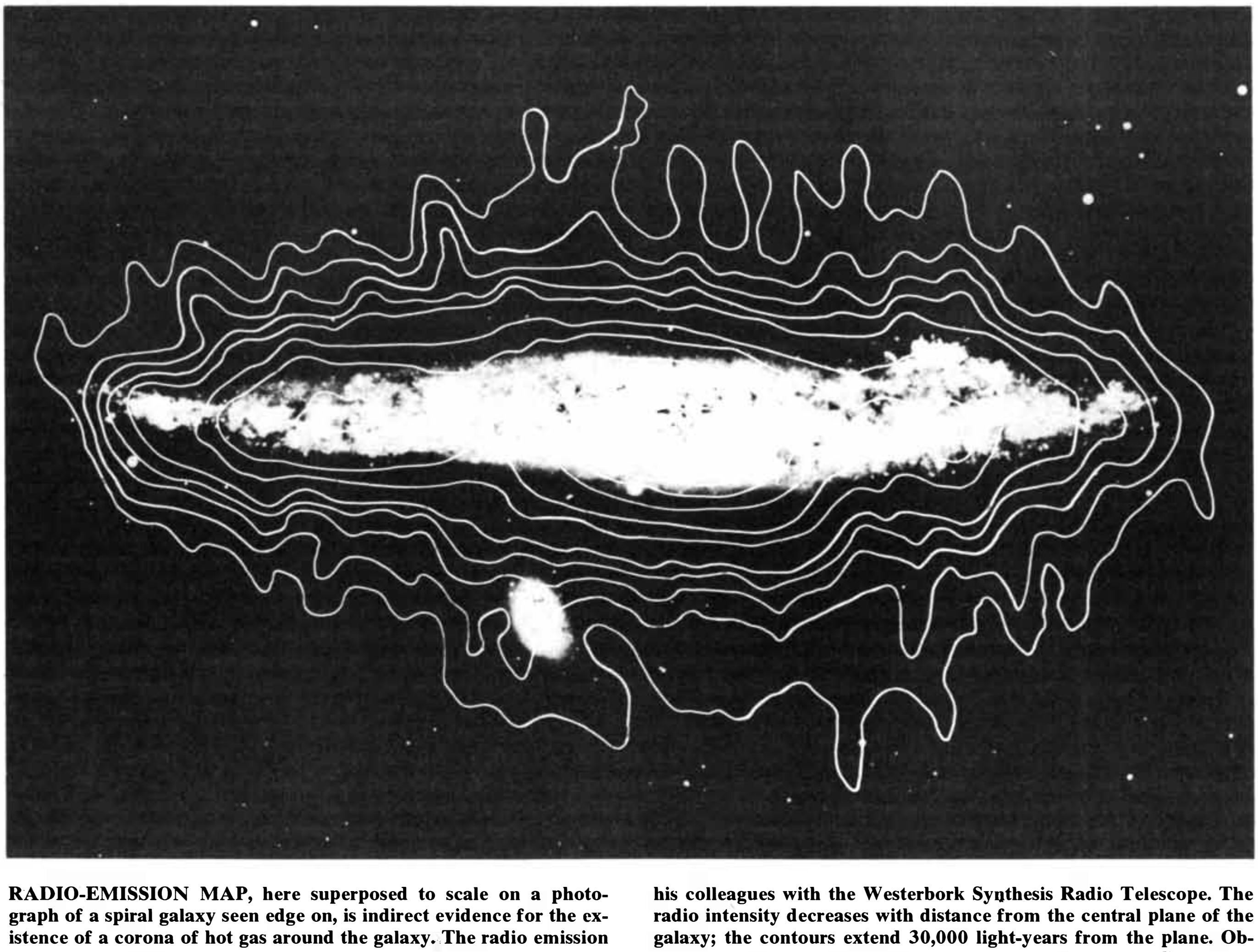
Like the electrons of other atoms, the single electron of a hydrogen atom is not free to move randomly around the atomic nucleus; it must remain in certain strictly specified patterns or states. The states differ according to the amount of energy the electron possesses in each one. The transition of an electron from one state to a state of higher energy requires that a specific quantity of energy be put into the atom; if the electron reverts to the original state, the same quantity of energy is released.

The energy needed to change the state of an electron may come from one of several sources, including collisions between atoms and electromagnetic radiation from nearby stars. If the energy is in the form of radiation, a particular wavelength is required. The reason is that the energy needed to change the state of the electron is rigidly specified. The energy of any form of radiation is inversely proportional to its wavelength; the higher the energy, the shorter the wavelength. Therefore only radiation of a certain wavelength can accomplish a particular electron transition. Since the same quantity of energy is released when an electron returns to its initial state, the emitted radiation will have the same wavelength as the radiation that caused the transition. Because a particular atom is capable of only certain transitions the wavelengths that are absorbed or emitted as radiation passes through a cloud of gas yield information about the atoms making up the gas.

In the case of the cool gas in the galactic disk the energy needed to change the state of the electron is furnished by collisions between hydrogen atoms. Because of the low temperature of the gas such collisions are followed by only the smallest electron transition. Therefore when the electron reverts to the lower state the radiation emitted has a long wavelength: 21 centimeters, which is in the radio range.

With radio telescopes 21-centimeter radiation has been exploited to detect disk gas at large distances from the sun, which is about 25,000 light-years from the center of the galaxy. Hydrogen has been detected at distances as great as 100,000 light-years. Since the disk is about 100,000 light-years across, such a distance implies that the gas is present throughout the disk.

Furthermore, the radio-telescope observations indicate that the gas of the disk participates in the rotation of the galaxy about an axis perpendicular to



RADIO-EMISSION MAP, here superposed to scale on a photograph of a spiral galaxy seen edge on, is indirect evidence for the existence of a corona of hot gas around the galaxy. The radio emission is not from gas atoms but from free electrons spiraling around lines of force in the galaxy's magnetic field. The galaxy is NGC 4631. The radio emission was mapped at a frequency of 1,412 megahertz by Renzo Sancisi of the Kapteyn Laboratory in the Netherlands and his colleagues with the Westerbork Synthesis Radio Telescope. The radio intensity decreases with distance from the central plane of the galaxy; the contours extend 30,000 light-years from the plane. Observations made by Adrian Webster of the University of Cambridge indicate that radio emissions may originate in a region extending 30,000 light-years from the plane of our galaxy. The region has about the same extent as that of the gas detected at ultraviolet wavelengths.

the central plane of the disk. The rotation is such that matter more than 10,000 light-years from the center of the galaxy moves in orbit with a velocity of about 220 kilometers per second. Given the diameter of the galactic disk, gas at the periphery of the disk may have revolved about the galactic center only some 25 times since the origin of the galaxy some 10 billion years ago.

Similar principles have been employed in the investigation of components of the disk gas other than hydrogen. When a telescope is aimed at a distant star, the radiation emitted by the star must pass through the gas in order to reach the telescope. The emission from the star generally consists of radiation of many different wavelengths, and as the radiation passes through the gas specific wavelengths will be absorbed. As we have seen, the wavelengths that are absorbed are those with precisely the quantity of energy needed to raise the electrons of the atoms of the intervening gas from one state to another. In the spectrograph the absorbed wavelengths are recorded as dark lines, representing sharp decreases in the radiation in the spectrum of wavelengths.

Absorption lines can yield information not only about the composition of the interstellar gas but also, by taking into account the Doppler effect, about its velocity. If the gas is moving away from the observer, the absorption occurs at a slightly longer wavelength. This "red shift" is directly proportional to the velocity of the gas along the line of sight, and so the velocity can be readily calculated.

The pattern of the absorption lines indicates that the interstellar gas includes many elements besides hydrogen. Some elements are found primarily in the solid particles of interstellar dust. The relative proportions of the elements in the gas and the dust combined are probably similar to those in the atmosphere of the sun. (The sun's atmosphere is mostly hydrogen, but it includes more than 50 other elements.)

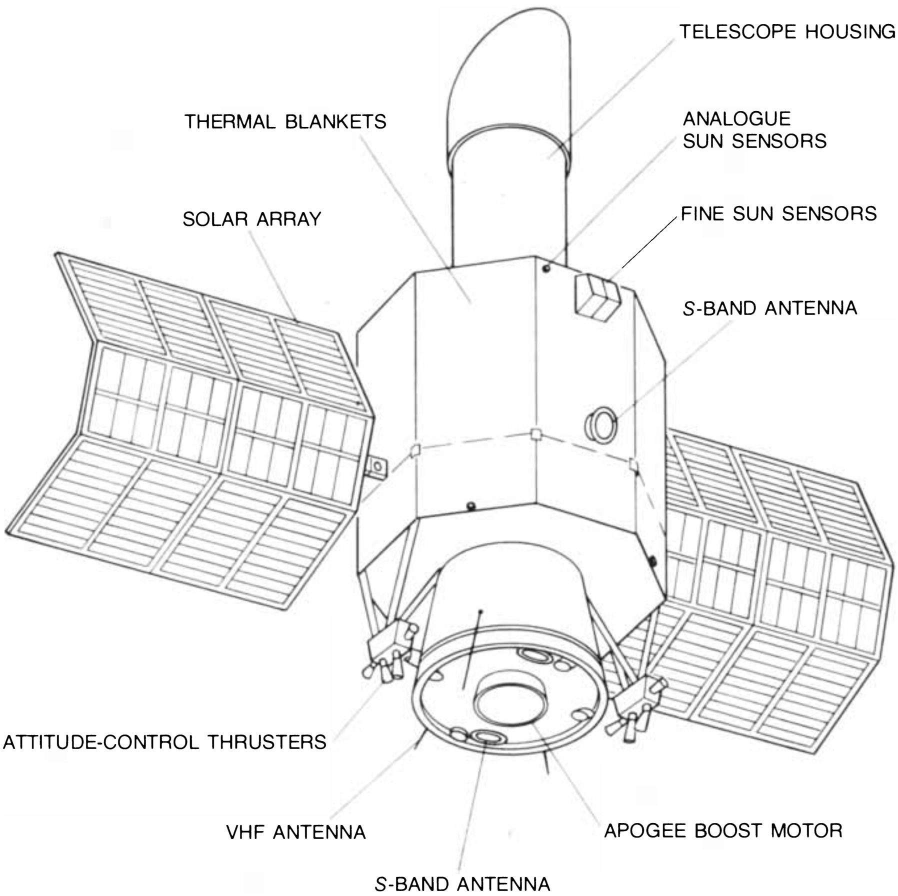
Each atom in the interstellar gas can assume several different states of ionization. An ion is an atom that has more or fewer electrons than the atom in the neutral (un-ionized) state. Most of the ions in the interstellar gas have had electrons removed and are therefore positively charged. Electrons can be removed from an atom by the same processes that raise them from one energy state to another, namely collisions with other atoms and exposure to electromagnetic radiation.

The degree of ionization depends on how many electrons have been removed. For example, if a single electron is removed from a neutral interstellar carbon atom, the atom acquires one unit of positive charge and is said to be once ionized. If another electron is removed, the atom, with two units of positive charge, is twice ionized. In atomic spectroscopy the neutral carbon atom is designated C I, the once-ionized atom C II and the twice-ionized atom C III. Much of the hydrogen in the interstellar gas is H I; the large clouds of neutral hydrogen between stars are known as H I regions. Near a hot star, however, energetic photons (quanta of radiation) remove the electron from the hydrogen atom; the result is the formation of an H II region.

The elements in the interstellar gas that are heavier than hydrogen are also influenced by radiation from nearby stars. For each of these atoms a particular quantity of energy is required to remove an electron; the quantity is referred to as the ionization energy. The differences in ionization energy result in varying degrees of ionization of the heavier elements under the conditions in the interstellar gas.

As in the neutral atom, each electron in an ion moves only in strictly specified ways. The wavelength of the radiation absorbed in the electron transitions can yield considerable information about the ions. Work done in the 1950's by Guido Münch of the California Institute of Technology with the 200-inch Hale telescope on Palomar Mountain yielded significant results on the interstellar gas. Münch directed the telescope at stars far above the galactic plane. Examination of the resulting spectra indicated absorption by Na (sodium) I and Ca (calcium) II. The absorption lines were quite complex; they indicated the presence of gas clouds with high velocities.

The high-velocity clouds observed by Münch have since been studied intensively by astronomers working with the 21-centimeter emission line of hydrogen. The existence of gas clouds at very large distances from the galactic plane raises several problems. The first problem is related to the gravitational forces exerted by the stars of the galactic plane. If a cloud of gas is to rise above the plane, it must overcome these forces. The fact that the high-velocity clouds had risen so far above the galactic plane indicated that they must have had a very



INTERNATIONAL ULTRAVIOLET EXPLORER (IUE), an earth satellite launched in 1978, was designed for astronomical observations at ultraviolet wavelengths. The satellite is now in geosynchronous orbit over South America, that is, it stays in position with respect to the continent. It is a joint project of the National Aeronautics and Space Administration (NASA), the European Space Agency and the British Science Research Council. The *IUE* carries a telescope and spectrograph. With the telescope pointed at a distant star the spectrograph records the intensity of ultraviolet radiation, including specific wavelengths absorbed as the radiation from the star passes through the intervening gas. Such data can indicate the gas's composition, temperature and velocity. The solar panels provide the power for the satellite's instruments. The sun sensors make it possible to adjust the satellite's attitude with respect to the sun. VHF antenna receives commands from ground stations; *S*-band antenna relays data to the ground.

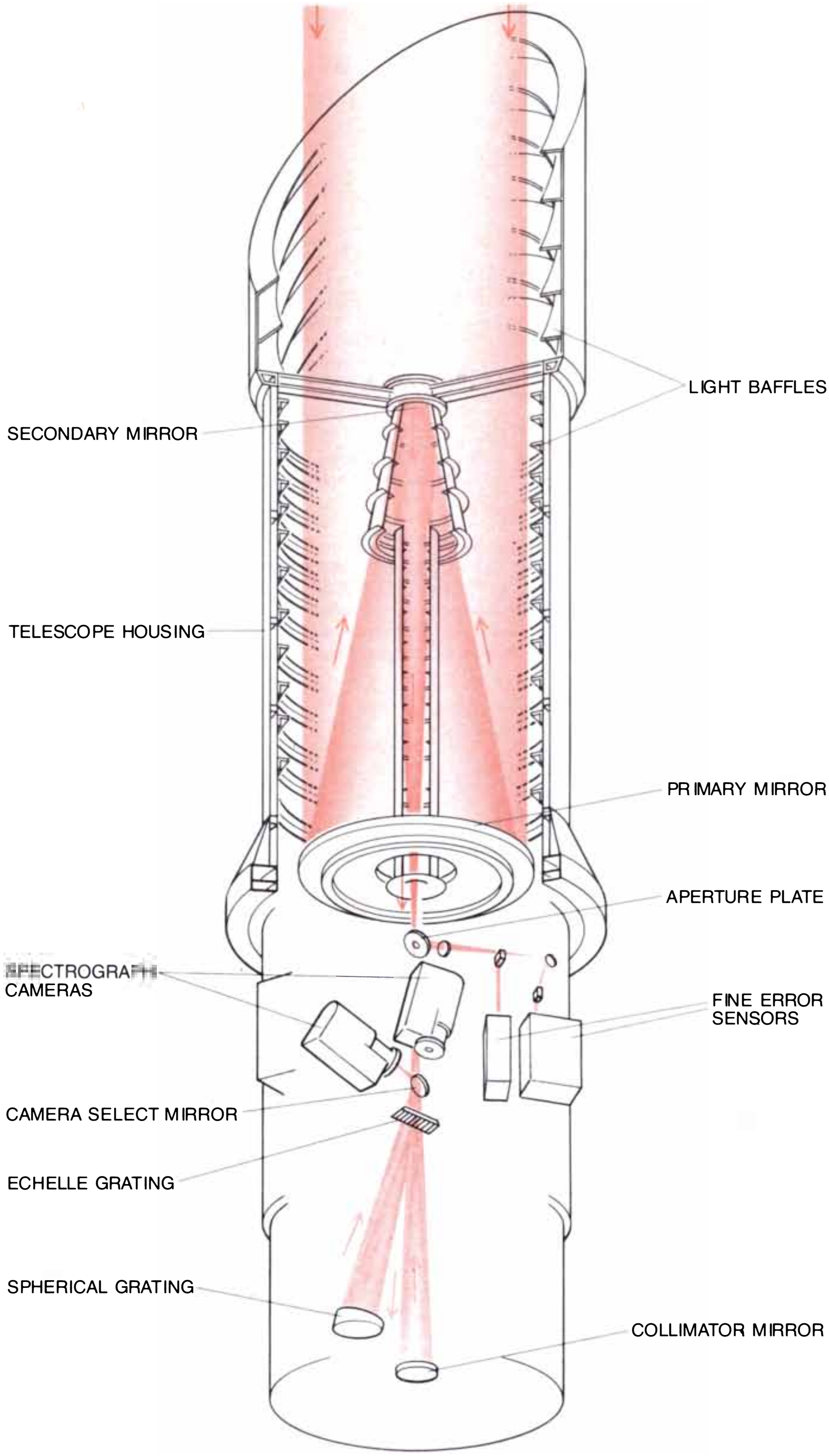
high velocity when they left the central disk of the galaxy.

Another problem was somewhat more perplexing. If the space around a high-velocity cloud were empty, the cloud would expand, disperse and disappear. The continued existence of the clouds implies that some force holds them together. In interstellar space the likeliest candidate for such a force is the pressure exerted by gas surrounding the high-velocity clouds. In 1956 Lyman Spitzer, Jr., of Princeton University proposed that the galaxy is surrounded by a corona of hot gas, and that it is the coronal gas that holds the high-velocity clouds together.

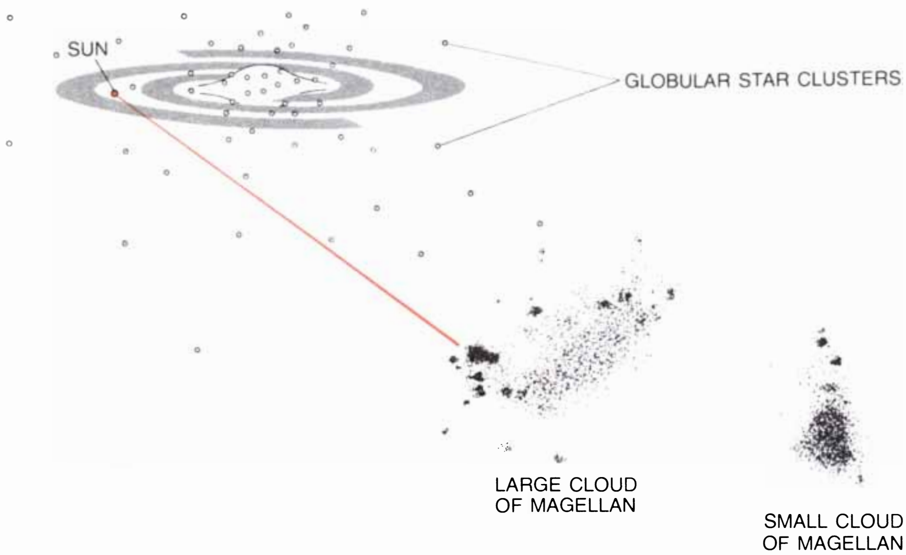
Spitzer had inferred the existence of the hot coronal gas from evidence gathered by Münch and others on the high-velocity clouds. In discussing the possibility of proving the existence of the coronal gas in a more direct way Spitzer noted that the presence of the coronal gas would probably be indicated only by the absorption of radiation with a wavelength of 2,000 angstrom units or less, which is in the ultraviolet range.

At the temperature predicted for the gas in the hottest part of the corona most atoms would have lost many electrons; the high temperature of the gas increases the kinetic energy of the colliding atoms and therefore the probability that one of the atoms will lose an electron. The highly ionized atoms include C IV, Si (silicon) IV, N (nitrogen) V and O (oxygen) VI. These ions all show absorption lines at wavelengths shorter than 2,000 angstroms. The absorption lines of the abundant atoms carbon, nitrogen, oxygen, magnesium, aluminum, sulfur, silicon, iron, nickel and zinc in their neutral or once-ionized states are also in the ultraviolet.

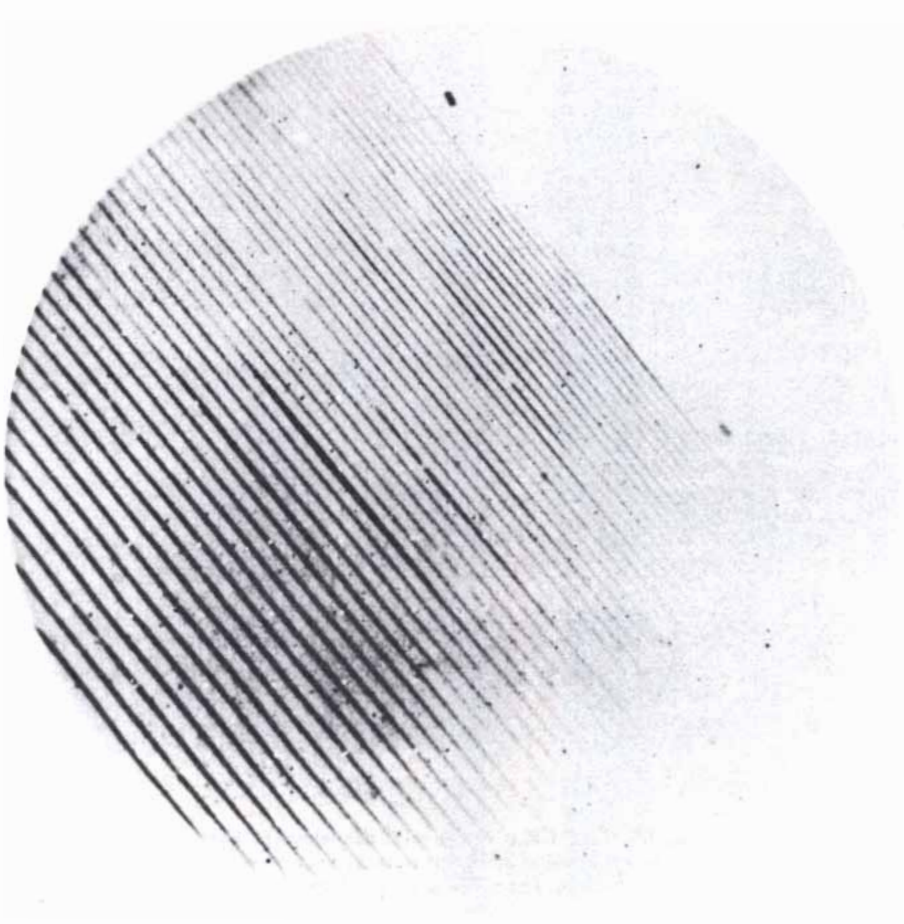
At the time Spitzer published his classic paper it was impossible to verify the existence of the corona; the necessary observations would have had to be made from the ground—from the bottom of the earth's ultraviolet-absorbing atmosphere. In order to gather more direct experimental evidence for the existence of the galactic corona an observational platform outside the atmosphere was needed. Earth satellites have supplied such a platform. The first satellites with instruments capable of detecting absorption in the ultraviolet part of the spectrum were launched by the National Aeronautics and Space Administration (NASA) in the early 1970's; they were known as the Orbiting Astronomical Observatories (OAO). The second successfully launched observatory in the series, *Copernicus,* was in operation from 1972 until the end of the OAO program in 1980. *Copernicus* carried experimental apparatus designed specifically for the study of lines in the ultraviolet absorbed by interstellar gas. The *Copernicus* project was directed by Spit-



INSTRUMENTATION OF THE IUE is shown with the telescope housing cut away. With the telescope pointed at a star the baffles inside the housing prevent light from the sun and the earth from entering the spectrograph. The primary mirror of the telescope is 45 centimeters in diameter. The radiation that strikes it is reflected to a secondary mirror near the top of the housing and back through an aperture in the primary mirror into the spectrograph compartment. Mounted behind the telescope are two spectrographic systems. One covers the wavelengths from 1,150 to 2,000 angstrom units; the other covers those from 1,900 to 3,200 angstroms. The two systems are quite similar; only the short-wavelength system is shown. The aperture plate behind the main mirror directs part of the beam to the fine error sensors, which keep the telescope directed at the target. Most of the beam passes through the aperture plate to the collimator mirror. The collimated beam is reflected onto the echelle grating, which breaks it down into its constituent wavelengths. The spherical grating separates the beam into spectral orders and directs it into one of two television cameras that are sensitive to ultraviolet radiation. The cameras make a television image of the spectrum that has been yielded by the echelle spectrometer. The television image is converted into digital form and relayed to the ground.



CLOUDS OF MAGELLAN are two small galaxies near our own. The Large Cloud of Magellan is about 160,000 light-years from the sun, the Small Cloud about 200,000 light-years. Hot stars in the two galaxies were exploited as a source of radiation for the study of the galactic corona with the *IUE.* The absorption indicates the presence of gas at temperatures between 10,000 and 200,000 degrees Kelvin. Other data suggest there is also gas at a million degrees K.



ECHELLE SPECTROGRAM made by the instruments of the *IUE* is a television image of part of the spectrum. The spectrum shown is for the star HD 38282 in the Large Cloud of Magellan. It includes radiation with wavelengths from 1,200 angstroms (*upper right*) to 2,000 angstroms (*lower left*). The gratings of the spectrograph break the radiation down into orders about 15 angstroms wide. Each dark diagonal line in this positive print corresponds to one order. The white space in certain orders indicates that radiation of a particular wavelength has been absorbed by the gas of the corona. The dark spots were made by cosmic-ray particles. The white dots are a grid for correcting the geometrical distortions in the television image.

zer, who also supervised the design of its instruments.

Of the many findings made with *Copernicus* one of the most important was the observation of O VI absorption in the interstellar medium near stars in the vicinity of the sun; the work was done by Edward B. Jenkins and his colleagues at Princeton University. The presence of O VI indicates that this interstellar gas has an extremely high temperature. William L. Kraushaar and his co-workers at the University of Wisconsin at Madison also concluded that components of the interstellar medium near the sun were extremely hot; their results were obtained by studying the emission of X rays.

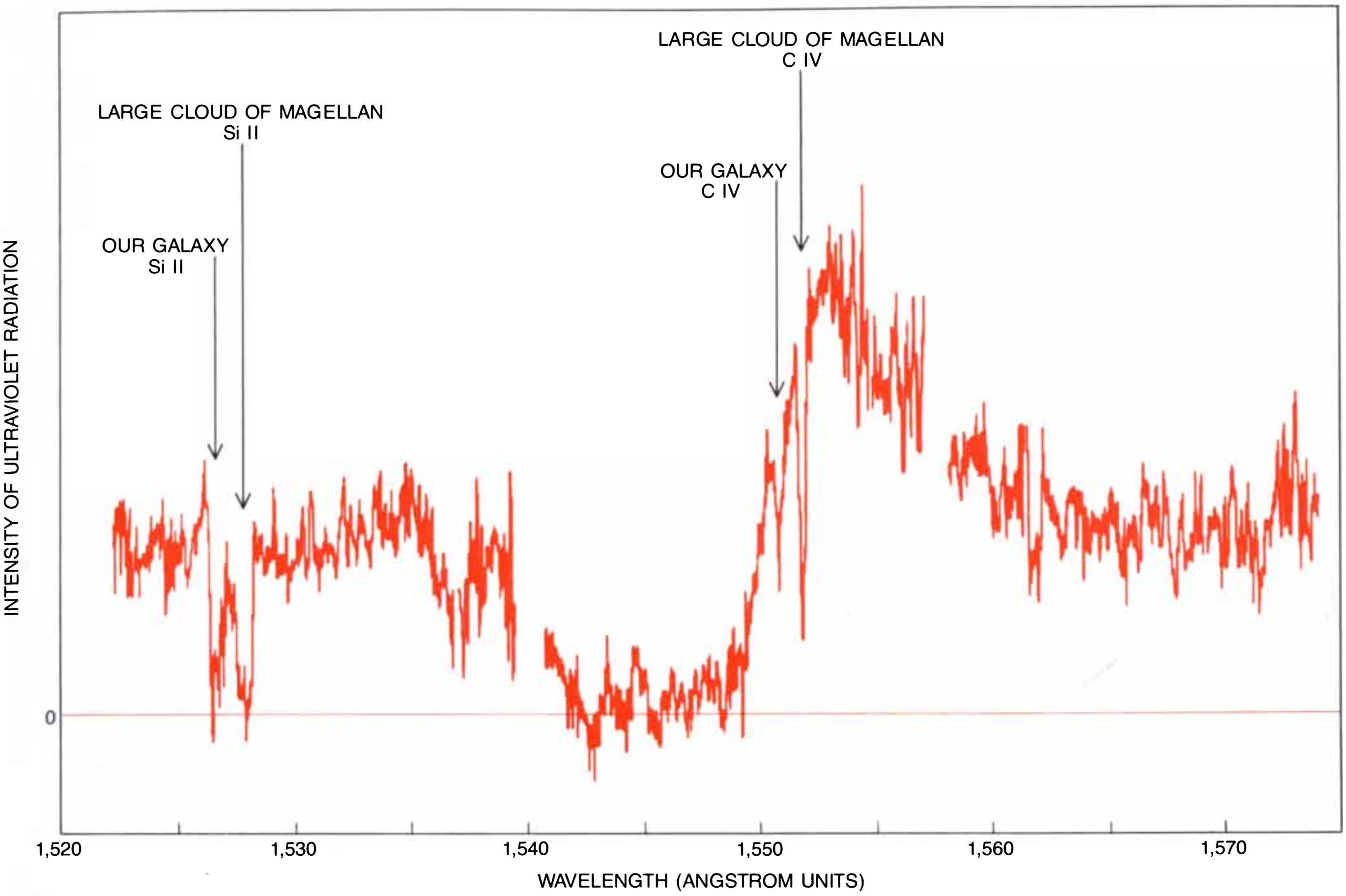
The hot components of the interstellar medium, which were discovered in the 1970's, have been the subject of much recent experimental and theoretical work. The best current hypothesis is that the gas is heated by shock waves from the explosion of supernovas. In a spiral galaxy these dramatic explosions come about once every 100 years.

The later work tended to support Spitzer's hypothesis that the high-velocity clouds are held together by pressure exerted by surrounding clouds of hot gas. *Copernicus,* however, did not have instruments sufficiently sensitive to measure the absorption of radiation from distant stars far above the plane of the galaxy. Measurements of this kind are needed to detect the galactic corona. It was not until the launching of the *International Ultraviolet Explorer* in January, 1978, that the corona could be actively studied.

The *IUE* is operated jointly by NASA, the European Space Agency and the British Science Research Council. It has served for observations of many phenomena other than the galactic corona. It is a geosynchronous satellite: its period of rotation about the earth is synchronized with the period of the earth's rotation so that the satellite remains stationary above South America, about 40,000 kilometers from control stations in Europe and the U.S. The satellite carries a telescope and a spectrometer. The beam of ultraviolet radiation from the spectrometer is directed toward television cameras sensitive to ultraviolet radiation; the cameras create an image of the spectrum that is subsequently relayed to the ground.

The resolving power of the spectrometer can be altered by the observer on the ground. In one mode details of the spectrum seven angstroms wide can be detected; in the other mode details .1 angstrom wide can be detected. The high-resolution spectrum is essential for the study of absorption by the galactic corona.

To examine regions of the corona far



ABSORPTION BY THE CORONA of our galaxy and the corona of the Large Cloud of Magellan is shown in a plot of the intensity of radiation detected by the *IUE* against the wavelength. The plot is for three orders of the spectrogram of the star HD 38282 at the bottom of the opposite page; the orders are separated by spaces. Absorption by the coronal gas appears on the plot as deep, narrow troughs. Each ion in the gas absorbs radiation of a particular wavelength. Two ions absorb radiation with wavelengths within the orders shown: the silicon atom with one electron removed (Si II) at about 1,527 angstroms and the carbon atom with three electrons removed (C IV) at about 1,550 angstroms. The absorption is due not only to coronal gas but also to gas in the central plane of the two galaxies. Because of the Doppler effect the wavelength of the absorption by the corona of the Large Cloud is longer than that of the absorption by the corona of our galaxy. Large trough and peak in middle order were made by matter flowing outward from the star; they constitute the P Cygni profile.

from the plane of the galaxy we directed the telescope of the *IUE* toward a star in the Large Cloud of Magellan. The Large Cloud and its close neighbor, the Small Cloud of Magellan, are small galaxies near our own: about 160,000 light-years from the sun. At that distance the stars of the Clouds of Magellan are so faint that an exposure of six hours was required to make a high-resolution spectrogram.

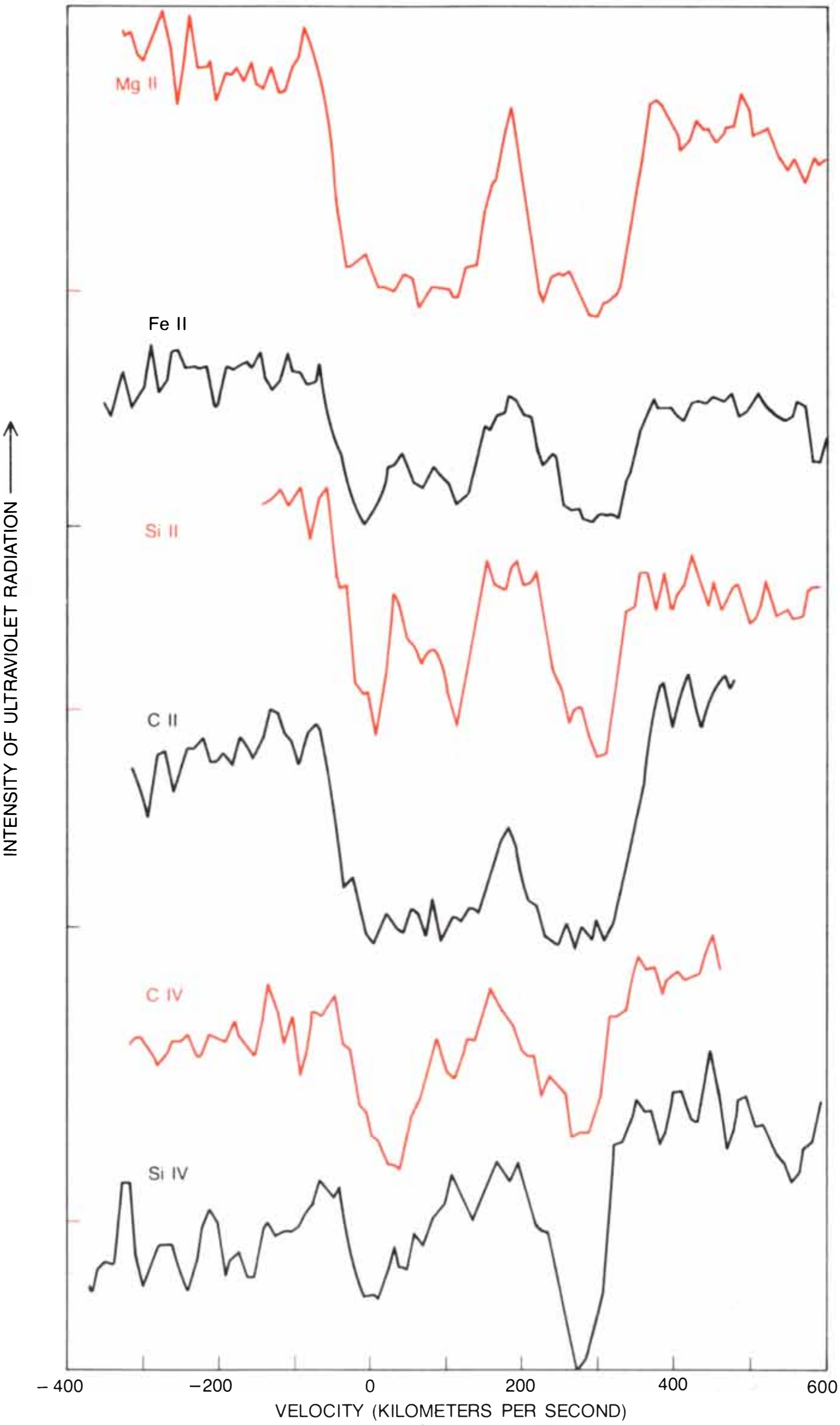
The Large Cloud of Magellan has an apparent velocity of 270 kilometers per second away from the sun. Much of the apparent velocity is actually due to the motion of the sun around the center of our own galaxy. When we examined absorption spectra made with the *IUE,* we found a complex pattern giving evidence of absorption by gas clouds with quite different velocities. We found evidence of absorption at a velocity of zero with respect to the sun and also at a velocity of 270 kilometers per second. The absorption lines were clearly due to gas near the sun and to gas in the Large Cloud of Magellan.

Surprisingly, we also observed pronounced absorption by gas with an intermediate velocity. The wavelengths of the absorption implied that the ions responsible were C II, Mg (magnesium) II and Si II; such ions are formed at moderate temperatures. Even more intriguing, however, was the finding of strong absorption by C IV and Si IV. Such ions are formed at a temperature of about 100,000 degrees K., so that their presence gives confirmation for the idea of a hot galactic corona.

In order to gain a more detailed understanding of the absorption lines the television image of the ultraviolet spectrum is converted into a plot of the intensity of radiation at each wavelength. By applying formulas related to the Doppler effect the velocity of the gas responsible for the absorption can readily be computed. If certain additional assumptions are made, the distance of a cloud can be estimated by measuring its velocity. Such a procedure makes it possible to determine the position of hot and cool clouds in the galactic corona.

Since the gas of the galactic disk rotates with the galaxy at 220 kilometers per second, that is the maximum apparent velocity with respect to the sun a cloud of interstellar gas in our galaxy can have. The motion of most points in the galaxy, however, is not directed precisely at the sun. The component of the velocity directed at the sun varies according to the distance from the sun. The sun is itself moving around the center of the galaxy; the apparent velocity of the object is therefore the difference between the velocity of the sun and the velocity of the object along the line of sight. The relation between velocity and distance makes it possible to calculate one quantity from the other.

Along the line of sight to the star designated HD 36402 in the Large Cloud of Magellan the maximum velocity an object in our galaxy can have is 175 kilometers per second away from the sun. When we pointed the telescope of the *International Ultraviolet Explorer* at HD 36402, we observed strong C II absorp-



VELOCITY OF GAS in the galactic corona can be calculated from the wavelength at which absorption takes place by means of Doppler-effect formulas. The intensity of the radiation detected can then be plotted against the velocity of the absorbing gas. This illustration shows such a plot for six ions in the coronal gas; the background source was the star HD 36402 in the Large Cloud of Magellan. The short lines on the vertical axis indicate the zero level of intensity. The Large Cloud has an apparent velocity of 270 kilometers per second away from the sun, much of it due to the sun's rotation about the center of our galaxy. For each ion strong absorption is seen by stationary gas and by gas with a velocity of 270 kilometers per second; this corresponds to gas near the sun and gas in the Large Cloud. Absorption is also seen at intermediate velocities. If assumptions are made about how the coronal gas moves in our galaxy, the position of the gas can be calculated from its velocity. Iron and silicon with one electron removed (Fe II and Si II) are formed in cool clouds. Absorption by these ions at 70 and 120 kilometers per second indicates the presence of cool clouds at intermediate distances from the central plane of our galaxy. Carbon and silicon with three electrons removed (C IV and Si IV) are formed in hot gas. Absorption by these ions at as much as 175 kilometers per second indicates gas with a temperature of about 100,000 degrees K. at large distances from the galactic plane.

tion at that velocity. We also observed absorption at higher velocities corresponding to gas near the Large Cloud of Magellan. In addition we found evidence of the highly ionized atoms C IV and Si IV at distances perhaps as large as 30,000 light-years. The absorption profile of C IV and Si IV suggests that their density decreases in direct proportion to distance; the density appears to be reduced by a factor of three over 10,000 light-years.

Our results suggest that there are clouds of ionized gas a long way from the galactic plane. The clouds are so far from the plane, in fact, that their position raises a theoretical problem. The gas of the galactic disk turns at 220 kilometers per second because of the gravitational forces exerted by the stars. The clouds observed with the *IUE* are so far from the galactic plane that one can ask whether the force exerted by the stars is sufficient to keep the gas in the distant clouds moving around the galactic axis of rotation at the same speed as the gas in the galactic disk. If the distant gas does not rotate at the same speed as the disk gas, the calculation of distances from velocities would be invalid.

Other work done with the *IUE*, however, has tended to confirm our estimate of the extent of the coronal gas. G. E. Bromage, A. H. Gabriel and Dennis W. Sciama of the University of Oxford and more recently Max Pettini of the Royal Greenwich Observatory and Kim West of University College London have studied C IV absorption along the line of sight to stars in the galactic halo that are as much as 10,000 light-years away. They found strong C IV absorption in the direction of all stars more than 3,000 light-years from the galactic plane.

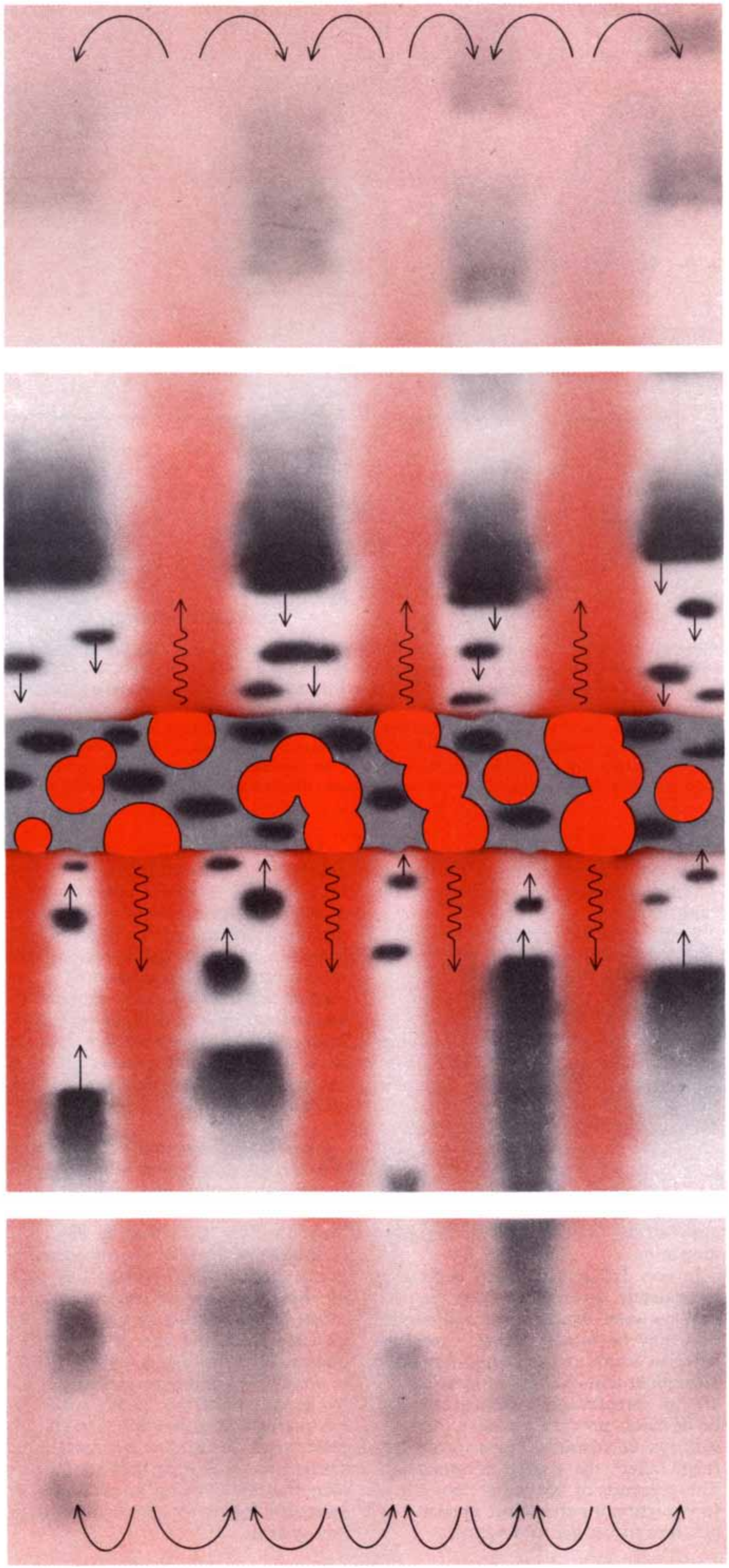
The cooler gas of the galactic corona also seems to be widely distributed. With the *IUE* the cool gas is best detected by means of C II absorption. With telescopes on the ground Ca II absorption is the best indicator. Donald G. York and his colleagues at Princeton and Chris Blades of the European Space Agency station of the *IUE*, working with investigators using the Anglo-Australian telescope in New South Wales, have observed Ca II absorption at very large distances from the galactic plane, thereby extending the pioneering work of Münch. Numerous cool clouds, in many different directions from the center of the galaxy, thus seem to be embedded in the hot coronal gas.

Detailed examination of absorption lines at the velocity of the Clouds of Magellan indicates that these two small galaxies also may be surrounded by envelopes of coronal gas much like that of our own galaxy. Our work suggests that all spiral and irregular galaxies have coronas. Further investigation will certainly be needed to confirm this hypothesis. Meanwhile much detail is being added to the general picture we already have of the gas in our galaxy.

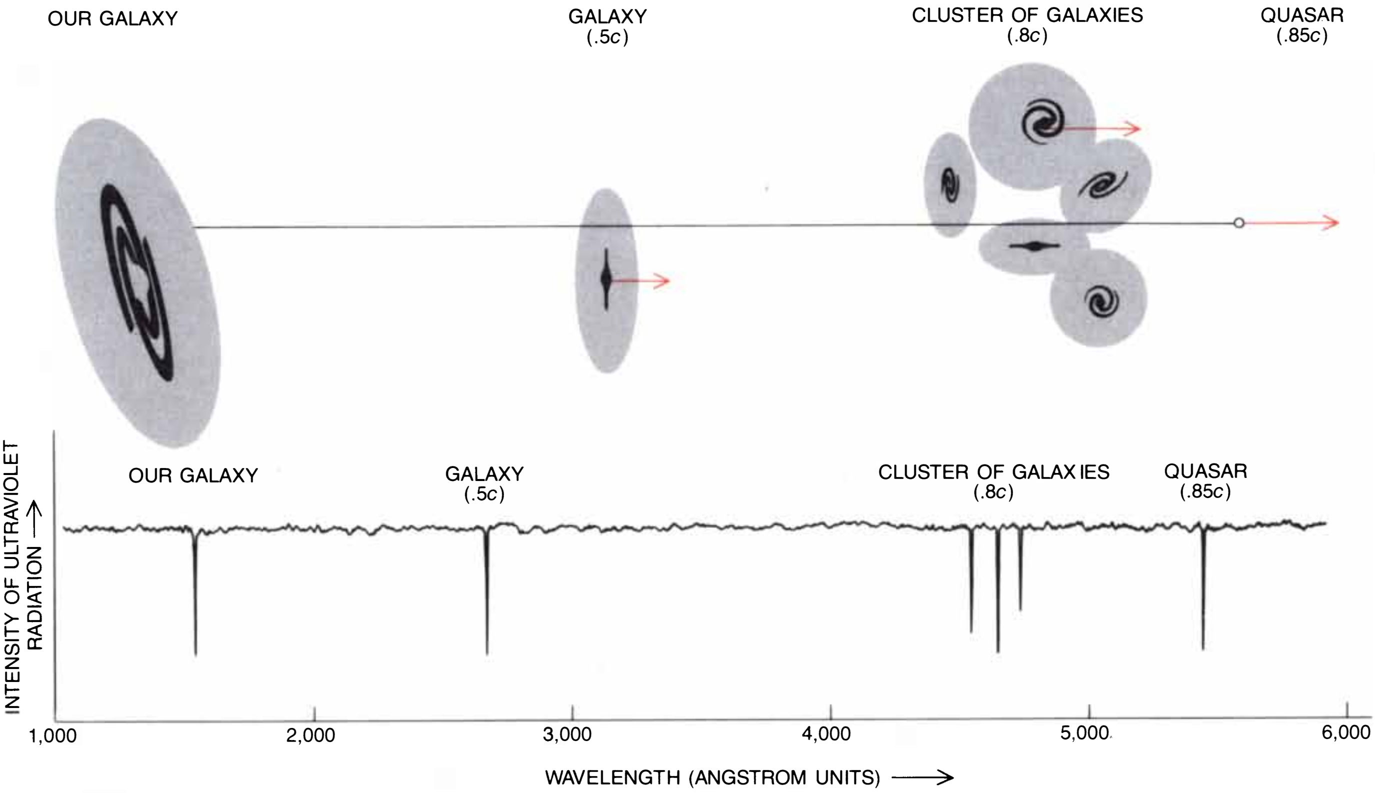
That picture has changed considerably over the past 25 years. Until about 10 years ago the interstellar medium was thought to consist of two main components: cool, neutral gas between the stars of the spiral arms and warm, ionized gas between the spiral arms. The detection of O VI absorption and X-ray emission suggested that on the contrary much of interstellar space is filled with highly ionized gas having a temperature of a million degrees K. Gas at such a temperature cools quite slowly; once it has been heated by the explosion of a supernova it remains hot. A subsequent explosion of a supernova would generate shock waves that could travel easily through the hot gas and keep it ionized.

Paul Shapiro and George B. Field of the Center for Astrophysics of the Harvard College Observatory and the Smithsonian Astrophysical Observatory have argued that the hot gas would not be confined to the galactic disk: it would tend to burst out of the disk and flow away from the central plane. The distance it traveled from the disk would depend on its temperature and on the gravitational forces exerted by the stars of the central plane. As the gas flowed outward it would gradually cool and become denser. Denser clouds would begin to descend; ultimately they might be observed as the cool regions of the corona. It is Shapiro and Field who have given the cycle the name galactic fountain.

Other workers have contributed to the effort to understand the processes in the galactic corona. Joel N. Bregman of New York University has extended Spitzer's work by constructing a model for the "percolating" motion of the gas in the fountain. Donald P. Cox of the University of Wisconsin at Madison has analyzed the large-scale structure of the interstellar medium and its energy balance. Results obtained by both workers support the hypothesis that the driving force of the fountain is the explosion of supernovas. If the hypothesis is correct,



"GALACTIC FOUNTAIN" is shown in a highly schematic cross section of the corona and the galactic disk. The dense gas of the disk is about 500 light-years thick. Hot disk gas rises to a height equal to about 50 times the thickness of the disk. Cooler gas in the form of clouds descends toward the disk to complete the fountain. The explosions of supernovas heat some of the disk gas to a temperature of a million degrees K.; such explosions take place in our galaxy roughly once every 100 years. The heated gas bursts out of the disk and moves away from the central plane. As the gas rises it begins to cool and condense, and then it descends toward the disk. Ultimately it is observed in the form of the clouds of cool coronal gas observed with the *IUE*. When the clouds reach the galactic plane, they become part of the cool gas in the disk and the cycle can begin again.



SPECTRUM OF A QUASAR (*shown schematically at bottom*) is strongly shifted toward the red, suggesting that the object is moving away from our galaxy at high velocity and therefore, in accord with the general expansion of the universe, is at a great distance. Superposed on such spectra, however, are groups of absorption lines that show smaller red shifts. These anomalous red shifts may be explained by the absorption of radiation from the quasar by the coronas of galaxies between the quasar and our galaxy (*schematic diagram at top*). In the spectrum at the bottom are six absorption lines made by the ion C IV. Stationary C IV absorbs radiation at 1,550 angstroms; absorption at that wavelength is due to gas in the corona of our galaxy. C IV absorption near 3,000 angstroms and between 4,000 and 5,000 angstroms corresponds to objects with respective velocities of .5 and .8 the velocity of light (*c*). Absorption at these wavelengths may be due to intervening galaxies. Absorption near the quasar appears at about 5,400 angstroms, corresponding to a velocity .85 that of light.

the vigor of the galactic fountain and the mixing of hot and cool gas in the disk would depend on the frequency of the explosions.

The gas with a temperature of 100,000 degrees K. detected by the *International Ultraviolet Explorer* is far from the hottest gas in the corona. It is difficult to determine whether the gas detected by means of C IV and Si IV absorption is ascending or descending. It is even difficult to tell whether the temperature ascribed to the gas is correct. In dense gases, such as those on the earth, raising the temperature of the gas adds kinetic energy that is quickly spread evenly throughout the atoms by collisions. As a result the temperature of the gas is closely connected to the degree of ionization.

In very diffuse gases such as those of the galactic corona, however, the relation between temperature and ionization may be much weaker. Collisions between atoms and electrons may be so infrequent that changes in temperature are not spread rapidly through the gas. In addition processes such as the absorption of cosmic X rays by the gas might affect the degree of ionization. The presence of particular ions therefore might not indicate the precise temperature of the gas as a whole.

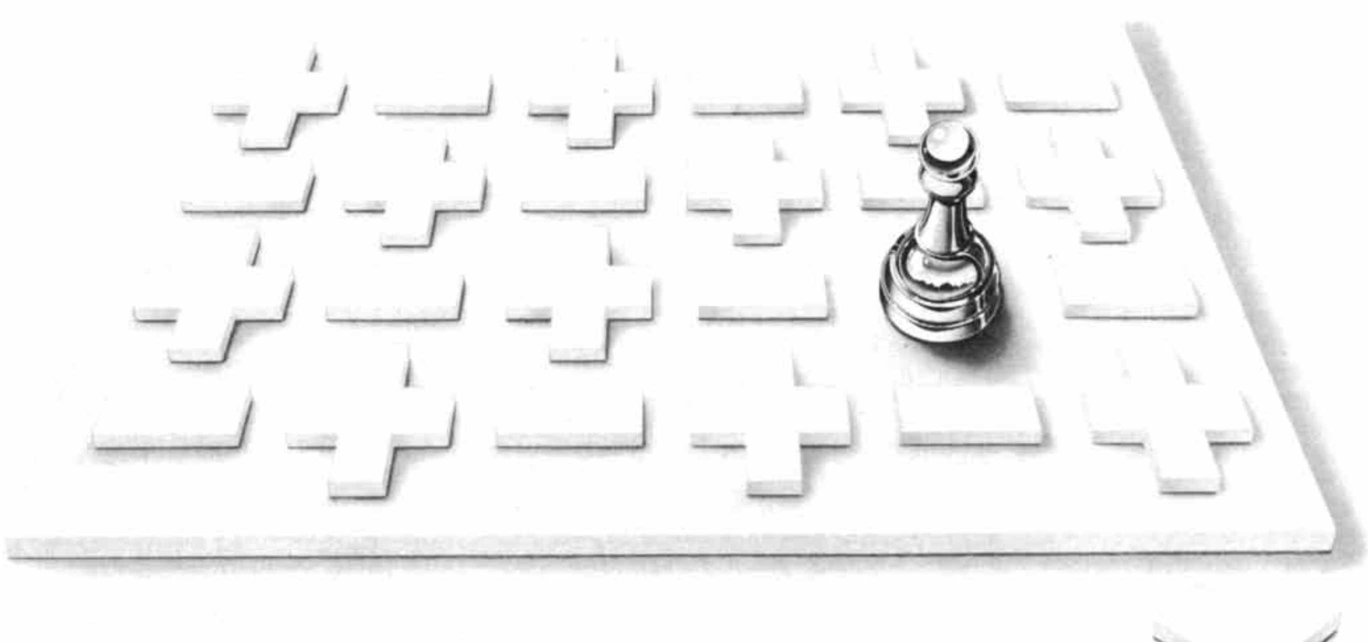
The discovery of the corona of our galaxy has important implications for a long-standing problem concerning the quasars. The emissions of quasars in the ultraviolet region of the spectrum are red-shifted into the visible region. Most workers interested in quasars have contended that the red shifts indicate the quasars are billions of light-years away in an expanding universe. Some workers have suggested that quasars are nearby objects in which peculiar phenomena are taking place.

A careful examination of quasar spectra has shown that in many of the spectra absorption lines can be put in groups that have red shifts smaller than the red shift of the quasar. In 1969 John N. Bahcall of the Institute for Advanced Study and Spitzer suggested that some of these systems of absorption lines might be introduced by the coronas of galaxies along the line of sight to the quasar. With the results from the *IUE* it is possible to compare absorption due to the corona of our galaxy with typical quasar absorption-line systems. The patterns of absorption are remarkably similar; the *IUE* data thus support the idea that intervening coronas are responsible for many quasar absorption-line systems.

If the coronas are indeed responsible for the observed absorption, it follows that quasars are at enormous distances and that they represent the most luminous objects known. And if the absorption-line systems commonly seen in the spectra of quasars are due to absorption by the coronas of galaxies, the coronas must be considerably larger than the images of galaxies seen in photographs.

The coronas required to yield the observed effects would need to be from 100,000 to 300,000 light-years in radius, or larger by a factor of three than the dimensions of galaxies in photographs. It has already been recognized, however, that galaxies are much larger when they are observed at the 21-centimeter emission line of neutral hydrogen than when they are observed at visible wavelengths. They would undoubtedly appear larger still if they were probed by means of absorption at the wavelengths of the highly sensitive ultraviolet absorption lines. It should soon be possible to expand greatly on these preliminary studies of galactic coronas. With the launching into orbit of the 2.4-meter Space Telescope in 1985 much additional information from the ultraviolet part of the spectrum on the corona of our galaxy and the coronas of other galaxies should be forthcoming.

Playing ionic chess to make silver sensitive.



Silver halides are the key to the photographic process.

Right.

Because silver halides respond strongly to light.

Wrong.

Pure and structurally perfect micro crystals of silver halide are marginally photosensitive. Unless and until "foreign material" is introduced, nothing happens.

The nature and placement of these impurities within the lattice structure of the silver halide crystal dictate such characteristics as speed and sharpness, contrast and grain.

For many years, we have been doing "impure" research into this critical doping of silver halide crystals. Never content with today's performance, we are constantly investigating new combinations, new arrangements, new triggers. Playing ionic chess on the lattice structure.

The procedure is complex simplicity. We begin by growing two Bridgman macro crystals, about 6″ long and ⅝″ across. One is pure silver halide, for control. The other contains an added impurity or group of impurities: transition metal ions such as iron, nickel, and cobalt.

Translucent sections, 4 mm x 1 cm, are cut from the crystals and exposed to visible and infrared light. During these separate exposures, electron spin resonance is used to check the light sensitivity of the ion structure being tested. The results are then checked against the pure control crystal.

Since the effect of different impurities is not additive, each possible permutation must be investigated individually. And since there is a large difference in the size of macro and production crystals, much sophisticated interpretation must precede the final extrapolation.

The result of this ongoing search is ongoing improvement in the ability of our photosensitive emulsions to record as fast and as faithfully as tomorrow's technological needs may demand.

For further information on this exclusive research by Kodak, send for reprint P82-006, "Pb^{2+} Influence on the Photochemistry of AgCl." Write: Eastman Kodak Company, Dept. GB-SA-4, 343 State Street, Rochester, NY 14650.

Kodak

Kodak. Where technology anticipates need.

SCIENCE AND THE CITIZEN

Saving Europe

In the course of the war in Vietnam an American military officer was reported to have said that in order to "save" a Vietnamese village it was necessary to destroy it. Much the same attitude is implicit in current plans on both sides of the East-West confrontation in central Europe to resort to nuclear weapons to defend European cities and towns against the threat of attack by the other side.

Recent official statements indicate that the military forces of the North Atlantic Treaty Organization have deployed more than 6,000 nuclear weapons in Europe for possible use in a "limited" nuclear war. The opposing forces of the Warsaw Pact are thought to have an equally formidable arsenal of nuclear weapons aimed at targets in Europe. Included in these tallies are both "theater" nuclear weapons deployed on intermediate-range aircraft and missiles and tactical, or battlefield, nuclear weapons such as short-range missiles, artillery shells and land mines. The bilateral nuclear-arms-control talks now under way in Geneva are directed toward reducing the number of theater nuclear weapons capable of reaching targets in Europe and reducing the number of strategic, or intercontinental, nuclear weapons targeted by the two superpowers at each other. Excluded from the negotiations are the large numbers of tactical nuclear weapons based in Europe.

A disturbing aspect of the shorter-range nuclear weapons is that many of them are deployed in forward positions in central Europe, particularly in East Germany and West Germany. Because the weapons would be close to the front lines in any likely war they would stand a good chance of being overrun by invading forces. Given a choice between using the weapons or losing them, the defending forces would presumably be under pressure to fire them first, thus escalating the conflict past the "nuclear threshold." Furthermore, the comparatively small size of some of the tactical nuclear weapons could have the psychological effect of lowering the threshold.

Both of these issues were addressed recently by the Independent Commission on Disarmament and Security Issues, an international group of present and former government officials formed two years ago under the chairmanship of Olof Palme, the former prime minister of Sweden. (The group, which was made up of members from Eastern, Western and Third World nations, included Cyrus Vance, former secretary of state of the U.S., and Giorgi Arbatov, a representative of the government and the Academy of Sciences of the U.S.S.R.) In a report published in June the commission called for "the establishment of a battlefield-nuclear-weapon-free zone in Europe" together with other "measures to strengthen the nuclear threshold and reduce pressures for the early use of nuclear weapons."

According to the report, *Common Security: A Blueprint for Survival,* the nuclear-weapons-free zone would start with central Europe, perhaps as a strip 300 kilometers wide centered on the border of West Germany with East Germany and Czechoslovakia. In time the zone would be extended, "ultimately from the northern to the southern flanks of the two alliances." No nuclear weapons of any kind would be permitted in the zone. Measures to ensure compliance, the members of the commission agreed, "would have to include a limited number of on-site inspections in the zone on a challenge basis."

In addition, the report states, "it is important to maintain a clear distinction between nuclear and conventional weapons. We urge the nuclear-weapon states to abstain from deploying weapons which blur the distinction by appearing to be more 'usable.'" The section of the report devoted to the nuclear threat in Europe also urges the U.S. and the U.S.S.R. to make "substantial reductions in all categories of intermediate-(medium-) and shorter-range nuclear weapons which threaten Europe."

The possible consequences of failing to remove nuclear arms from Europe are spelled out in another recent publication. Writing in a special issue of *Ambio,* a journal of the Royal Swedish Academy of Sciences, three American arms-control specialists, William Arkin, Frank von Hippel and Barbara G. Levi, calculate the effects of a "limited" nuclear war in the two Germanys. Their analysis shows that because of the density of the population in central Europe, the presence of military targets near civilian centers and the substantial "areas of death" of even the smallest battlefield nuclear weapons, "the purely military use of nuclear weapons in Europe on any scale large enough to achieve militarily significant results would result in the unintentional deaths of many millions of civilians."

The Education Gap

The education in science and mathematics that American students get in elementary school, junior high school and high school has declined in both quality and quantity in the past decade. The decline may have become severe enough to affect the capacity of American society to produce a competent labor force. A number of scientists, educators and members of the Reagan Administration agree on these points. The Administration, however, differs sharply from the other groups over how the problem should be solved. The disagreement was in evidence at the National Convocation on Precollege Education in Mathematics and Science held in Washington in May.

The data cited most frequently in the controversy over science education are students' scores on standardized tests. Paul D. Hurd of Stanford University reported to the convocation that between 1963 and 1980 the mean score on the mathematical part of the Scholastic Aptitude Test, which serves as a college-entrance examination, decreased from 502 to 466. The results of another national testing program show a steady decline between 1969 and 1977 in the general level of scientific knowledge among 17-year-olds. One factor contributing to the lower scores may be the minimal quantity of science and mathematics education in the country's 16,000 school districts. For example, only about a third of all students take three years of mathematics in high school.

Hurd went on to suggest that the future of scientific education may be bleaker than its present. Between 1971 and 1980 the number of candidates training to become mathematics teachers decreased by 77 percent; the number training to become science teachers decreased by 65 percent.

The current concern over American science education is reminiscent of that provoked by the launching of the first earth satellite by the U.S.S.R. in 1957. Much of the subsequent improvement in education was due to efforts by the National Science Foundation in collaboration with workers in the sciences and mathematics. The NSF established programs aimed at improving courses and textbooks and at reforming the training of teachers. The programs had considerable effect: in 1978 more than half of all high school science teachers were using materials developed by the NSF.

The Reagan Administration, however, is unsympathetic to a revival of the collaboration between the Federal Government and scientific workers. The Administration contends that responsibility for further improvement in public education lies with local government and private industry. Two years ago the NSF spent about $70 million on programs intended to improve science education. The agency asked for only $15 million in the current budget, all of it for fellowships for graduate students.

A fact often overlooked by both proponents and critics of Federal intervention is that the decline in science education, although widespread, has not been

Who will be first to see through the walls of the future?

It could be you and Hughes Electro-Optical and Data Systems.

We built the first working laser, and today our electro-optics, fiber optics, thermal imaging and infrared systems are like second sight.

Like a cryogenic rocket-borne infrared telescope that's extending man's vision farther than ever before.

Like advanced monolithic focal plane arrays that will improve the sensitivity, resolution and range of future thermal imaging systems.

Like precision scanning mirrors that map the earth every eighteen days.

Who will be first to see into the depths of space? Who will be first with intelligent sensors? Who will be first with the electronics of the future?

It could be you and Hughes, a company with over 1,500 different projects in progress.

At Hughes Electro-Optical and Data Systems, we'll introduce you to people, ideas and jobs that could change your world. And maybe ours.

It could be you and Hughes **Electro-Optical and Data Systems**

Call or send resume to:

Hughes Electro-Optical
& Data Systems
P.O. Box 913, Dept. AS-8
El Segundo, CA 90245
(213) 616-3373

Current openings:

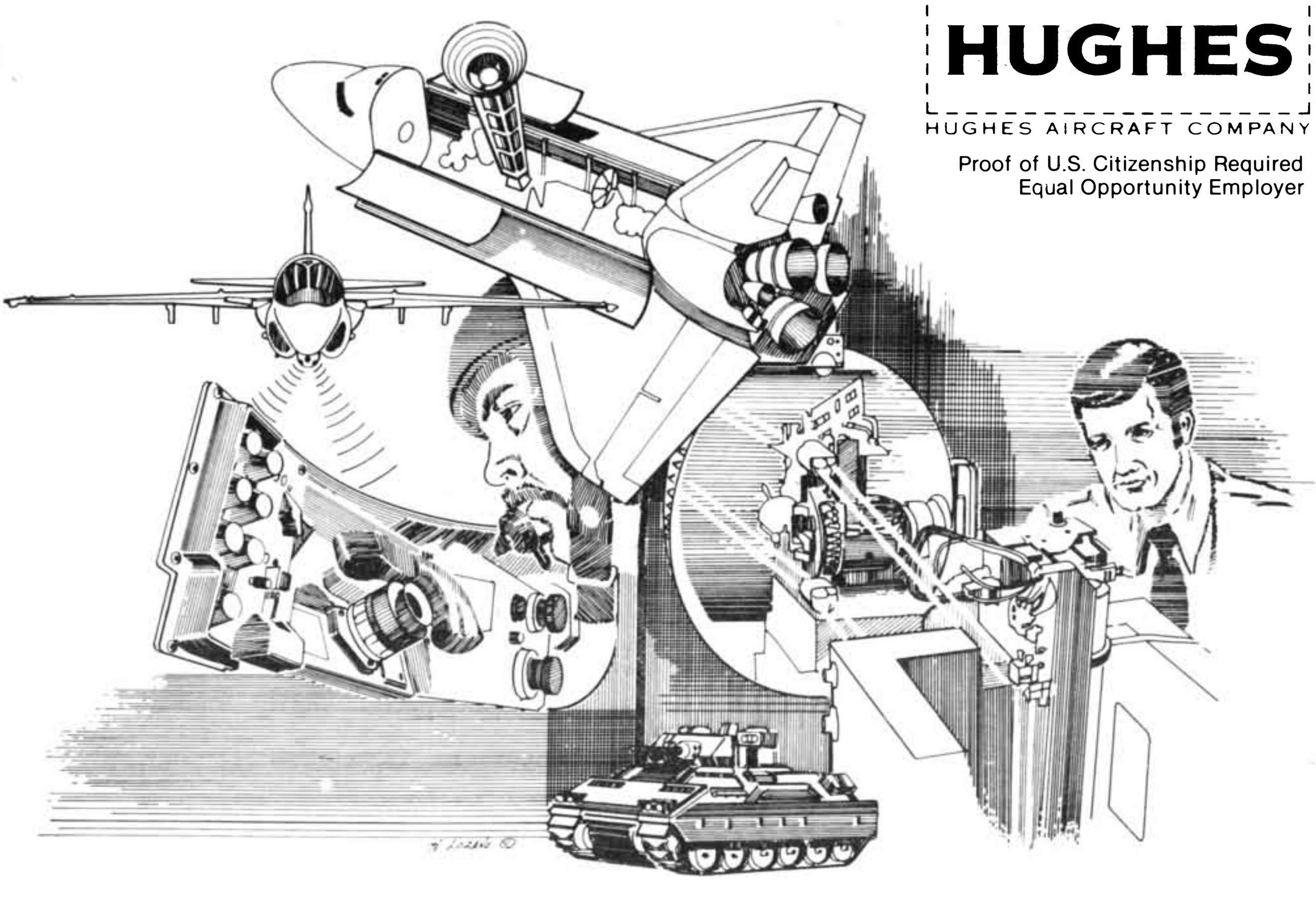
Analog/Digital Design
CAD/CAM/CAT
Components & Materials
Computer Science
Designers/Drafters
Electromechanical Packaging
Facility
Industrial
Materials & Processes
Planners
Production
Production Control
Program Controls
Quality Assurance
Systems/Procedures
Test Equipment Design

Creating a new world with electronics

HUGHES

HUGHES AIRCRAFT COMPANY

Proof of U.S. Citizenship Required
Equal Opportunity Employer



uniform. It is the education of lower-middle-class and working-class children that appears to suffer the most. Indeed, the education received by an elite of middle-class students seems to have improved in the past decade. Hurd noted that test scores of children of couples who live in the suburbs and have had at least some college education showed little decline in the past decade. Furthermore, the number of students taking advanced-placement examinations in science and mathematics increased from 24,000 to 50,000 between 1969 and 1979, and the average scores increased in each of those years. Advanced-placement tests give a high school student the opportunity to receive credit for a year or more of college work.

Izaak Wirszup of the University of Chicago has compared the effectiveness of the science-education systems in the U.S. and the U.S.S.R. For many years Russian schools offered an excellent education for a technical elite but little education for the masses. The system was reformed in 1966, however, and today 98 percent of all students in the U.S.S.R. complete secondary education, compared with 80 percent in the U.S. Russian students are required to take 10 years of geometry, five years of physics, six years of biology and four years of chemistry, compared with one year of each subject in most American schools.

Cosmological Objects

On a photographic plate of the sky a quasar looks like a star. It is nonetheless proposed that each quasar is the extremely luminous center of a distant, hyperactive galaxy. In the main the supporting evidence has been simply the most salient feature of a quasar's spectrum: peaks in the spectrum, which result from the emission of electromagnetic radiation at specific wavelengths by specific types of atoms, are offset markedly toward the red end of the spectrum. On the basis of the red shift a quasar is presumed to be rushing away from our own galaxy as part of the expansion of the universe. It follows that the quasar is a distant (or "cosmological") object, and so it must be prodigiously bright or it could not have been detected.

Recently new lines of evidence have added support to the proposal that quasars are cosmological. One line of evidence is the discovery that the light from at least a few quasars passes near a distant galaxy on its way to our own. The first such quasar was found in March, 1979, by Dennis Walsh of the Jodrell Bank Radio Astronomy Observatory, Robert F. Carswell of the University of Cambridge and Ray J. Weymann of the University of Arizona. They detected what they took to be two quasars only six seconds of arc apart that had identical spectra and red shifts. Eight months later photographs made by Alan N. Stockton of the University of Hawaii at Manoa and by a group of investigators at the Mount Wilson and Las Campañas Observatories revealed a faint elliptical galaxy between the quasar images. The galaxy's red shift places it about half as far away as the quasars. It is now accepted that the intervening galaxy acts as a gravitational lens: its gravitational field bends the light from a single quasar in such a way that light passing on opposite sides of the galaxy is redirected toward the earth. Hence a single quasar gives the illusion of being two. Since then two more instances of a quasar image split by a gravitational lens have been noted.

The discoverers of gravitational lenses find it disconcerting that with more than 1,000 quasars known only three instances of lensing are confirmed. They find it equally disconcerting that if lensing is common, it must focus the light of many quasars, leading observers to misjudge their luminosity. One extreme surmise is that quasars do not exist at all as a distinct class of objects; they are images of bright galaxies further brightened by an interposed gravitational lens. J. Anthony Tyson of Bell Laboratories evaluated this hypothesis in view of the distribution of known galaxies across the sky and the probability that a lensing galaxy will intervene between a quasar and an observer. He found that the idea "is not totally crazy," although the mass of the postulated lensing galaxies had to be made "unreasonably large."

A less extreme surmise is that lensing preferentially affects the images of very distant quasars because they are the ones most likely to have a lensing galaxy interposed in the path of their light toward our galaxy. Thus, it is proposed, the brightened images of distant quasars lead investigators to conclude that quasars are commonest at great distances. This conclusion, which suggests that quasars were abundant in the early universe but are rare today, is in turn a basis for the broader hypothesis that the universe changes with time—in a word, that it is evolving.

Does gravitational lensing account for the apparent "cosmological evolution" of the quasars? In *The Astrophysical Journal* Yoram Avni of the Harvard-Smithsonian Center for Astrophysics and the Weizmann Institute of Science notes that lensing can redirect the photons (quanta of electromagnetic radiation) from a quasar but cannot change their quantity. "This means," he writes, "that while the fluxes from some of the quasars are amplified, the fluxes from the other quasars are deamplified, and there is no net mean amplification." Avni constructs hypothetical distributions of quasars and lenses under the condition that the net flux from quasars is conserved. He concludes "that lenses can in principle mimic an apparent evolution," but he finds it unlikely that lensing accounts completely for the statistics of quasars and their red shifts.

A second line of evidence that quasars are cosmological is reported in *Nature* by Todd A. Boroson and J. Beverly Oke of the California Institute of Technology. Their evidence is direct: they propose that the wispy nebulosity (they call it fuzz) observed just north and south of the quasar 3C 48 is the distant galaxy in which the quasar is embedded. Last year Boroson and Oke measured the spectrum of narrow strips of the fuzz two arc-seconds north and south of the quasar. The spectra turned out to be notable for absorption lines: dips at certain wavelengths characteristic of the spectrum of blue-giant stars, which populate the arms of spiral galaxies. Boroson and Oke note that the lines are red-shifted by an amount that matches the red shift of a number of lines in the spectrum of the quasar itself. They conclude that the quasar lies at the center of a spiral galaxy, albeit an unusually luminous and blue one.

While Boroson and Oke have been examining 3C 48, Tyson and William A. Baum and Tobias J. N. Kreidl of the Lowell Observatory have recorded detailed images of the quasar 3C 273. It is the quasar closest to us, or at least the one with the smallest red shift; it is receding from our galaxy at merely a sixth the speed of light. Along the line of sight to 3C 273 Tyson and his colleagues have found a galaxy. It is too large to be an ordinary elliptical galaxy, yet its light is much too red (in color, not in red shift) for it to be a spiral galaxy. Thus Tyson and his colleagues doubt that gravitational lensing by an ordinary galaxy along the line of sight to some other object gives rise to 3C 273. Tyson now suspects that 3C 273 has no lens. What he and his colleagues have found may be the second established instance of a galaxy in which a quasar is embedded.

Backward Flow the Genes

New evidence of the complexity and plasticity of the genetic apparatus in higher organisms continues to emerge. Five years ago the big surprise was the discovery of split genes. Whereas in bacteria a protein chain is encoded by a continuous array of the DNA subunits called nucleotides, in most of the genes of amphibians, birds and mammals the coding sequences are interrupted by noncoding intervening sequences, or introns. To express a split gene the enzymatic machinery of the cell first transcribes the DNA (introns and all) into RNA. Then the primary transcript is processed. The upstream end of the gene is modified by the addition of a "cap." At the downstream end a "tail" is added: it is a long string of just one kind of nucleotide, adenylate, and so it is called a poly-*A* tail. Finally enzymes cut out the introns and splice together

the coding sequences to make a coherent messenger RNA that is translated into protein.

Split genes were just the beginning. Further analysis of the genome (the full complement of genes) in mammalian cells revealed that there can be more than one copy of a gene or even an entire family of copies. In most cases the introns of the copies are different but the coding sequences are very similar, so that any of the copies can be transcribed to yield a functional transcript. Some members of gene families are pseudogenes, however, which have been altered by mutation and do not yield a functional transcript. In some gene families all the members are grouped on a single chromosome, but in others pseudogenes are found dispersed to a chromosome other than the one carrying the functional genes.

In 1980 Philip Leder and his colleagues at the National Institute of Child Health and Human Development were studying the family of alpha-globin genes in the mouse, which code for one of the proteins of hemoglobin. When they determined the nucleotide sequence of one member of the family, they found not only the random mutations typical of a pseudogene but also something much more striking. The gene's two introns had been cleanly excised; the coding sequences were contiguous, just as they are in messenger RNA. Moreover, the intronless pseudogene was not on mouse chromosome 15 with its functional siblings. It had moved to chromosome 11. The combination of precise splicing and gene movement suggested that the gene is a DNA copy of processed RNA that has somehow been reintegrated into the mouse genome. By now a number of other genes that look like messenger RNA have been discovered. Leder calls them processed genes.

Recent reports from Leder's laboratory and from that of Nicholas J. Cowan of Princeton University describe processed genes that display additional evidence of having passed through an RNA phase. Gregory F. Hollis, Philip A. Hieter, O. Wesley McBride, David Swan and Leder report in *Nature* on a pseudogene member of the human gene family encoding the lambda light chain, a component of antibody molecules. The pseudogene is not on the same chromosome as the known functional versions of the lambda light-chain gene, and it lacks an intron characteristic of all the functional genes.

There is another significant hallmark of RNA processing. Near the downstream end of the gene there is a long sequence of adenylates suggestive of the beginning of a processed RNA's poly-*A* tail. The string of *A*'s is followed by a nine-nucleotide sequence that is a precise repeat of a sequence far upstream, near the beginning of the processed gene. Such repeats are often generated at sites where a stretch of DNA has been cleaved and a new segment of DNA has been inserted into the break. The implication is that a processed RNA derived from a lambda gene was somehow copied into a single strand of DNA that was converted into double-strand DNA, and that the double-strand version was inserted into the chromosome. If the lambda pseudogene was indeed made from RNA, however, the RNA was a highly aberrant transcript: the entire upstream end of the gene is missing.

In another report in *Nature* C. Deborah Wilde, Carol E. Crowther, T. P. Cripe, Mary Gwo-Shu Lee and Cowan describe a processed gene encoding the protein beta tubulin, a component of the cablelike organelles called microtubules. Again the introns are missing, there is a poly-*A* tail and there are repeated segments suggestive of insertion. In this case, moreover, the entire sequence of the pseudogene is roughly homologous with that of a functional beta-tubulin messenger RNA. The beta-tubulin pseudogene, in other words, appears to be derived from a complete processed-RNA transcript.

The only enzyme known to make DNA from an RNA template is reverse transcriptase, an enzyme carried by certain viruses (retroviruses) that have an RNA genome. It has been suggested that retroviruses may have mediated the reverse transcription and reintegration of some processed genes, and indeed there is evidence of viral sequences in some of them. It is also possible that some mammalian enzymes may transcribe RNA into DNA under special circumstances. The reversed flow of genetic information must be rare, but over the long span of evolution even a rare process might suffice to explain the presence of a large number of processed genes.

Spring Cleaning

The particles called intermediate vector bosons seemed almost within grasp. A machine for making them had been assembled at the European Organization for Nuclear Research (CERN) in Geneva. Two large detectors stood ready to record evidence of their brief passage. In a test of the apparatus last fall all had gone well. In April the search was to begin in earnest, and by summer it seemed possible that at least preliminary results would be announced. Instead there was a mishap, and the search has been delayed by six months.

Intermediate vector bosons are the hypothetical particles that are thought to act as carriers of the weak nuclear force. They are much heavier than any particles yet observed, and the CERN device, in which protons and antiprotons collide head on, is the only one now operating with enough energy to create them (see "The Search for Intermediate Vector Bosons," by David B. Cline, Carlo Rubbia and Simon van der Meer; SCIENTIFIC AMERICAN, March).

The mishap damaged the larger of the two detectors, designated UA1. In the course of preparations for the April experiments the annular vacuum tube of the proton-antiproton collider was being "baked" to remove stray vapors. The detector was being protected from the heat by a flow of air when an accumulation of rust, oil and grime was blown into it. The most serious contamination was on the outer surface of the electronic particle counter called a drift chamber, which forms the core of the detector. On the surface there are some 50,000 connections to wires threaded through the chamber; each connection had to be cleaned individually.

The cleanup was completed in June. By then, however, the accelerators at CERN were committed to other experiments, and the boson hunt had to be put off until September.

Torn Knapsack

Two-key encryption systems were a mathematical novelty in 1975; they are now being applied to the coding of banking transactions, electronic mail, signals specifying the routing of telephone calls and other kinds of electronic information that would otherwise be vulnerable to the forger or the eavesdropper. The two-key ciphers, which are also called public-key ciphers, seem to offer both convenience and security. Now Adi Shamir of the Weizmann Institute of Science in Israel has shown that one of the first two-key systems is insecure. He did so by proving that a mathematical problem called the knapsack problem, which had been considered exceedingly difficult, can be solved rapidly by a simple computer algorithm.

The practicality of a two-key encryption system depends on the ability of the authorized user (or his computer) to quickly carry out relatively easy mathematical operations, such as finding the product of two large prime numbers. The security of the system is based on the difficulty of doing the operations in the opposite direction; for example, the task of the forger or the eavesdropper might be to find the two prime factors of a large composite number. Such operations can be so time-consuming that they are computationally impractical.

The knapsack problem is a venerable mathematical puzzle whose application to cryptography was suggested in 1976 by Ralph Merkle and Martin E. Hellman, who were then at Stanford University. Given a large set of objects, each of which has a different weight, the problem is to determine which subset of the objects will fit exactly into a knapsack with a given carrying capacity. In a cryptographic system the objects are simply numbers and the total weight is their sum. The authorized user merely chooses a subset of numbers and finds their sum, which is easy. The unauthorized user, however, knows only the sum and must deduce which numbers were added in order to form it.

Shamir was able to solve the knapsack problem after he noted that it could be transformed into a mathematically equivalent problem in integer programming. Integer programming is a well-established means of assigning scarce resources to several products in order to maximize some quantity, such as profit. Shamir showed that the integer-programming problem to which the knapsack problem is equivalent can be solved by an algorithm recently invented by Hendrik W. Lenstra of the University of Amsterdam. Hence Lenstra's algorithm can also solve the knapsack problem.

Shamir's successful attack on the knapsack problem does not jeopardize any of the two-key cryptographic systems that are now in service or that are being considered for commercial applications. The system invented by Merkle and Hellman has never been adopted for coding purposes, largely, it appears, because mathematicians suspected that it could be cracked. Several other two-key systems, however, may soon face similar challenges.

Of these the most important is the technique based on factoring. It was devised by Ronald L. Rivest of the Massachusetts Institute of Technology, Shamir and Leonard M. Adleman of the University of Southern California. Factoring has been a central concern of mathematics for at least 2,500 years, and lately it too has been yielding some ground to persistent assaults by mathematicians. With a large computer prime factors that have as many as 50 digits can now be extracted from a composite number in several hours. Gustavus J. Simmons of the Sandia National Laboratories projects that within a year prime factors will have to exceed 100 digits in order to be secure.

In spite of the lively game between code makers and code breakers, the convenience of the two-key systems still outweighs the fear that they will become insecure. In a conventional cryptographic system security is compromised if the key to the cipher is not kept secret. In a two-key system an encoding key is generated by every potential recipient and is made publicly available. In the method proposed by Rivest, Shamir and Adleman each recipient would publish the product of a pair of prime numbers. Anyone who wanted to send a message to a particular recipient could employ the published composite number to encode the message, but only the intended recipient, who alone knows the factors, could decode it.

Two-key systems are now being incorporated into fast, low-cost microelectronic chips. The Hewlett-Packard Company is testing a chip capable of preventing electronic "forgeries," in which the origin of data transmitted electronically is misrepresented. The chip employs a technique invented by Hellman and Stephen Pohlig of the Massachusetts Institute of Technology that enables two people to exchange information over a public communications channel about how they intend to encode messages to each other, without revealing the encryption method to a third person who might be listening in.

Probably the most intriguing application of a two-key system is one suggested by Simmons. If a ban on the underground testing of nuclear weapons could be negotiated, compliance might be verified by on-site seismic monitoring stations. Each station would have to transmit information that could be read by both sides but would be impossible to forge. Simmons points out that a two-key code based on the multiplication of four prime numbers could meet all the requirements. Each country could be certain that only seismic data were being transmitted to the other side and yet would be assured that the data received had not been tampered with.

256 Kilobits

Random-access memories, or RAM's, which are the microelectronic chips made in the largest quantities, have been steadily growing more capacious. The growth is now accelerating. An early RAM could store roughly 1,000 bits of information; it was designated a one-kilobit, or 1K, RAM. Each succeeding generation has been larger by a factor of four, progressing from 4K to 16K to 64K. The next generation, the 256K RAM, was not expected to be commercially available until about 1985. Now Hitachi Ltd. and Fujitsu Ltd. in Japan and Motorola, Inc., in the U.S. have indicated that they may get versions of the 256K RAM on the market next year.

Among problems to be overcome in this endeavor are yield, size and redundancy. Yield is the percentage of chips that prove to be satisfactory at successive stages of the manufacturing process. In most cases a chip of a new generation is not brought to market until its yield is 10 percent; by then the yield of chips of the previous generation is generally about 50 percent. The 64K RAM is now being made in the U.S. at yields of from 25 to 30 percent. Manufacturers apparently will have to speed up the pace of development if they are to achieve economic yields of the 256K chip by next year.

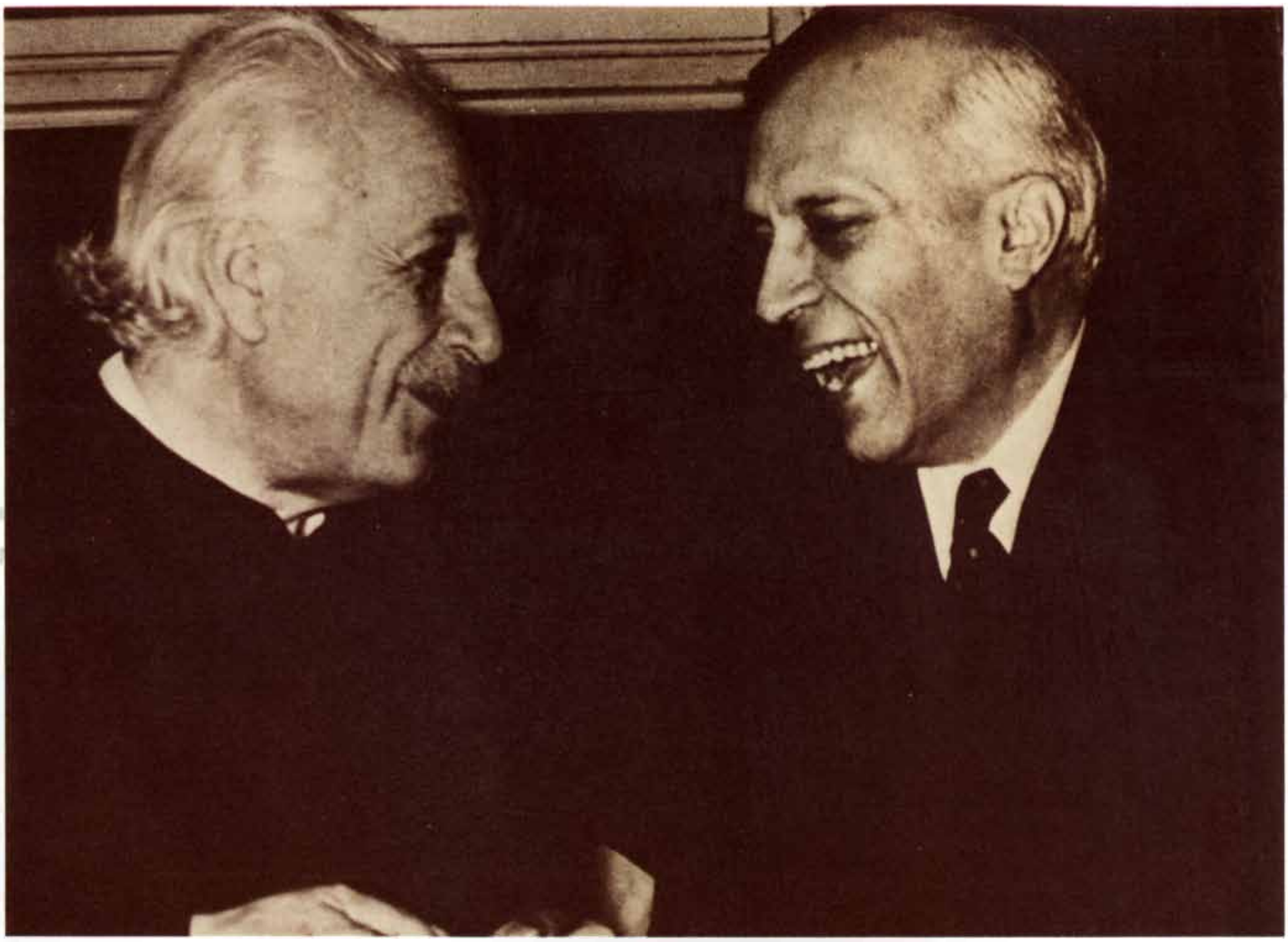
Size is related to yield, because a larger chip is more likely to have defects. The 256K chips made in the U.S. are expected to be from 30 to 60 percent larger than typical 64K chips. That would give them an area of 40 or 50 square millimeters, compared with the 25 or 30 square millimeters expected for the Japanese 256K devices. On the other hand, a larger chip has larger memory cells, which could make each cell less susceptible to failure. The U.S. manufacturers plan to compensate for lower yields by building their 256K chips with extra rows of memory cells that can be switched in to replace malfunctioning ones. The value of such redundancy is under debate because it may raise the cost of testing chips and reduce their long-term reliability.

The Botanical Light Guide

A growing seedling can respond to light even when most of the plant is still covered by the soil. How is the signal that triggers the response transmitted from the exposed shoot to the rest of the plant? Evidence is at hand that in at least three kinds of plant (corn, mung bean and oat) the tissues of the seedling's new stem conduct light over considerable distances to potential sites of photoreception. In doing so the tissues function much as a bundle of optical fibers does. The work is described in *Proceedings of the National Academy of Sciences* by D. F. Mandoli and W. R. Briggs, each of whom is associated with both Stanford University and the Carnegie Institution of Washington's Department of Plant Biology in Stanford.

Mandoli and Briggs were spurred to their investigation by observing the extreme photosensitivity of etiolated oats: young oat plants growing in the absence of light, a condition in which the plant develops an immature white stem and a few pale yellow leaves. Tiny amounts of light induce photomorphogenesis (a change in the pattern of growth) while the plant is still underground. Mandoli and Briggs found that even a brief exposure to "safe" illumination, a green light botanists have employed thinking that it will not evoke a response in plants, can induce photomorphogenesis in the oat seedling. They thought light-guiding by tissues might enable photomorphogenesis to begin well before most of the seedling had emerged from the soil.

To test the hypothesis Mandoli and Briggs directed light from a .5-milliwatt helium-neon laser onto one end of a deliberately curved segment of etiolated tissue and measured the output of light at the other end with a photomultiplier. They found that the tissues of several plants transferred the light over distances of 45 millimeters or more. The curving of the tissue was one of several tests to determine whether the transmission was the result of simple light-scattering in the tissue or of internal reflection, the mechanism in optical fibers. Scattered light cannot follow a curve, but internally reflected light can. All the tests were consistent with transmission by internal reflection.



A meeting of minds; Dr. Albert Einstein and Jawaharlal Nehru at Princeton University in 1949.

"It is science and technology which has made other countries wealthy and prosperous, and it is only through the growth of technology that India shall become a wealthy and prosperous nation."

Jawaharlal Nehru

Science and technology in INDIA the past, present, and future.

"Jawaharlal Nehru clearly recognized the importance of science. It is due to him that we as a people have at least some consciousness of science. He founded a chain of national laboratories and established what is now the Bhabha Atomic Research Center in Trombay. It was under his inspiration that the Government adopted a Scientific Policy Resolution in 1958.

Science has long ceased to be considered an esoteric pursuit. It has to be part of the life of every Indian, the soldier, the farmer, the worker, the housewife and the student."

Indira Gandhi

CONTENTS

Text by Michael Frenchman

The past: cradle of science in the Indus Valley.

"INDIA has a population of 660 million speaking 15 main languages and is a vast and complex sub-continent. It ranges from the snow-capped Himalayas in the north down to the Ganges Plain and the lush green of the tropical south. Its development stretches back almost to the dawn of time with its first technological revolution around 4,000 BC. The people of the Indus Valley civilization made use of the wheel and possibly the plow, they smelted and forged metal, and they developed elaborate fire and flood controls for their farms and villages. As then, oxen still plow the land, but today's scientists send rockets hurtling into the same skies mapped more than a thousand years ago by the first astronomers. Science and technology in India is a continuing story."

IT BEGAN with the technological revolution that took place between 4,000 BC and 3,000 BC when India, along with Egypt and West Asia, emerged from the mists of antiquity. During this period India was both technically and socially well advanced. Archaeological excavations have shown that there were then thriving cities and centers of commerce. The towns were well planned, the buildings were constructed of standardized fired bricks, and drainage systems were installed. Ports such as Lothal were developed as export centers for early manufactured products of smelted copper and bronze; polychromic paints were used to decorate pottery with a variety of animal and bird designs.

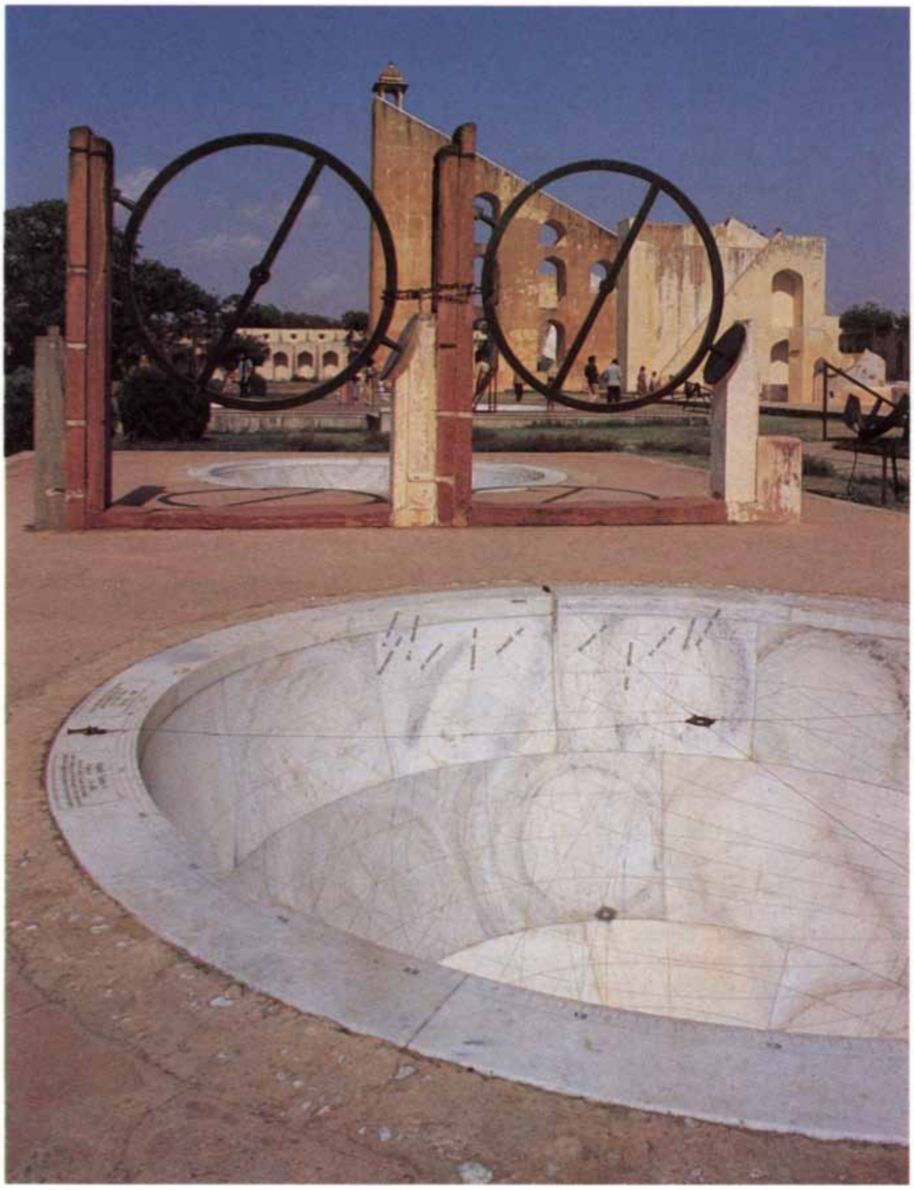
It was not until around 800 BC that the smelting of iron began and then, as today, it was the result of a technological spin-off from the primitive arms industry that existed at the time. In order to improve early weaponry, metallurgical experiments took place which later led to the use of iron in agriculture and for other crafts. The proficiency of the ironsmiths still stands today in the form of the 24 ft. 3 in. high six-tonne Iron Pillar in Mehrauli, Delhi. It is made of wrought iron with an iron content of 99.72 percent and appears to have been protected against rusting by the application of a thin coating of manganese oxide. An even larger pillar, from a later period, about 1200 AD, was some 42 ft. high but now lies in several pieces at Char.

Several centers existed for iron founding and copper smelting, and huge copper and brass images of the Buddha were frequently cast during the Classical Period from 400 AD to 800 AD. According to an account by a Chinese pilgrim, one image of Buddha towered 80 ft. high. Other Buddhas and statues were produced by the *cire perdue* (lost wax) process which is fully described in Sanskrit texts of the time. Construction engineers in the 12th century were using iron girders and beams on a scale unknown in any other part of the world. The most significant use is in the temples at Puri and Konark in Orissa. The Puri temple contains 239 iron beams, and one of the beams at Konark is 35 ft. long by about one foot square. All are 99.64 percent iron and are produced in a similar manner to the Delhi iron pillar.

Religion played a vital role in the early days, too. Some of the oldest recorded mathematical works such as "Sulva-Sutras" showed how geometry was used in the design and construction of altars. The early mathematicians of the 5th century, like Aryabhata and Bhaskara I, and later mathematicians like Brahmagupta, Mahavira, and Bhaskara II, used and developed most mathematical formulas that we know today. A unique development followed in the 6th century—the use of the zero to represent a group of 1–9 digits or decimal place numeration. A symbol, in this case a dot, was inserted to represent the missing unit value (ten, hundred, thousand, etc). This breakthrough was described by a 7th Century Syrian writer as "a computing that surpasses description." By about the 9th century a zero was used instead of the dot. Most historians agree that the use of the zero originated in India and spread to other cultures, an early transfer of "know-how." This emphasis on numbers occupied most of the great minds of the day through successive ages, even down to modern times.

Aryabhata I, a leading exponent of his time, dealt with square roots, cube roots, areas of triangles, volumes of pyramids, and the area of circles. The most important breakthrough was related to the geometry of the circle and the value

Fine engraving on an astrolabe made in Punjab in 1862.



One of the great observatories at Jaipur which was built by Raja Sewai Jai Singh II between 1718-34.

of π. Before Aryabhata's time, approximate values had been established for π, but it was he who gave it the value of 3.1416, which was the sum used centuries later by the Arabic, Chinese, and European mathematicians. Indeed, the Indian mathematicians were responsible for establishing most of the ground rules in use at the time and that were then further developed by others who followed. Bhaskara II is perhaps best known for his algebraic work and the compendium "Siddhartasiromani."

Allied with mathematics was the development and use of astronomy, whose earliest days were reported by Aryabhata and Varahamira. Many later scientific works owe their origins to what are called the "Five Siddhantas," of which the "Surya" had a major influence on astronomical traditions in India. Aryabhata was the first to adopt a scientific approach to astronomy, establishing it as a separate discipline from mathematics. He also conceived the theory that the earth was a sphere rotating on its own axis. This and other contemporary radical ideas were the basis for the Aryabhatiya, an astronomical textbook which was to have wide influence for several centuries. Astronomy and its development was essential for both religious and practical purposes. Calendars and feast dates had to be worked out and the seasons identified for growing crops.

Great observatories were built in the late medieval period—Benares (now Varanasi), Ujjain, Mathura, Delhi, and Jaipur—by Raja Sewai Jai Singh II between 1718 and 1734. As lunar and solar calendars were tabulated, new instruments for measuring the stars and the planets were developed to add to the discussion on whether the earth was flat, spherical, or revolving, but the geocentric concept continued to be the dominant one.

With the introduction of the astrolabe to India following Islamic influence, the instrument became the most important astronomer's aid. This led to Lahore (now in Pakistan) becoming an early scientific instrument-making center in the 17th century. About 40 astrolabes from that period still exist today. Other instruments of the period included quadrants and armillary spheres.

The leading exponent of the age was undoubtedly Jai Singh, who produced an elaborate set of astronomical tables, the "Zij Mohammad Shahi," which were based on extensive research and astronomical knowledge not just in India but in other parts of the world from China to Europe.

The amount of interest in mathematics, astronomy, and medicine can perhaps be gauged by the fact that between the 12th and 18th centuries more than 10,000 manuscripts were written in Sanskrit, Arabic, or Persian on these or allied subjects. Even more works were produced in the other languages of India.

There were many varied early science writers who not only produced original works but also wrote respected commentaries and translations of foreign observations. Two of the best known were Ustad Ahmad Al-Mi'mar Lahori, builder of the Taj Mahal, and Lutufu'llah, one of his sons, who was known as "the geometer."

In parallel with the development of mathematics and astronomy were the advances being made in the medical field, principally the "Ayurveda" or "science of life", which included the development of dietetics and food technology. The emphasis in the beginning was on diagnostics—cause and symptoms. In the Ayurveda theory, the body was governed by three factors called the "tridosa"—wind, bile, and phlegm—which corresponded to the humors of the Greek and medieval physicians of Europe. Good health depended on maintaining the balance between these three humors, aided by the physician if the balance was upset.

The ancient physicians not only attended to the sick but also observed nature as a whole in the search for curative powers from flowers, fruits, seeds, leaves, and minerals. Although the physician-philosopher had to have an immense breadth of knowledge and perception of Ayurveda, he would also specialize in one particular aspect. Good hygiene, nursing, drugs, and acceptance of the physician—principles which still stand today—were essential elements of health care. The early surgeons were masters of their craft, too, and developed more than 120 instruments including scalpels, catheters, syringes, and forceps.

Surgical expertise was something which the early Hindus excelled at and their knowledge spread far and wide and was known to the Arabs, Greeks, and Egyptians. Operations included those for cataracts, laparotomy, lithotomy, trepanning, and rhioplasty. Some of the earliest known Ayurvedic practitioners were Charaka and Sushruta in the early centuries before Christ, and the eminent clinician Vagbhata in 600 AD. Much of their work has been recorded. One of the best known later works in existence is the treatise on the human eye by Zarrin Dut (1087–1088) which identified visible ailments and those which could not be seen, suggesting prophylactic cures for the latter. A little later the invading Persians brought an important Islamic influence to bear on medical knowledge of the time. Continued research and exploration went into finding and using new drugs for both prevention and cure. This often acted as a catalyst for development and progress in alchemy and chemistry, which was often meticulously recorded. There is a wealth of ancient manuscripts, many of them written on dried palm leaves.

Because of the climatic conditions of the subcontinent, efforts were continuously being put into developing new methods of food preservation such as chutneys, murabbas, and the use of sherbet. From the earliest known days great importance was always attached to hygiene and sanitation and the diseases which might result if proper standards were not observed. Studies were also made of plant genetics and diseases such as those recorded in works like "Vrksayurveda" published in the late 15th or 16th century.

The development of chemistry, apart from its use in producing drugs for the physicians such as compounds of mercury, arsenic, and herbal extracts, also led to applied technological processes in dye, paper, perfume, pyrotechnics, and sugar production. Allied with this development were the experiments in the use of new mineral ores and mixtures of copper, iron, tin, gold, silver, and lead.

It was war and the need for better weapons that led to a comparatively rapid development of the armaments industry and the use of bronze, iron, and steel for cannon-making from the 16th century onwards. A couple of centuries earlier the custom of coating copper cooking vessels with tin had been introduced. A special alloy ware, Bidari, was formed, and the extraction and use of zinc were developed; none of which were known in Europe until many years later.

Industrial technology reached its highest peak during the 16th and 17th centuries with the manufacture of cannons and hand guns of all kinds. These achieved considerable mechanical sophistication as well as aesthetic appeal. The 17th century hand guns had flint locks; and cannon barrels were made by rolling and twisting flat iron with heat edge welding so that they could be bored from inside. Artillery techniques moved rapidly and expanded in many directions with the introduction of bamboo shafted iron rockets.

During the 16th century the armorers were producing cannons of immense size in bronze and iron. The Great Gun of Agra, in brass, weighed over 53 tonnes and the even bigger Malik-i-maidan at Bijapur had a bronze barrel 14 ft. 3 in. long with a diameter of 4 ft. 10 in. at the mouth—the largest in the world at the time.

Religion, mathematics, and agriculture from the earliest recorded days were the dominant factors in everyday life. This was to a large extent governed by the seasons of nature. The need to measure rainfall and predict possible weather changes were as important then as monsoon warnings are today. Calendars for crop sowing and rain gauges were evolved; proper farming practices were established and various implements for tilling the soil developed. As well as the physician-astrologer, the blacksmith and carpenter also assumed greater roles in the life of the village. Heavier ploughshares of up to 2 kg. in weight enabled deeper cultivation of the soil to take place and iron continued to be the supreme material right up to the introduction of British rule in the 19th century.

As Professor M. G. K. Menon, eminent scientist and a member of the Planning Commission, points out, science and technology were already part of Indian culture and tradition when British rule began. In many instances, Indian developments were ahead of those in Europe. Under the British, emphasis on science and technology was to foster the needs and requirements of the then government by increasing trade, communications and defense. This resulted in several important advances such as the first scientific survey of the country, and a network of meteorological stations, the building of canals and the establishment of the railway system. Various academic institutions were created such as the Asiatic Society in Calcutta in 1784, the Indian Association for the Advancement of Science founded in 1876 by Mahendra Lal Sircar, and many others. All these helped create a greater awareness of science in the country and eventually led to the birth of modern science in India between 1890 and 1940. As Dr. B. V. Subbarayappa, of the Nehru Center, explains, "There is an enormous breadth of intellectual ability in India and the Indians have always had the ability to assimilate knowledge from the very earliest days—we have, if you like, been aware of the transfer of technology for a very long time."

When one visits research institutes, laboratories, and scientific establishments, it is impossible not to be impressed by the enthusiasm for science which exists at all levels. But, according to Professor Menon, there is very scant outside appreciation of the fact that the country has based itself on science and that the people have benefited from the following wake of technology.

The wrought iron pillar at Delhi which has remained rust-free for 1,500 years. Linking arms around the 23 ft. high pillar is supposed to bring luck. It has an inscription (below) in Brahmi script.

S. Ramanujan

Sir J. C. Bose

Sir C. V. Raman

The renaissance of science

THE END of the last century and the first fifty years of the present one was a period of renaissance in science for India. During this time some of its most able scientists were produced. The earliest of the great pioneers was Sir J. C. Bose (1858–1937) who demonstrated the production and detection of 5-25mm electromagnetic waves. He went on to develop sensitive measuring instruments for monitoring plant growth and behavior.

Perhaps the greatest discoverer of all was C. V. Raman, who was born in 1888, the son of a mathematics teacher. In 1928, Sir C. V. Raman, as he later became, discovered the "Raman effect"—the quantum theory of diffusion of light. He was awarded the Nobel Prize for his outstanding work. He was a man of considerable all-round ability and, as well as being a scientist of distinction, he was a musician and artist. He once remarked: "The history of science has shown that real fundamental progress is always due to those who had ignored the boundaries of science and who treated science as a whole." This could not have been more true than in his case. Most of his experiments were done in Calcutta and later in Bangalore where he established the Raman Research Institute. He died in 1970.

The development of mathematics and physics played a vital role during this renaissance period, a reflection on the great heritage of science in the country. Another example of this was the work of S. N. Bose which led to the Bose-Einstein statistics. An outstanding astrophysicist was Meg Nad Saha (1893–1956) known best for the Saha equation relating to stella spectra, his research into ionization, and the propagation of radio waves in the upper atmosphere. He was founder-director of the Institute of Nuclear Physics in Calcutta in 1955 and a leader of modern astrophysics.

Srivanasa Ramanujan, who lived only from 1887–1920 was a mathematical prodigy of great originality and was largely self-taught. He was one of the first Indians to become a Fellow of the Royal Society in 1918. In more modern times, one person who was to have the greatest impact was Dr. Homi Bhabha, the father of India's nuclear program. He was a man of vision and campaigned ceaselessly for more resources for research. He was instrumental in establishing the Tata Institute of Fundamental Research which was later incorporated into the Bhabha Atomic Research Center. He was born in 1909 but his life ended tragically in an air crash in 1966, just two years before the opening of India's first nuclear power station.

India's space program was first envisaged by Vikram Sarabhai (1919–1971), who studied cosmic ray effects under C. V. Raman. His later research led to the discovery of new solar relationships of cosmic ray variations.

Another Nobel Prize winner is Har Gobind Khorana, a biochemist who was born in 1922. He has won many awards for his medical research, including the production of artificial genes. He is now living and working in the United States where he has held the Sloan Chair of Chemistry and Biology at the Massachusetts Institute of Technology since 1970.

Homi Bhabha

V. Sarabhai

H. G. Khorana

The present: the role of science today and its father, Jawaharlal Nehru.

"Science has developed at an ever-increasing pace since the beginning of the century, so that the gap between the advanced and backward countries has widened more and more. It is only by adopting the most vigorous measures and by putting forward our utmost effort into the development of science that we can bridge that gap. It is an inherent obligation of a great country like India, with its traditions of scholarship and original thinking and its great cultural heritage, to participate fully in the march of science, which is probably man's greatest enterprise today."

Bringing the grain to market by camel-drawn cart. As part of a nationwide scheme millions of traditional carts are being fitted with rubber tires to prolong their life and make them more efficient.

THESE PROPHETIC words are from the 1958 Scientific Policy Resolution adopted by Prime Minister Jawaharlal Nehru's government a decade after independence had been achieved. It was Nehru who had the drive and vision of a great India based on science. More than anyone else, he laid down the foundations for a 20th century science and technology policy that would enable India to be self-reliant.

Professor S. Nurul Hasan, vice-president of the Council of Scientific and Industrial Research (CSIR) tells how Nehru constantly emphasized the importance of science to everyone. During his time many of the research laboratories were established, he made himself president of the CSIR, the Atomic Energy Center was created, and the University Grants Commission formed.

"He was convinced," says Professor Hasan, "that without modern science he could not change the quality of life of the people." Nehru developed science and technology in a conscious way as a major force for social and economic change. He himself remarked once: "It is science alone that can solve the problems of hunger and poverty. The future belongs to science and to those who make friends with science."

Nehru also gave a great boost to education and established five institutes of technology at Bombay, Delhi, Kanpur, Kharagpur, and Madras. He also set up agricultural universities, medical institutions, and colleges. He believed that the only way India could catch up with development was by having trained scientific manpower. He pushed hard for overseas education with Britain, the Commonwealth countries and the United States. He insisted that in all bilateral agreements there should be a clause to allow Indian students to study in foreign universities. Vigorous measures were taken for the development of science through increasingly large financial allocations after the first five year plan of 1951–56. More universities and laboratories have been formed and, according to Professor M. G. K. Menon, scientific and technical manpower in the country is now estimated at 2.5 million. About 160,000 qualified scientists and technicians are produced each year.

Research expenditure under the current 1980–85 plan amounts to approximately $1.8 billion, of which 70–80 percent goes to agriculture, atomic energy, space development, industrial research, and general science and technology. The total expenditure represents about 0.6 percent of the gross national product.

India's R&D infrastructure is made up of about 130 laboratories and institutions within the framework of the Council of Agricultural Research, the CSIR, Council of Medical Research; the Departments of Atomic Energy, Space, Electronics, Science and Technology; and the Defense Research and Development Organization. There are also some 700 special purpose R&D units set up within various public and private sector establishments. A new development has been the growth of consultancy organizations, which number about 150 and employ some 20,000 experts.

The stimulus for much of this growing scientific infrastructure came from Nehru, who not only encouraged people like Dr. Homi Bhabha to develop nuclear energy, but played an important decision-making role himself.

This lead has been followed by the present Prime Minister, his daughter Indira Gandhi, who chairs the Cabinet

Experimental light industrial solar collector (left) and a tank for growing blue-green algae which can be used for nitrogen fixation, instead of chemical fertilizer, in rice paddy fields.

Committee on Science and Technology. She is also Minister of Science and Technology, and is directly responsible for the nuclear, space, and electronics departments. She is herself a firm believer in the value of science and the importance of self-reliance for India because of its size and the diversity and complexity of its problems. She remarked earlier this year during a visit to Britain to explain some of her country's achievements that, "Smaller countries could perhaps develop by acting as feeders to the economies of the advanced countries, 'by plugging into their grid' as the phrase goes. India simply cannot meet all her requirements that way. It is too vast to be bailed out by any one country or group of countries."

This was why self-reliance, as opposed to self-sufficiency, had to be at the very heart of science and technology planning. Neither "true defense nor true development can be bought or borrowed. We have to grow them ourselves... self-reliance does not mean making everything ourselves but acquiring the capacity do so."

This is an important distinction emphasized by Professor Menon, who says that self-reliance "demands national commitment and political will and involves many facets, including technology policy..." A crucial element was a self-reliant base of science and technology itself—which could not be built without significant basic research. He goes on, "That is the only way to generate basic understanding which is not restricted to specific knowledge in an area, but provides the ability to attack and solve problems over a wide spectrum; and it is this ability that basic research, more than anything else, develops to the highest degree."

Commenting on the need for basic research shortly before her departure for the United States, Mrs. Gandhi stressed that it was important that it should be a continuing process. It was not "something which could be left in the air." It helped in the application of science and technology, which was more intelligible to the common people.

She, herself, is not a believer in appropriate or intermediate technology if it means sacrificing the goal of higher research. During the conversation, she added: "I don't believe that small is beautiful in all cases, but I do think that while you are doing the bigger things which are essential you have to do the medium and small scale as well. The bigger should not shut off the other, either. You must balance it. While there are lots of things we would like to change on a bigger scale, there are many small things which can have an immediate effect on the life of the average rural worker."

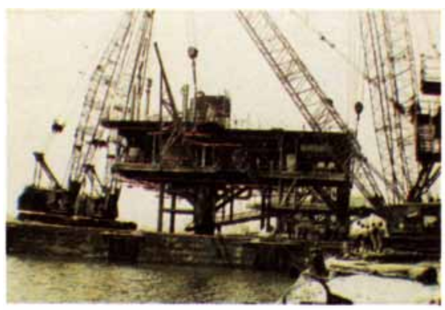
She feels that, unfortunately, sometimes scientists do not want to give too much time to these small things like improving bullock carts by fitting them with rubber tires and putting ball bearings in corn mills. She would like more research done on housing, for example, something which could make a tangible difference to people's lives. Generally, Indians wanted progress and had a tremendous ability to assimilate and adapt to new ideas and to "Indianize" them, but to some extent she felt there would always be a conflict between immediate gain and long-term interests. In an earlier reference to advanced science, she commented that space and nuclear programs were not just a matter of vanity, but of necessity—as they were of immense practical value in providing energy and offering the ability to map the country better, and provide desperately needed communications. Speaking about last year's Cancun conference, which she felt had kept the door open for a continuing North-South dialogue, she thought that the entire developing world had had a raw deal from the colonial nations which had themselves only become industrialized at the expense of exploiting the developing countries in the past.

"It was much easier for them as they could build up quickly, but we are harassed at every step... the West does not realize what extra burdens we have to carry because of our late start, and because of obligations—which they did not have to fulfill—with regard to human rights and so on. In today's world we are at a disadvantage compared to other countries," says Mrs. Gandhi.

"You might say, isn't that asking for charity, but we don't think so. Unless there is a better economic balance in the world it is bound to increase tension and affect the economies of the industrialized West. And the same thing holds good for the situation within our own country unless we can create a better balance and solve the problems of poverty and economic backwardness."

India regards science and technology as one of the keys to tackling these issues. There is a feeling that others in the world do not really understand the reason behind the country's drive for science, and that not only its past, but its present achievements are being over-

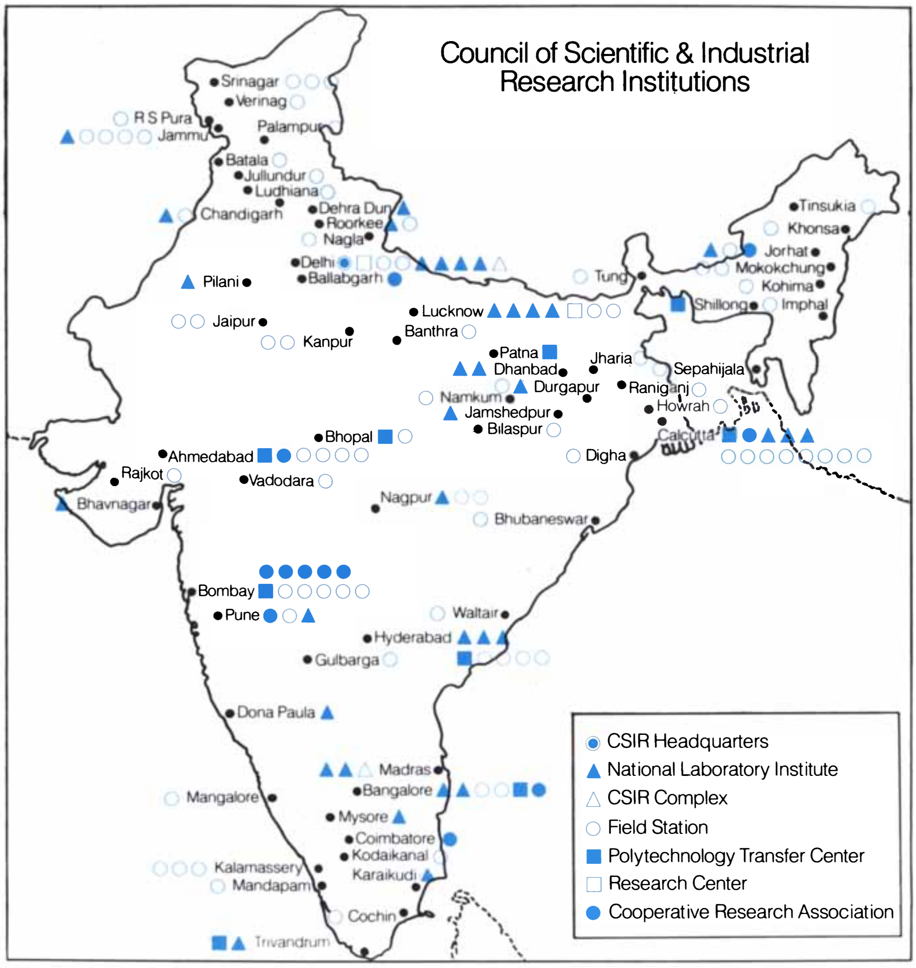
An Indian-designed and -built offshore oil platform at the Bombay Shipyard.



looked. Impressive progress is being made in plant breeding and genetics; in the fight against disease and sickness; in providing more food grains (production has risen from 50 million tonnes to 135 million tonnes in thirty years); in experiments with birth control by immunization; and by concerted efforts to break new barriers in nuclear physics and space technology. Indians are now plumbing the ocean depths and this year sent their first expedition to the Antarctic—all in the cause of self-reliance.

Thousands of Indian scientists are now working abroad to gain experience and continue research work, although efforts are being made to attract them back home where their skills are desperately needed in the thrust for more knowledge and development.

As the celebrated Indian scientist C. V. Raman put it: "Unless the real importance of pure science and its fundamental influence in the advancement of knowledge are realized and acted upon, India cannot make headway in any direction and attain her place among the nations of the world.... There is only one solution for India's economic problems and that is science, and more science, and still more science."



Advances in agriculture and health

SINCE INDEPENDENCE, major strides have been achieved in the development of a strong industrial base for the country, which has been supported by a policy of planned investment. Rapid progress has been made in the production of steel, non-ferrous metals, cement, paper, chemicals, fertilizers, as well as heavy engineering and manufactured goods. These include shipbuilding (India is among the world's top ten shipbuilders), the construction of offshore oil drilling platforms and equipment, locomotives, earth-moving machinery, tractors, buses, and other commercial vehicles. India has also developed its own technology for petrochemicals and plastics.

Significant changes have taken place in textile production, which is one of the oldest industries in the country. Traditional silks and cottons, which account for the major proportion of production, have now been augmented by many man-made fibers. But, in order to bring more economic activity to rural areas there has been a major revival in the handloom weaving industry, which has about ten million workers producing 3 billion meters of cloth a year. The mills, which have another one million workers, produce 7 billion meters a year.

Steel production from six major steelworks has more than doubled in the last 20 years to 7 million tonnes a year, and provides the basis for a whole range of manufacturing industries. This has enabled a large degree of self-sufficiency to be reached in certain production areas, such as manufacturing machinery for cement making, sugar refining, material handling equipment, and the engineering industry in general. There has also been some progress in producing more consumer goods, many of which form valuable exports.

The expansion of the industrial base of the country has meant a strengthening of the energy sector, particularly the supply and transmission of electricity to both rural and urban areas. Thirty years ago it is estimated that only some 3,000 villages received electricity. Today the number has risen to 300,000. The main sources of electrical generation are from hydro and coal-burning power stations. There is also a contribution from nuclear stations. The total installed generating capacity has increased from 2,300 MW in 1950 to approximately 33,000 MW.

Science and engineering have been instrumental in developing the energy program, where considerable efforts have been made to improve both generation and transmission by building new dams for both hydroelectric powers and for irrigation purposes. Hydroelectric generating capacity now totals nearly 12,791 MW compared to only 560 MW in 1950.

Major new investments continue to be made in the energy sector, especially in terms of coal production which is currently just over 100 millions tonnes. This is expected to rise to 400 million tonnes by the end of the century.

Oil is now becoming increasingly important. The Bombay High offshore field is the most significant field with a current production of around 5 million tonnes. This is expected to rise to about 13 million tonnes by the end of the current Five Year Plan in 1984–85.

IN THE FOLLOWING PAGES developments in other sectors will be examined. These include achievements in the use of nuclear power for peaceful purposes, a look at manpower resources and technical skills in engineering and electronics, agricultural and medical research, and the space program.

The future: Bangalore, a brain bank for tomorrow's technology

BANGALORE is a brain bank for India. It is the nearest equivalent to California's "Silicon Valley" in the United States. In the laboratories and research institutes "chips" are as common as chappattis, and Bangalore probably has one of the highest concentrations of scientific and technological ability across the board in the sub-continent. The town is the home of many engineering and electronic establishments, as well as basic scientific research institutes. These include the Indian Space and Research Organisation, the Indian Institute of Science, Hindustan Aeronautics, Hindustan Machine Tools, Bharat Electronics, and the Raman Research Institute, to name but a few.

Progress check: automatic thermosonic bonding of silicon chips at Bharat Electronics Ltd. The machine can process 1,200 chips a day—three times the manual level.

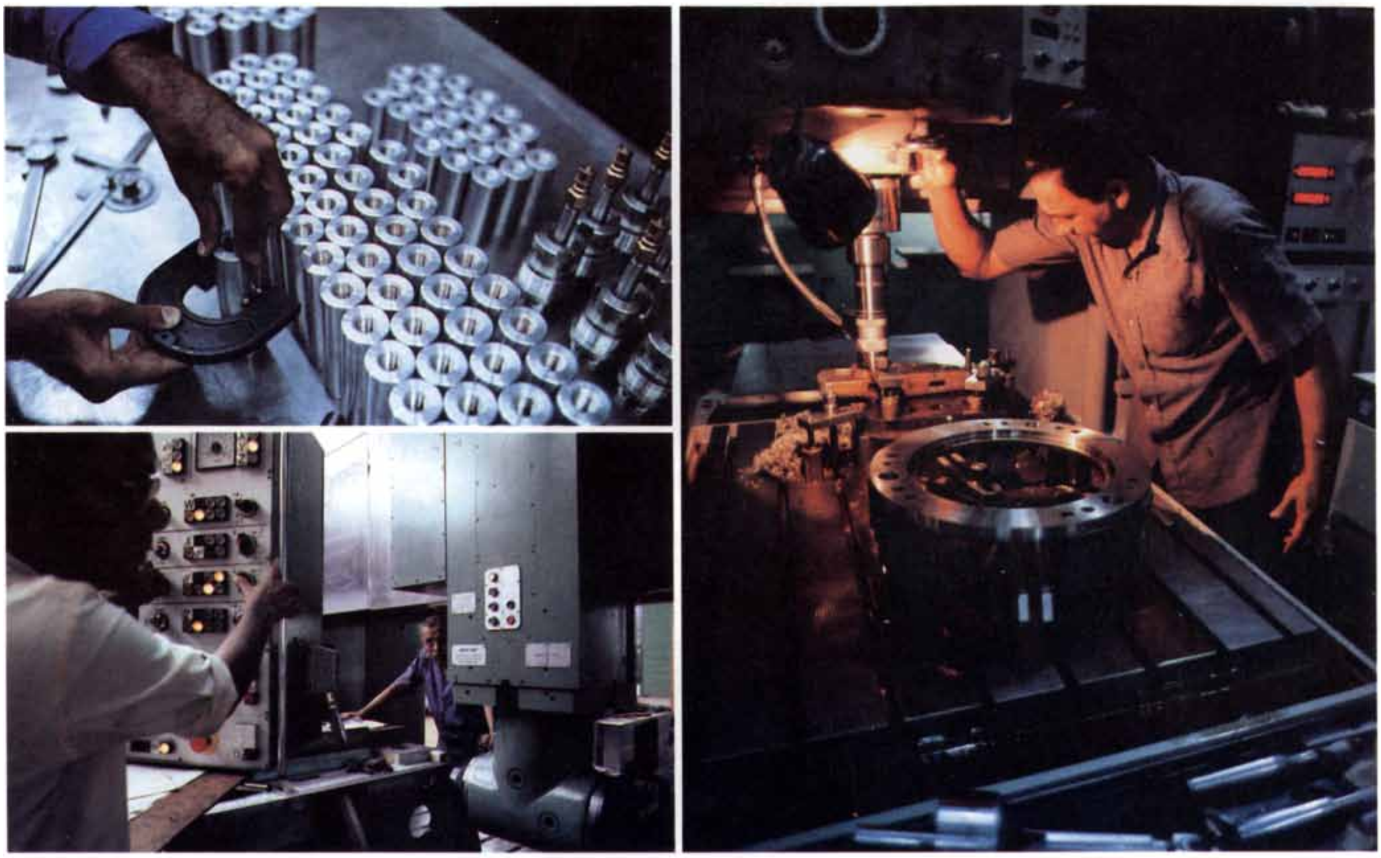
Part of the $100 million petrochemical complex at Baroda which has been designed and built by Indian engineers.

Formerly a garrison town under the British, Bangalore's technological tradition stems from the establishment of Hindustan Aeronautics Ltd, commonly known as HAL, which serviced aircraft during the Second World War. Since then it has gone on to design and build its own prop, jet training, and other aircraft like the Kiran, Ajeet, and Marut. It has more than 2,000 trained design engineers and specialists among its total workforce of 40,000.

Today it is going into production of the British designed Jaguar fighters, and HAL has already been building an improved version of the Soviet MiG-21 fighter. In addition, its engine division manufactures under license a number of power plants, from the Rolls-Royce turbo-prop Dart 531 and Orpheus jet engines, to Soviet R-11 and F2S-300's. HAL also manufactures engines for rotary wing aircraft, and produces airframes for the Cheetah helicopters that it exports to France.

HAL has a substantial reservoir of technological engineering expertise and innovative skill which is probably the foremost in South Asia. As well as producing its own indigenous designs, its engineers are constantly striving to improve and modify engines, aircraft, and components which are being built under license. Its activities cover the whole range of aeronautics from avionics and engineering systems to producing inter-stage assemblies for space rockets and satellites—achievements matched by few other developing countries.

Although total self-reliance in aviation is still some way off, this is not quite so in another vital area: machine tools. In 1953, Hindustan Machine Tools (HMT) was established with Swiss collaboration. Today the company employs 27,000 and is one of the world's largest manufacturers of machine tools, exporting to more than 50 countries. It has also established joint production and other facilities in Kenya, Algeria, Sri Lanka, the Philippines, and the United States. Mr. T. V. Mansukhani, managing director, speaks about the company's latest move into manufacturing computer numerically controlled (CNC) machines: "This is really only just starting in India and there is a big potential for us as we can also provide a complete back-up."



Precision engineering; India is now a leading manufacturer of machine tools and exports to more than 50 countries. Operators at Hindustan Machine Tools Ltd. (bottom, left, and right) carrying out calibration checks and (top left) zircon fuel elements made at the Bhabha Atomic Research Center.

Following a lead from Jawaharlal Nehru, who once said, "Make watches and you can do any precision work," HMT has diversified into watchmaking and has one of the largest and most sophisticated automatic watchmaking production lines in the world. The company also produces tractors, printing machines, and injection molding equipment.

On a more philosophical note, Mr. Mansukhani adds, "We are still finding out how to put things together in this country and we need everything from the bullock to the Boeing."

Another technological leader which is fast building itself a future for tomorrow and is closely allied with almost every area of development, is Bharat Electronics Limited (BEL). This is the country's fastest growing electronics group and employs some 17,000 persons, of which 3,000 work in its Delhi plant. More than 1,000 are in R&D departments. The main emphasis has been to develop self-sufficiency in communications equipment and systems.

Commenting on R&D, Mr. N.L. Krishnan, BEL's managing director, says, "Knowledge is something which we cannot restrict, but the problem is that it outpaces our infrastructure here in India."

It is difficult to adjust to the different levels of technology in a country as vast and diversified as India. Ion implantation and advanced production of integrated circuits at BEL are in sharp contrast to the humble bullock cart which has undergone its first major technological improvement in 2,000 years or more with the fitting of rubber tires to its wooden wheels. This prolongs the life of the cart, makes it easier to draw, and causes less damage to roads—the value of which, according to Mrs. Indira Gandhi, is very difficult to explain to someone from an advanced industrialized nation like the United States. Applied rural technology, of which the bullock cart modification is typical, is one of the most important areas of research in the country.

Improved experimental biogas plant with solar collector for family use.

One of the centers for this work, Application of Science and Technology for Rural Areas (ASTRA), is at the Institute of Science in Bangalore. Here the work involves studying a whole range of problems from water-lifting windmills, different kinds of solar collectors for domestic and light industrial use, soil cement building blocks, development of lime kilns, and biogas plants.

Much of the research in these fields may at first appear crude in relation to other, more glamorous, high-tech areas, but they are just as important, and sometimes more immediately so, when it comes to bringing improvements to the village communities. Developing practical biogas plants, which produce methane gas for cooking purposes by fermenting cow dung in an underground tank, can be of significant practical value in areas where there is little or no wood for fuel. Making improvements to clay cooking stoves or chullahs so that they are more economical in fuel consumption is just as meaningful to the farmer's wife as developing more energy-efficient car engines in Europe or the United States.

The atomic age: taming the tiger in search of cheap power

WHEN MOST COUNTRIES saw atomic power primarily as a force for destruction during the Second World War, India realized that it could be a valuable source of energy. It was one of the first countries to think in these terms, under the lead taken by its most eminent scientist, Dr. Homi Bhabha, the father of India's nuclear program. Today, this is one of the most advanced and sophisticated areas of science and technology. It covers the whole nuclear cycle from the mining and processing of uranium and manufacture of heavy water, the construction of all reactor parts themselves, to the production of fuel elements and the reprocessing of spent fuel. Now fast-breeder reactors are being built and thorium fusion studied.

From the beginning in 1944, India has developed a completely indigenous nuclear industry for peaceful purposes with a wide range of associated research activities, from medicine to agriculture, and a well-established program of radio-isotope production. Studies are also being carried out on irradiation treatment of vegetables to prolong their active storage life, cryogenics, magneto hydropower techniques, and cosmic ray research. As in other branches of science, the emphasis on self-reliance remains the key issue.

There are now four research reactors in operation, and a 100 MW high-flux natural uranium fueled reactor, completely built by Indian technology, is due to come on stream next year at the Bhabha Atomic Research Center (BARC). Also being studied are programs for the development of a medium energy heavy ion accelerator, laser induced fusion and related high density plasmas, and a 500 MW thermal reactor using heavy water. Currently the country has four nuclear power reactors in operation. Two of them are 220 MW boiling water type reactors using slightly enriched uranium at Tarapur near Bombay, which were purchased from the United States in 1964, and two more 220 MW stations at Rajasthan, one of which was imported from Canada. The second one has been built almost completely indigenously. Two more nuclear power stations of twin 235 MW capacity are under various states of construction at Kalpakkam near Madras and Narora in Uttar Pradesh. Both of these are natural uranium, heavy water types. Four heavy water plants with a combined output of 300 tonnes a year are also being built.

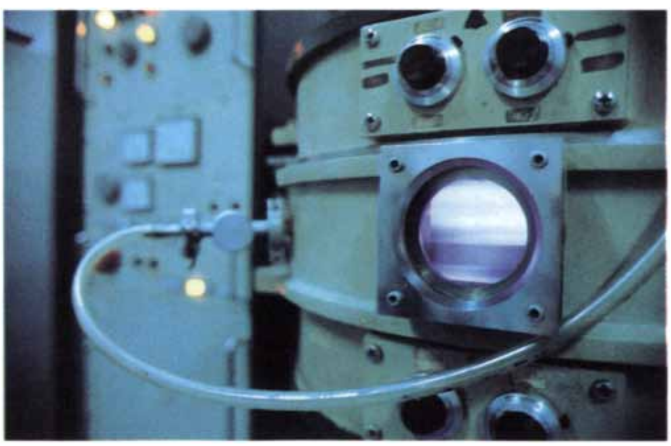
Looking back at the early days, Dr. Raja Rammana of the Atomic Energy Commission speaks of the key decision that Dr. Bhabha took to bring India into the nuclear age. "It was his vision and the need for self-reliance that has helped give us many of our scientists and technicians today. He insisted, and this was very important, that we had to do as much for ourselves as possible." The primary objective for the nuclear program had been the search for a means of producing cheaper electrical power. Dr. Ramanna explains: "A lot of work was done during the 1950's calculating the feasibility of nuclear energy as a power source.

"It was decided that nuclear power was best because the capital costs and gestation period were more advantageous than hydropower and we could build the stations more or less anywhere. We did not have to worry about coal transportation to conventional thermal stations from coalfields distant from the areas needing the power."

The nuclear power stations have not been without their problems but, as Dr. Ramanna explains: "It was all part of the learning process." The most exciting areas of the nuclear program are in research, which is centered on BARC, the Tata Institute of Fundamental Research, and the Tata Cancer Hospital in Bombay. There are a number of other nuclear R&D institutions elsewhere in the country, including the Saha Institute of Nuclear Physics at Calcutta and the Reactor Research Center at Kalapakkam.

The leading institution is the huge garden complex, BARC, at Trombay just outside Bombay. The BARC is part of the Atomic Energy Commission, whose current chairman is Dr. H. N. Sethna. Nuclear studies began at the Tata Institute for Fundamental Research in 1945, but they soon outgrew the available facilities there. Major work was transferred to the newly formed Atomic Energy Establishment at Trombay in 1954, which was later renamed Bhabha Atomic Research Center—in memory of its founder—in

Copper coating zircalloy (left) by the "sputtering" technique at the Bhabha Atomic Research Center. One of the many fields of activity is into prolonging storage life of vegetables by exposing them to radiation from cobalt 60.



1967. More than 2,500 nuclear scientists and technicians have been trained at BARC since 1957.

With its four working reactors and a fifth nearing completion, BARC has the major responsibility for developing the infrastructure for nuclear technology. The latest R-5 100 MW high flux natural uranium reactor now nearing completion will be used for isotope production and for the development of improved power reactor technology.

Basic research work at BARC covers all major disciplines. These include reactor engineering, radiological protection, instrumentation, radiation medicine, isotope production for various uses, agriculture, metallurgy, manufacture of zircon fuel elements, and reprocessing.

Spent fuel reprocessing began in 1964, making India only the fifth country in the world to undertake this operation. The plant, which initially handled 30 tonnes of fuel a year, was designed and built by BARC technicians. Based on experience from this pilot project, a 100 tonne per year plant has been established at Tarapur to handle spent fuel from the power station there, and another similarly sized processing works is going ahead at Kalpakkam. This will have a separate stream for reprocessing fuel from the fast breeder reactor being built.

When the fast breeder becomes active it will be producing more plutonium as well as Uranium 233 using thorium; the third phase would be Uranium 233-thorium as fuel. India has substantial uranium resources estimated at well over 53,000 tonnes and its thorium deposits total more than 363,000 tonnes—the largest known deposit in the world. Because of these enormous fuel resources India does not have to rely on outside supplies.

The fast breeder reactor, according to Dr. Ramanna, will be the most significant step forward in India's nuclear program. "Nothing can be more exciting for us than these breeder reactors, it will be like taming the tiger," says Dr. Ramanna.

Industry forges ahead

INDIA'S ADVANCES in agriculture are the result, not only of improvements in seed varieties, but also of extensive investment in irrigation, flood control, better use of fertilizers and insecticides, as well as more adequate extension services to the rural areas.

All these factors have helped contribute to the country's self-sufficiency in grains, which now total about 135 million tonnes a year. Progress has also been achieved in improving the production of coarse grains like ragi, bajra, jawar, pulses, and oil seeds, as well as fruits and vegetables.

In the last 30 years rice production has increased from nearly 23 million tonnes to 54 million tonnes a year, wheat from 6.6 million tonnes to 36 million tonnes, and sugar cane from 56 million tonnes to 176 million tonnes, which makes India one of the largest cane producers in the world.

The breakthrough on crop production and better yields is the result of systematic efforts by scientists working at the many research institutes and universities. They have developed new methods of plant protection, breeding resistant varieties of crops, and producing more efficient insecticides.

Microbiology and genetic engineering in plants are key factors in the struggle to improve agricultural output. Much of this research is going on at the Indian Agricultural Research Institute (IARI) just outside New Delhi. There, the scientists have made wheat grow shorter and fuller; they try to determine the sex of the papaya seedlings to prevent wastage, as only the female has fruit, and breed new mutations of valuable cash crops. Triple dwarf varieties of wheat can now give a harvest of 6 tonnes per hectare compared to only 2 to 3 tonnes before.

Problems with importing or manufacturing chemical fertilizers (although facilities have been greatly extended) are still a major constraint on improving agricultural yields. For these reasons an intensive program into developing nitrogen-fixation plants sponsored by the Department of Science and Technology is underway at one of the IARI laboratories. Blue-green algae are being examined as possible alternative sources of nitrogen for rice-growing areas.

Apart from improved food production, the other great problem facing the country is health and family planning. More research has been going on in these two areas than in any other. Professor V. Ramalingaswami, director-general of the Indian Council of Medical Research (ICMR) in New Delhi, says these problems present a challenge for high science and technology to develop their own solutions.

The substantial expansion of medical services and improvements in infant nutrition and major reductions in communicable diseases have already increased the average life span by about 20 years over the last quarter century. Similarly the fight against malaria (once there were 75 million cases a year, now there are less than 100,000), the eradication of smallpox, and the virtual elimination of cholera, have all contributed to greater life expectancy. Half a million people still die annually from TB, an age-old scourge, but new methods of domicillary treatment and chemotherapy have been particularly effective.

There is also a new sense of hope for treating the three million cases of leprosy in the country, about one fifth of which are infectious, and much fieldwork is being done on anti-leprosy vaccines.

One of the major problems involving long treatments for various diseases, often for years, is getting patients to continue taking drugs. The ICMR has developed sub-dermal drug implants (in the soft upper arm) which can be effective for several years. It is also possible to implant biodegradable capsules.

New advances in sub-dermal implants of drugs at the Indian Council of Medical Research. A patient is about to have a drug capsule inserted into her arm.

FAMILY PLANNING: this is one of the crucial issues and great efforts are being put into researching new contraceptive measures, including implants to overcome the problem of forgetting to take the "pill." Another innovation is to take contraceptive drugs by inhalation through the nose. If more family planning is not introduced the present population of 660 million will have doubled in the next 37 years.

The space age: eye in the sky keeps watch on monsoons

FOR THOUSANDS of years the Indian people have marked the seasons for planting and harvesting by the stars. That tradition continues today—but with the use of satellites and rockets, many of which are built and launched in India. Today's space program, which began only eighteen years ago, is the most advanced of any developing country and may at first seem somewhat incongruous against a background of India's teeming millions, of whom 48 percent are subsisting below the official poverty line. But the point of the program, as Professor U.R. Rao, director of the Indian Space Research Organization (ISRO) at Bangalore explains, is to bring the maximum benefit to the maximum number of people in the shortest time. One of the alternative methods is to use satellites.

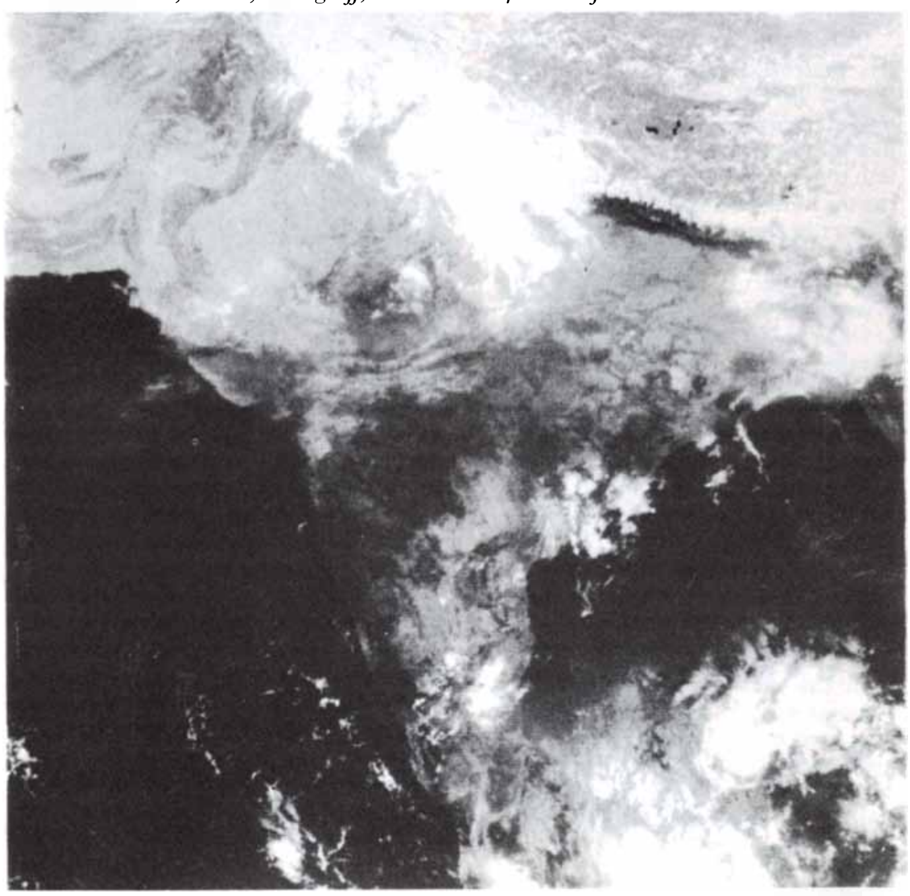
Photographs of the sub-continent showing the path of the monsoon as well as normal weather factors are already available on an increasing scale. Other satellite pictures will enable agro-economists to make more accurate predictions. But one of the major aspects will be an almost immediate improvement in communications, particularly by telephone.

A major problem with the conventional land line system or micro-wave links for radio and telephones is that they are subject to constant damage from natural hazards and disasters in the sub-continent. So one of the current philosophies on communications is to bounce signals into space and back, rather than trying to go over the land surface by more conventional methods. This will mean first class communications quality with 100 percent reliability during the life of the satellites.

"With the present experimental APPLE geostationary satellite we are almost ready to start our first telephone network using 30 ground stations which have already been built," says Professor Rao, who is a staunch advocate of the space program. "We have to have satellites, there is no other way for us," he adds.

Already some experimental services for television have been run but, according to Professor Rao, bringing television via satellite to the whole nation will take a long time. "After all, we have 560,000 villages to reach and they must have electricity first, so we will have to gradually phase in village television," he explains. According to Professor Menon the thrust of the space program has been to develop the necessary skills and capabilities to build both the satellites and the launch vehicles. In the last six years considerable progress has been made, and the space program has maximum priority over the next eight years. More than a thousand solid-fuel rockets have been launched from the two

India's first own designed and built satellite launcher, SLV-3, taking off; and satellite picture of the sub-continent.



ranges at Thumba near Trivandrum and Sriharikota, an island off the southern coast of Andhra Pradesh. These have been used for scientific, meteorological, and technical studies.

Sriharikota or SHAR, which means an "arrow", is the most extensive space research station in India, covering some 33,000 acres on the island where 1,500 people live and work on the space program. SHAR has facilities for launching multi-stage rockets and for statics tests of launch vehicle motors.

So far six satellites have been designed, built indigenously, and successfully launched into space. Two have been launched with India's own rockets, the SLV-3.

The first into space in 1975 was Aryabhata, which was launched by the Soviet Union to test satellite technology and instrumentation. Two others, Bhaskara 1 (1979) and Bhaskara II (1981) were also launched by the Soviet Union. Bhaskara I carried a micro-wave radio meter and television camera system for earth observation. Bhaskara II was a slightly improved version of the first and is also used for earth observations. It weighs 436 kg. and its Samir radio meters are able to analyze data on water vapor content over the earth's surface. Its two television cameras are able to take pictures covering an area 341 km. x 341 km. at a time with resolution down to 1 km. These pictures can show the extent of water on the land area, geological formations, as well as forest coverage.

Another important function of the satellite, which orbits the earth at a height of 525 kms. every 92.5 minutes, is to test thermal paint and solar cells under prolonged exposure in a space environment. Both these items have been developed indigenously. All the satellite's mainframe systems have been designed and produced at the ISRO center in Bangalore, which was also primarily responsible for the mission planning and in-orbit operation. Many public and private sector companies and organizations have been assisting.

Other space application experiments have been made using the United States NASA ATS-6 satellite in 1975–76 and the French German Symphonie in 1977–79. The ATS-6 was used for a Satellite Instructional Television Experiment (SITE) which formed a vital test run for what will eventually become a major application of satellite television in the future. During the experiment, special television programs were beamed direct to 2,400 villages spread throughout six states. Low cost dish aerials picked up the signals as the satellite passed overhead.

Another similar satellite, but designed by Indians and built by Ford Aerospace and Communication Corporation, INSAT-1, has recently been launched by NASA but has been having some teeth problems. It is intended to be part of a multipurpose operational satellite system for India. Apart from television and radio transmissions it is hoped to be able to broadcast early warnings of monsoons and floods to rural areas. By the end of the decade India hopes to replace all foreign built INSAT spacecraft with satellites of its own design and construction.

A solid-state satellite camera undergoing thermal vacuum and caliberation tests in a space simulation chamber at the ISRO center, Bangalore.

Efficient communications are essential not only for disaster warnings but also for educational purposes generally in rural areas. A working rural community television network would be a considerable breakthrough which would have immediate benefits from an educational point of view for both children and farmers, as well as spreading information about family planning. The satellite center at Bangalore is the main research and fabrication plant for the space industry. It is there that most of the indigenously designed satellites have been built. These include the RS-1, the first in the Rohini series—a small 38 kg. satellite used for technical and monitoring purposes.

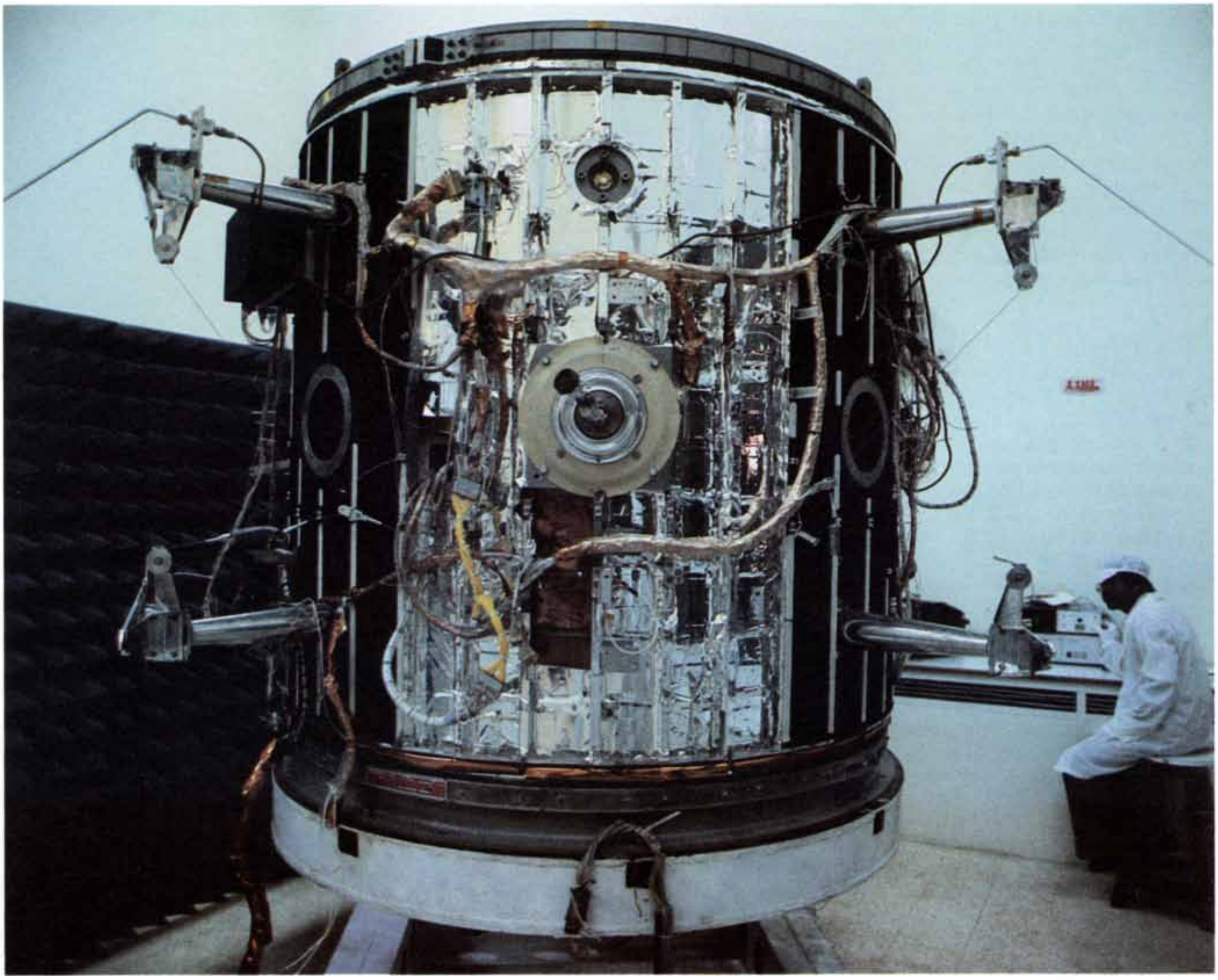
The most dramatic development in the space program was when the Rohini satellite was launched into space on July 18, 1980 on the SLV-3 launch vehicle which had also been designed and constructed by ISRO—the first wholly Indian-made rocket and satellite launch. The SLV-3 was launched from SHAR and its successful firing was a tribute to the skill of Indian scientists and technicians who depended solely on themselves for every stage of the construction and launching. The rocket weighed 17 tonnes of which 13 tonnes were solid propellant. The four-stage SLV-3 is similar to the American Scout rocket and has a total thrust of 176,000 lb. The first two stages of the rocket had metal casings and the last two were made of fiberglass. The main role of the satellite itself was to monitor the firing and operation of the small liquid fuel control rockets used in the fourth stage.

There were 10,000 major components and 44 sub-systems in the SLV-3 and if any single item had failed the mission would not have succeeded. This shows the high standard of skill and innovation that exists in the space industry.

More powerful launch vehicles are currently being developed. This will include the Polar Satellite Launch Vehicle (PSLV), which will be able to carry 1,000 kg. satellites into sun-synchronous orbits. Another is the smaller Augmented Satellite Launch Vehicle (ASLV) which will carry 150 kg. satellites, basically for research purposes.

One of the latest satellites launched last year was the Ariane Passenger Payload Experiment (APPLE), which was India's first three axis body stabilized satellite to be placed in geostationary orbit for communication experiments. The 670 kg. satellite was lifted aboard the European Space Agency's (ESA) ARIANE Launch vehicle last July together with an ESA meteorological satellite and test capsule.

The significance of its launching was that after APPLE had been placed into an intermediate elliptical orbit it was boosted into its correct position by its apogee rocket motor, which had been developed as the fourth stage of India's own SLV-3 launch vehicle. It has been placed into stationary orbit 22,000



Stripped-down APPLE satellite in the ISRO center "clean" room at Bangalore.

miles above the island of Sumatra with its antennae directed towards India.

Placing APPLE into position has given India's space technicians valuable experience for handling, station keeping, and post-launch operations. The design and development of APPLE itself involved building five full scale models in order to test all main function, engineering, and flight systems. The spacecraft's main frame systems such as structure, thermal, power, electronics, telemetry, telecommand and tracking systems were all developed at Bangalore. The apogee boost motor, deployment mechanism and antennae were made at the Vikram Sarabhai Space Center in Trivandrum. Other parts and components were developed and made at the Space Applications Center at Ahmedabad and other research institutes.

APPLE embodies some of the latest satellite technologies and is an important milestone in the space program. Together with the Rohini I and Bhaskara II, it has shown that India can be in the forefront of space technology in spite of its lack of resources compared with the Western industrialized nations.

The priorities for the rest of the decade are to accelerate the indigenization of the INSAT space craft for communications and other purposes, and to develop a remote sensing satellite system tailored to the country's own needs in land usage.

The space program is only one part of the transitional development which has taken place in the country since independence. India, with her vast resource of technical manpower, and her ancient scientific tradition, is poised for greater growth in science and technology. She has a leadership, both political and scientific, that recognizes with vision, the challenges of tomorrow. The world will hear more of India's continuous achievements and progress in the future.

This special section has been sponsored by the following:

Department of Science and Technology
Council of Scientific and Industrial Research
Department of Atomic Energy
Department of Space
Indian Council of Medical Research
Indian Council of Agricultural Research
Department of Electronics
Ministry of Defense Production
External Publicity Division of the Ministry of External Affairs, Government of India

---

Main photographs by Graham Finlayson; picture of Indira Gandhi by Madan Arora; others provided by Government of India.


dimension

Allergy

It can now be understood as immunity gone wrong. Why hay-fever victims make antibody to a normally innocuous pollen is not clear, but the cellular and biochemical agents of the response are known

by Paul D. Buisseret

It is best to begin a discussion of allergy by defining the word and establishing the scope of its meaning. "Allergy" is from the Greek *allos* (other) and *ergon* (work). The derivation implies an unusual or inappropriate reaction to a stimulus. Not every adverse response to a normally harmless external stimulus is an allergy, however, and the distinction is not always clear. One person may develop an itchy rash after an injection of penicillin, another whenever he wears a certain wristwatch. Hay-fever victims sneeze in certain seasons. Some women are nauseated by oral contraceptives. A person may have abdominal cramps and diarrhea after eating a certain food. Are all these people "allergic"? Some are indeed, but some may not be. For example, many people have an unusual or exaggerated response to a drug, but they are not necessarily allergic to the drug.

Allergy is a disorder of the immune system; it is immunity gone wrong. In the immunological world of a vertebrate organism the great divide is between the self and the nonself. The organism distinguishes nonself tissues, cells, proteins and some other large molecules from those of the self; it identifies them as being either harmless or potentially harmful and thereupon either ignores them or defends the organism against them. The cells with the capacity for recognizing, evaluating and neutralizing or eliminating nonself material constitute the immune system. The first step is recognition of a specific antigen, which is accomplished by the white blood cells called lymphocytes. They are abundant not only in the blood but also in the lymphatic system, the lungs, the liver and the spleen. There are two kinds of lymphocytes: *T* cells and *B* cells. Each clone, or genetically identical line, of *T* and *B* cells is preprogrammed to recognize a highly specific combination and conformation of atoms, called an epitope, on a nonself molecule. The cells do so by displaying on their surface specific receptors shaped to bind one particular epitope. A substance carrying enough recognizable epitopes to evoke an immune response is called an antigen.

The type of immune response depends on the nature of the antigen. Some antigens characteristically stimulate a response that is mainly cellular. Such antigens are "presented" to *T* cells by the large specialized cells called macrophages. The antigens bind to receptors on the *T* cells, causing the cells to secrete the complex substances called lymphokines. At the site of an antigenic challenge certain lymphokines cause small blood vessels in the vicinity to become dilated and carry more blood, so that the tissue gets redder and warmer. Other lymphokines make the vessels leaky, so that fluid and blood cells move into the surrounding tissue and cause swelling. The classic triad of the inflammatory reaction is redness, warmth and swelling. Still other lymphokines activate macrophages and cause them to congregate at the site of inflammation. One function of macrophages is phagocytosis: they engulf the antigen, neutralizing it or destroying it enzymatically.

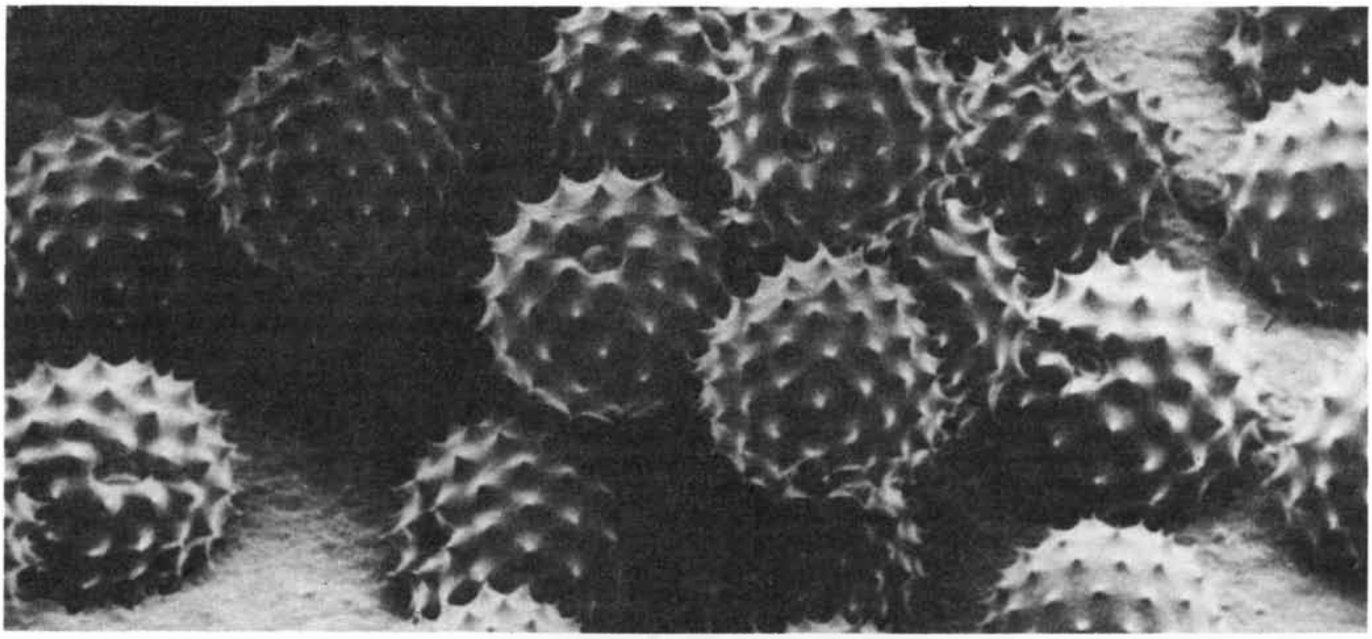
Under normal circumstances the *T* cells are able to evaluate the threat posed by an antigen: to distinguish between harmful foreign substances (such as pathogenic bacteria) and harmless ones. If the recognition system fails, a harmless substance may be mistaken for a harmful one. It acts as an antigen, provoking the cellular response described above. That is what happens, for example, when a person develops a contact dermatitis—a rash or other inflammation of the skin—after contact with a normally innocuous material such as a cosmetic, a necklace or the leaf of a plant. Because this type of allergic reaction takes about 48 hours to develop it is commonly referred to as delayed hypersensitivity.

Immediate Hypersensitivity

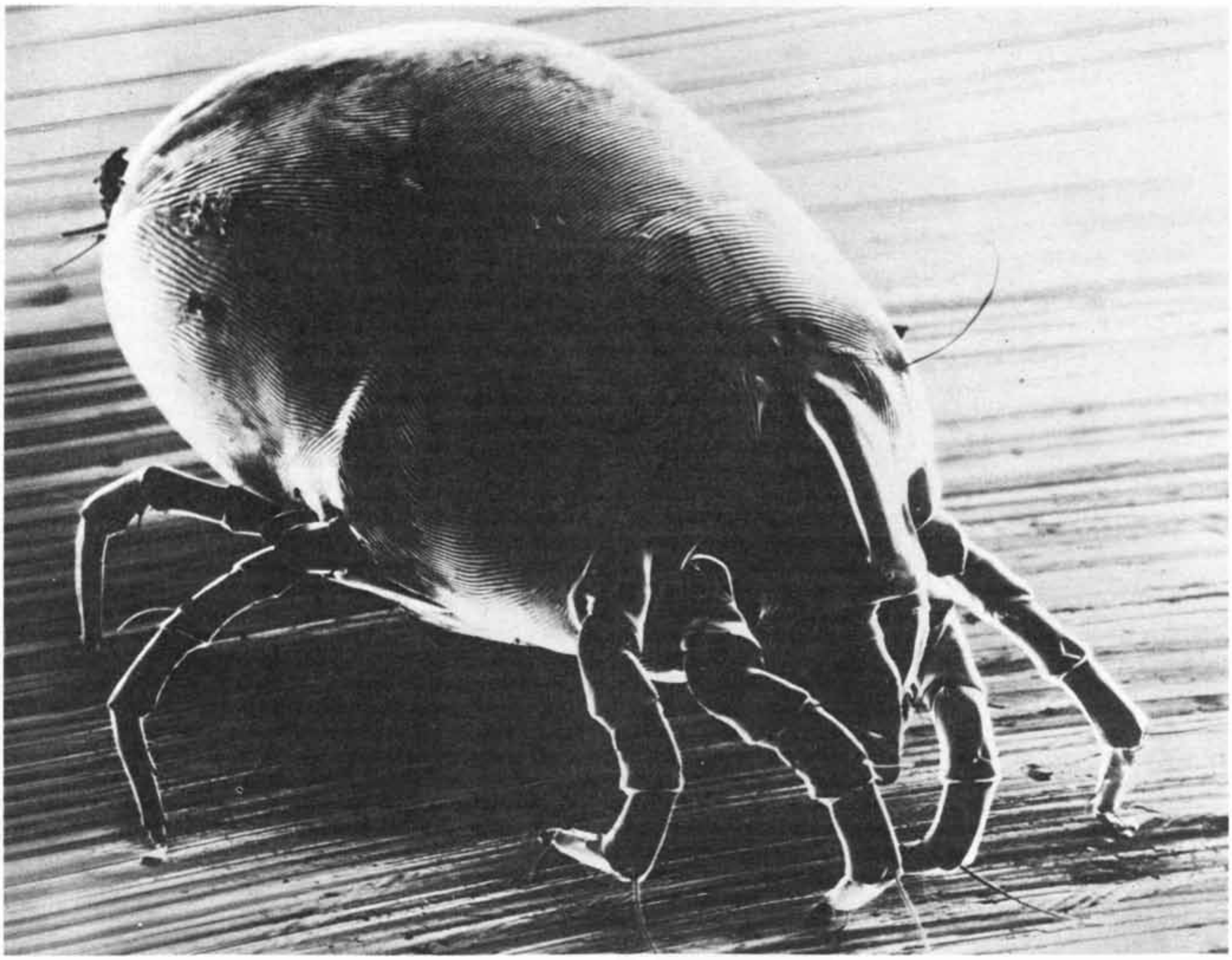
There is another kind of allergic reaction that develops more explosively, within seconds or minutes after an encounter with an antigen. It is called an immediate-hypersensitivity or anaphylactic response, and it is mediated not by *T* cells and macrophages but by antibodies, which are made by *B* lymphocytes. The receptor displayed on the surface of an immature, quiescent *B* cell is a sample of the specific antibody it is prepared to synthesize in quantity. When an antigen binds to such a receptor, it triggers a transformation of the *B* cell. The cell proliferates and differentiates to form a clone of plasma cells, in which the protein-synthesizing structure called the endoplasmic reticulum is greatly enlarged. The plasma cells synthesize and secrete into the bloodstream millions of identical antibody molecules, thus generating a humoral immune response (as opposed to the cellular immune response).

Again proper recognition of the threat posed by an antigen is crucial. With every breath of air or mouthful of food a person takes in millions of potentially antigenic nonself molecules; the uncontrolled transformation of *B* cells into plasma cells in response to this onslaught could keep the body in a permanent state of immunological uproar, with antibodies being made against many harmless or even beneficial molecules. Normally that does not happen because the transformation of *B* cells is controlled by subclasses of *T* lymphocytes. "Helper" *T* cells prompt antigen-triggered *B* cells to transform when the synthesis of antibody is appropriate, and "suppressor" *T* cells inhibit inappropriate antibody production. It seems likely that the failure of this regulatory system is what leads to an allergic response. Either the suppressor cells fail to act or their chemical message does not get through; *B* lymphocytes that have bound an innocuous antigen are stimulated to become plasma cells and secrete antibody.

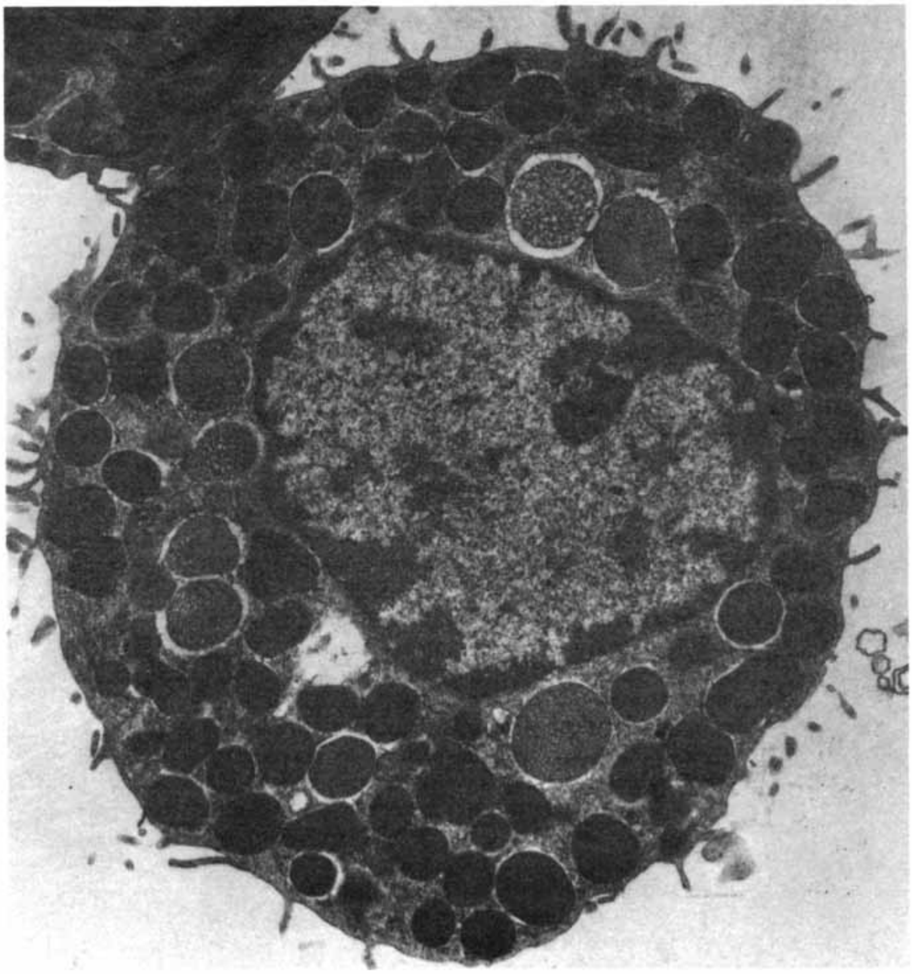
There are five classes of antibody molecules: the immunoglobulins IgM, IgD, IgG (of which there are several subclasses), IgE and IgA. Each immunoglobulin has a particular role in the body's defense against infection or in some other immunological task. In most people IgE seems to have a limited immunological role: it helps to repel inva-



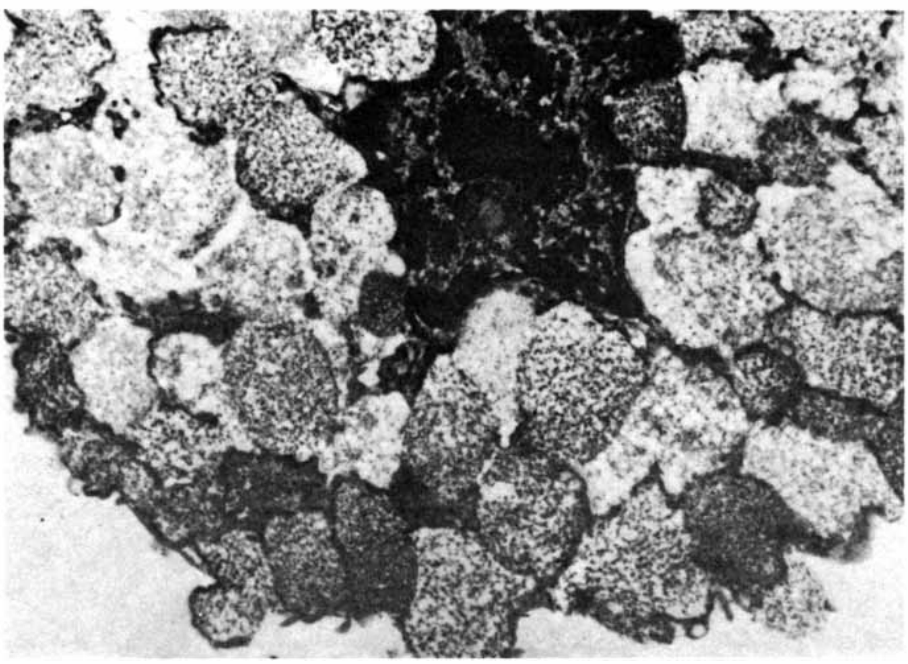
RAGWEED POLLEN is a notorious allergen, that is, an ordinarily innocuous substance to which many susceptible individuals nonetheless develop an inappropriate immune response. In the case of ragweed pollen and most other airborne allergens it is an "immediate hypersensitivity" response mediated by antibody of the class designated IgE and by connective-tissue cells called mast cells. Pollen grains are enlarged 1,850 diameters in this scanning electron micrograph made by David J. Lim of Ohio State University College of Medicine.



HOUSE-DUST MITE is responsible for much year-round allergic rhinitis and asthma. It lives in the dust in dwellings, and particularly in the dust of mattresses and pillows, where the human skin scales on which it subsists are shed in quantity. This is the mite *Dermatophagoides farinae,* enlarged 500 diameters in a scanning electron micrograph made in the Acarology Laboratory of Ohio State University.



MAST CELL is the source of histamine and certain other chemical mediators of the immediate-hypersensitivity response. A thin section of a rat mast cell is enlarged about 7,000 diameters in this electron micrograph made by Ernesto O. Hoffmann of the Louisiana State University School of Medicine in New Orleans. The cell is crammed with perhaps 1,000 dense granules containing the mediators. An allergic person makes antibodies against an allergen; they bind to receptors on the surface of mast cells. When the allergen is next encountered, it binds to the antibodies, triggering the release of the mediators and the ensuing allergic reaction.



DEGRANULATING MAST CELL is seen in this micrograph made by Hoffmann. Some granules have fused with cell membrane and discharged their contents. Others seem to have fused with adjacent granules and to have passed their contents along toward cell membrane.

sion by parasitic worms. Elevated IgE levels are therefore often found in the blood of people living in tropical regions where infestation with worms is common. Elsewhere there appears to be no adaptive, or "good," function for IgE, and most people synthesize very little of it; IgE antibodies account for only some .001 percent of all the antibody in a normal person's blood.

The situation is different in an allergic person. It is IgE that mediates immediate hypersensitivity. In the absence of proper modulation by *T* cells a clone of *B* lymphocytes is transformed by the binding of an antigen (an "allergen") that would not ordinarily excite an immune response but to which the allergic person is about to become sensitive. The plasma cells make antibodies against the allergen—preferentially antibodies of the IgE class, for reasons that are still not known. Allergies are specific because the IgE is specific. A person is allergic to cats because he makes IgE in response to the binding of a particular protein of cat hair or skin; he may tolerate other animals perfectly well. A person allergic to oysters makes IgE that recognizes and interacts with a protein in oysters; he may be able to eat any other food with impunity. Some unfortunate individuals, however, have such a tendency to develop allergies that they make IgE against more than one allergen. I shall have more to say about what makes people "atopic," or generally allergic, and why some people have more allergies than others.

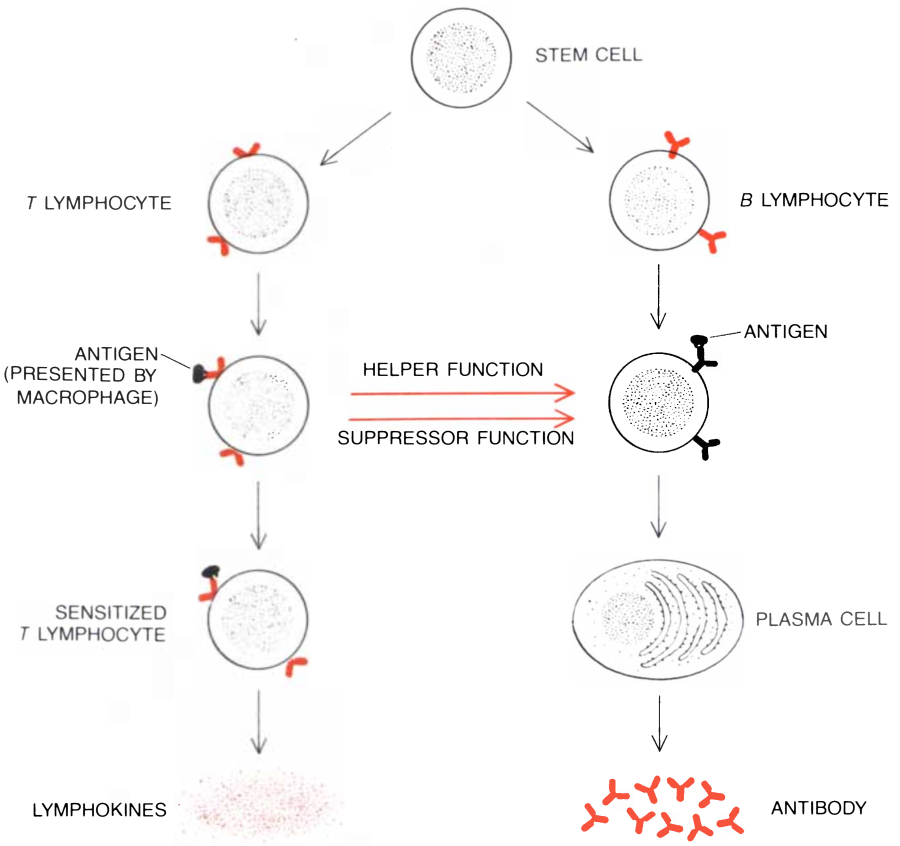
The Allergic Reaction

IgE does not mediate the allergic reaction alone; what it does is prime an effector cell. In 1877 the German bacteriologist Paul Ehrlich described a curious connective-tissue cell seen most abundantly in the skin, in lymphoid organs, in membranes of the eye, nose and mouth and in the respiratory tree and intestines. He noted that the cells appeared to be stuffed with large granules, which stained brilliantly when the cells were treated with certain dyes. It was assumed that the granules had been engulfed by the cells, which were therefore designated mast cells, from the German *mast* (a fattening feed). The name stuck, even though the granules in fact are probably not engulfed but rather produced within the cells. There are about 1,000 of them in each mast cell; they are vesicles filled with the chemical mediators of the allergic reaction.

The surface of a mast cell is studded with between 100,000 and 500,000 receptors for IgE molecules. When specific IgE antibodies are synthesized by the millions in response to the binding of an allergen, they move through the bloodstream to mast cells in connective tissue and become firmly fixed to the receptors. As yet, however, there is no allergic reaction. That comes only when the allergic individual next encounters the same antigen. This time the antigen need not go through the process of triggering *B* cells to transform. It goes straight to the IgE fixed to mast cells and binds to the antibody. As a result the mast cell degranulates: the granules move to the surface of the cell and release their contents into the surrounding tissue.

The sequence of events that leads to degranulation has been worked out largely by Michael A. Kaliner of the National Institute of Allergy and Infectious Diseases, his colleague Robert P. Orange, K. Frank Austen of the Harvard Medical School and Teruko Ishizaka of the Johns Hopkins University School of Medicine. Two adjacent IgE molecules on the surface of a mast cell are bridged by two epitopes on an antigen molecule. The bridging somehow activates a molecule that is a precursor of a protein-digesting enzyme, serine esterase; when the precursor, serine proesterase, is activated, it is converted into the enzyme. Serine esterase initiates a chain of reactions that result in the generation of phosphatidyl choline.

The sequence of events somehow modifies the structure of the mast cell's outer membrane. The membrane becomes permeable to calcium ions. Calcium enters the cell and activates the enzyme phospholipase A_2, which promotes the further metabolism of phosphatidyl choline to form lysophosphatidyl choline and arachidonic acid. The entry of calcium ions also activates enzymes that release energy and promotes the assembly in the cell of cablelike microtubules and the contraction of threadlike microfilaments. These contractile structures cause the granules to move to the cell membrane, fuse with it and disgorge their contents: histamine, heparin, serotonin and chemical factors that activate blood platelets and attract the white blood cells called eosinophils and also phagocytic cells. Each of these mediators contributes in its own way to the allergic reaction.



IMMUNE RESPONSE is mediated by two kinds of lymphocytes, both of which are descended from primitive stem cells in the bone marrow. *T* lymphocytes effect cellular responses, including delayed-hypersensitivity allergic reactions; *B* lymphocytes make antibodies, which effect humoral responses, including immediate-hypersensitivity allergic reactions. Both classes of lymphocytes have receptors on their surface that can recognize a specific epitope, or antigenic determinant, on a nonself molecule. Some *T* cells, sensitized by contact with an antigen, liberate lymphokines, which give rise to various inflammatory reactions. A *B* cell's receptors are samples of the antibody it can synthesize. Contact with a specific antigen can trigger the transformation of a *B* lymphocyte into an antibody-secreting plasma cell. The transformation is regulated by subclasses of *T* lymphocytes that function as either "helpers" or "suppressors."

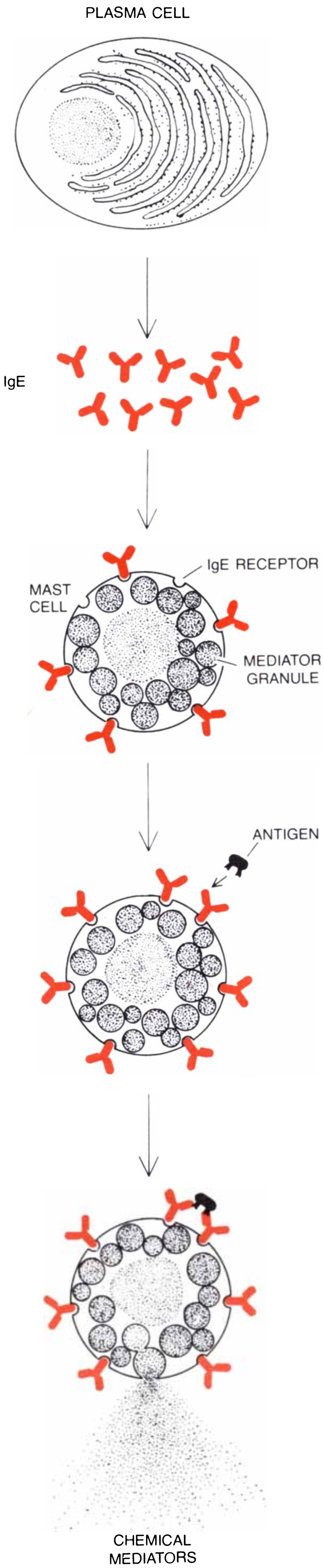
Prostaglandins and Leukotrienes

In addition to the preformed mediators stored in mast-cell granules there are two important groups of potent biological mediators that are synthesized not only in mast cells but also in several kinds of leukocytes (white blood cells). They are the prostaglandins and the leukotrienes. The prostaglandins are a family of fatty acids whose importance as cellular messengers and as agents that affect various physiological activities has been known for some time.

The leukotrienes have only recently been characterized chemically, but their role in allergic reactions was noted several decades ago. In the late 1930's a substance was found in the extracellular fluid of the lung that was clearly not one of the mediators liberated from mast-cell granules but that caused a slow, long-lasting and profound constriction of the airways in experimental animals. It became known as the slow-reacting substance of anaphylaxis, or *SRS-A*. Within the past three years *SRS-A* has been shown to consist of a mixture of three substances with an unusual chemistry. They are thioethers: fatty acids linked by a sulfur atom to one or more amino acids. Their structure was worked out by Bengt Samuelsson of the Karolinska Institute in Stockholm, who called them leukotrienes because they are made by leukocytes and have three conjugated double bonds (double bonds separated by single bonds) in their parent molecule.

The prostaglandins and the leukotrienes share a common origin. They are derived, by two different enzymatic pathways, from arachidonic acid, which is formed in cells whose cell membrane has been disrupted. The enzyme phospholipase breaks down the phospholipid molecules (such as phosphatidyl choline) of the disrupted cell membrane to form arachidonic acid. There ensues a two-branched "arachidonic acid cascade." The prostaglandin branch is initiated when the enzyme cyclooxygenase converts arachidonic acid into two cyclic endoperoxides, the prostaglandins G_2 and H_2. These are modified by enzymes called synthetases to make the biologically active prostaglandins D_2, E_2, $F_{2\alpha}$, I_2 and thromboxane A_2. Prostaglandins $F_{2\alpha}$ and thromboxane A_2 are potent but short-lived constrictors of smooth muscle in the bronchi of the lungs. Prostaglandin E_2 has the opposite effect: it dilates the bronchi. Members of the prostaglandin family also affect the activity of mucous glands and the viscidity of their secretions, the stickiness of blood platelets and the nonstick property of the lining of blood vessels.

Another enzyme, lipoxygenase, initiates the leukotriene branch by converting arachidonic acid into 5-hydroperoxy-eicosatetraenoic acid, which is modified to become leukotriene A_4. (The subscript indicates that there are four double bonds in all in each leukotriene molecule.) Leukotriene A_4 is converted by the addition of water into leukotriene B_4 or, by the addition of glutathione, into leukotriene C_4. Glutathione is a tripeptide: a chain of the amino acids



cysteine, glycine and glutamic acid. The loss of glutamic acid converts leukotriene C_4 into leukotriene D_4; the loss of glycine from D_4 forms leukotriene E_4. The mixture of C_4, D_4 and E_4 constitutes *SRS-A*. The leukotrienes are from 100 to 1,000 times as potent as histamine or the prostaglandins in constricting the smallest airways of the bronchial tree.

The link between the triggering of an allergic reaction and the initiation of the arachidonic acid cascade in leukocytes is still not known. Stimuli other than the IgE-mediated allergic response can also initiate the cascade: hormonal signals, trauma and infection. Certainly the prostaglandins and the leukotrienes are important factors in many nonallergic conditions. In allergy they join with the mediators liberated from the mast-cell granules to give rise to the contraction of smooth muscle in the airways or the intestine; the dilatation of small blood vessels and an increase in their permeability to water and plasma proteins; the secretion of thick, sticky mucus, and (in the skin) the stimulation of nerve endings that results in itching or pain.

Allergens

Different people are allergic to a wide range of seemingly unrelated substances, but the list of common allergens is curiously consistent. The reason is not at all clear. Why should the pollen of one plant, ragweed, cause millions of susceptible individuals to develop hay fever when they inhale it, whereas the pollen of another plant, the nettle, rarely bothers anyone? Both pollens are abundant in the air at certain times of the year and both are easily inhaled. In a two-year period nearly 3,000 patients seen at the allergy clinic of St. Mary's Hospital in London were allergic to the pollen of common grasses; not a single patient was allergic to the pollen of the nettle or the plantain. Evidently allergenicity depends in part on some unknown physicochemical properties of the pollen itself. The same thing is apparently true of the spores of molds and fungi. Some of them are major sources of hay fever and what appear to be very similar ones are not.

The allergenicity of drugs also varies widely. Most drugs are chemically rather inert and do not readily combine with proteins in the body, and such drugs are ordinarily not allergenic. More reactive drugs (and the breakdown products of their metabolism in the body) may or may not cause an allergic response. The explanation is that drugs tend to be small molecules—too small to serve as antigens even if they are epitopes. When such "haptens" combine with a protein in the body, the resulting complex may then be antigenic or, in the case of an allergic person, allergenic. Even then there may be enough uncomplexed hapten molecules to occupy most of the antigen-combining sites on IgE, blocking the binding of drug-protein complexes and preventing the development of an allergic response. This phenomenon of hapten inhibition may be one reason for the relative rarity of allergic drug reactions. Nonallergic adverse effects of drugs are commoner.

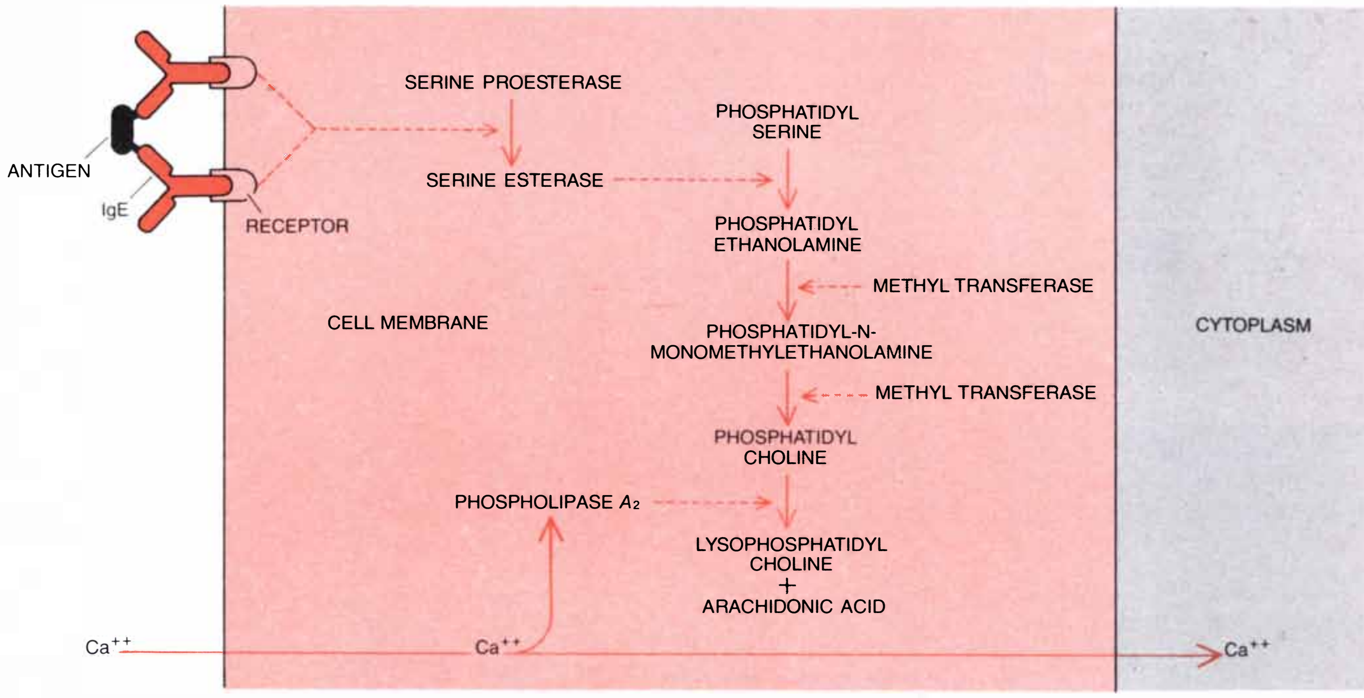
Although seasonal hay fever due to pollens or spores is the commonest manifestation of IgE-mediated allergy, a substantial proportion of allergic individuals develop similar symptoms the year round. Some people with year-round allergic rhinitis (which may progress to asthma) are allergic to a domestic pet. If there is no pet, one looks elsewhere in the home environment. Often the offender is the minute house-dust mite *Dermatophagoides,* which lives in house dust and subsists (as its name implies) on human skin scales. Everyone sheds large quantities of dead skin cells daily, particularly in the bedroom while dressing and undressing, so that the house-dust mite is commonest in the dust of mattresses, upholstery and pillows. People who develop symptoms when they sleep on a feather pillow often think they are allergic to feathers (rather than to the microfauna inhabiting the feathers).

A simple skin test can often resolve doubt about which airborne allergen causes an individual's hay fever or asthma. A single drop of an extract of the suspected allergen is pricked through the skin of the arm. If the patient has synthesized IgE against the allergen, mast cells in the skin are triggered to degranulate and within about 10 minutes an acute local allergic reaction is manifested by an itchy wheal surrounded by a flare of reddened skin.

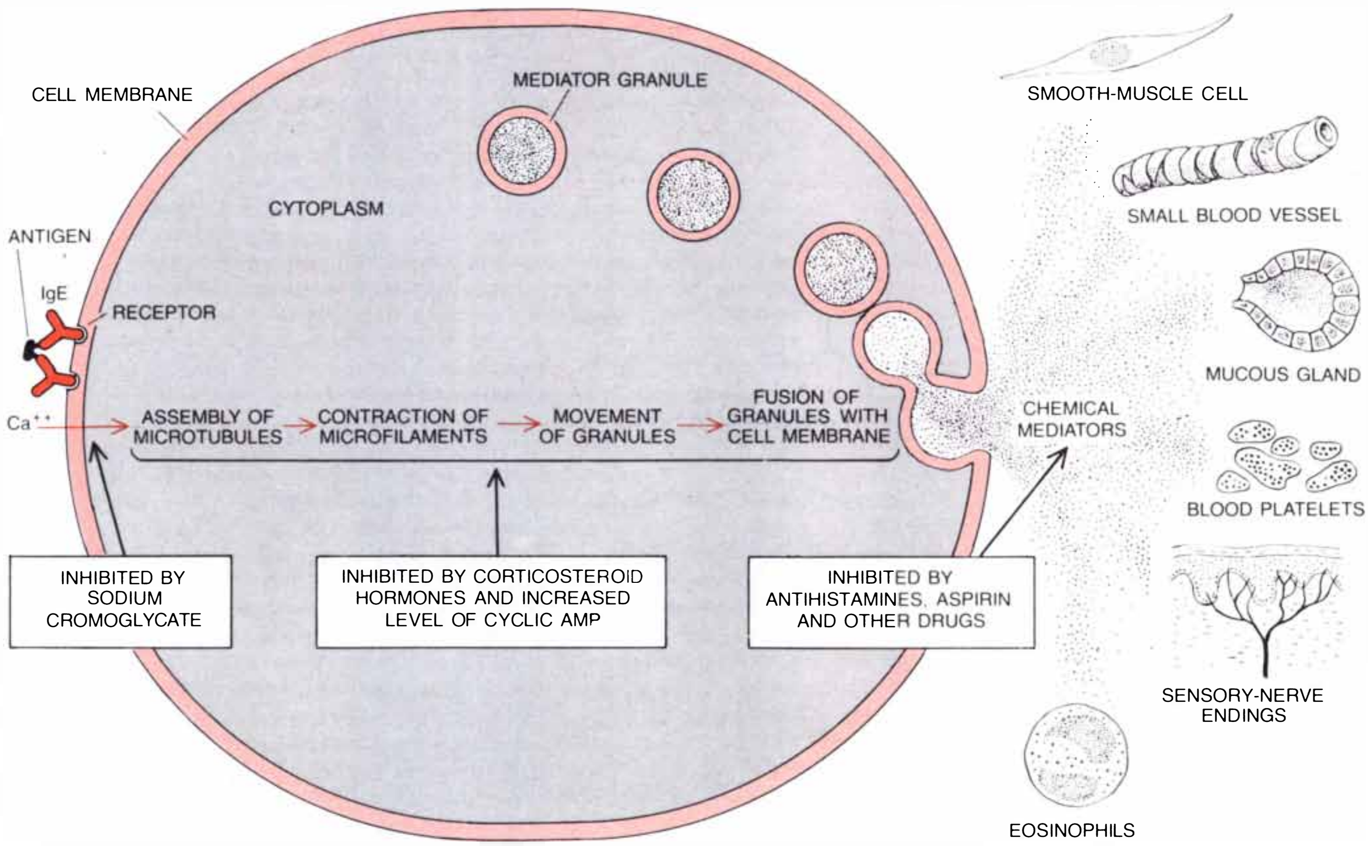
Food Allergy

Allergy to airborne material is fairly common, affecting more than 10 percent of the population. Allergy to foods is less common, but its true incidence is uncertain because food-induced symptoms are often less well defined than the hay fever or asthma induced by an airborne allergen. The rapid development of urticaria (hives) after the ingestion of a particular food is an indication of a probable allergic response, but many symptoms are less obvious and may mimic those of disorders other than al-

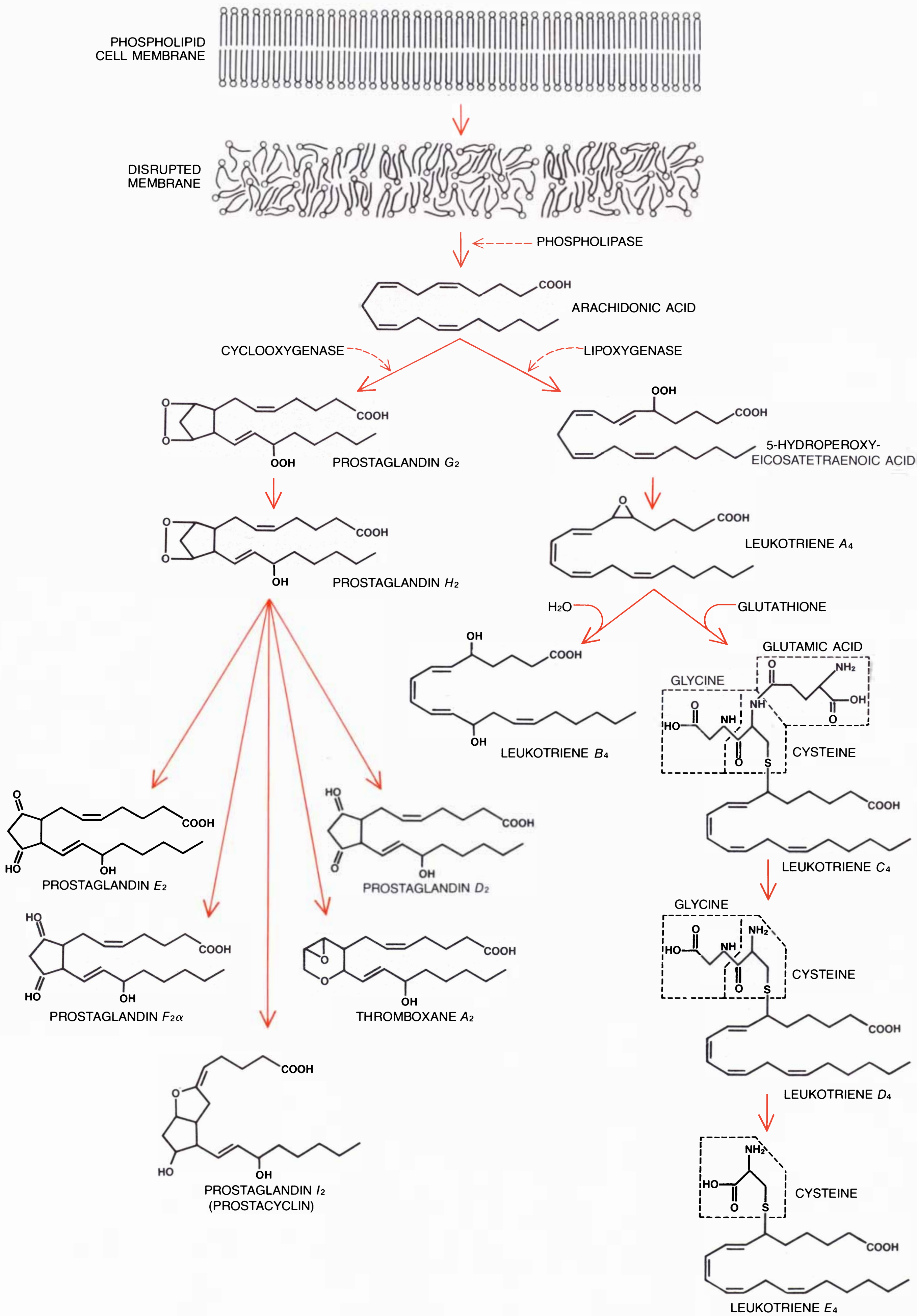
MAST CELL'S ROLE in immediate hypersensitivity is summarized. IgE is secreted by a plasma cell in response to the first appearance of an antigen. The IgE becomes fixed to receptors on the surface of a mast cell. When the antigen is again encountered, epitopes on an antigen molecule are bound by two adjacent IgE molecules. The bridging of the IgE molecules initiates a process whereby mast cell is degranulated and its chemical mediators of the allergic reaction are liberated.



TRIGGERING OF MAST CELL begins when two IgE molecules on its surface are bridged by an antigen molecule. The event activates serine proesterase to form serine esterase, an enzyme that converts phosphatidyl serine into phosphatidyl ethanolamine. The latter is methylated in two steps to form phosphatidyl choline. As a consequence of the changes in the cell membrane there is a marked influx of calcium ions (Ca^{++}) into the cell. The calcium activates the enzyme phospholipase A_2, which promotes the breakdown of the phosphatidyl choline to form lysophosphatidyl choline and arachidonic acid. Other calcium ions enter the cell, where they promote degranulation.



DEGRANULATION AND ITS EFFECTS are summarized. Calcium ions entering the cytoplasm regulate enzymes controlling microfilaments whose activity makes the granules move to the cell membrane, fuse with the membrane and discharge their contents: chemical mediators of the allergic reaction. The mediators act on target tissues and cells. For example, they contract smooth-muscle cells, dilate small blood vessels and make them leaky, stimulate secretion by mucous glands, activate blood platelets, irritate nerve endings in the skin and attract the white blood cells called eosinophils. The process is susceptible to drugs at various stages. The entry of calcium ions can be blocked by sodium cromoglycate. The movement of granules can be inhibited by corticosteroid hormones or by agents that increase the cytoplasmic concentration of cyclic adenosine monophosphate (cyclic AMP). Once released, the mediators can sometimes be neutralized by various drugs, including antihistamines and aspirin, or their activity can be counteracted by drugs such as epinephrine.


PHOSPHOLIPID CELL MEMBRANE
DISRUPTED MEMBRANE
PHOSPHOLIPASE
COOH
ARACHIDONIC ACID
CYCLOOXYGENASE
LIPOXYGENASE
OOH
PROSTAGLANDIN G2
5-HYDROPEROXY-EICOSATETRAENOIC ACID
PROSTAGLANDIN H2
LEUKOTRIENE A4
H2O
GLUTATHIONE
GLUTAMIC ACID
GLYCINE
CYSTEINE
LEUKOTRIENE B4
LEUKOTRIENE C4
PROSTAGLANDIN E2
PROSTAGLANDIN D2
PROSTAGLANDIN F2α
THROMBOXANE A2
LEUKOTRIENE D4
PROSTAGLANDIN I2 (PROSTACYCLIN)
LEUKOTRIENE E4

lergy. Abdominal pain and diarrhea, for example, are most often the result of bacterial or viral infection of the intestines. When these symptoms become recurrent or chronic and no infecting organism can be identified, other causes must be sought.

At Guy's Hospital Medical School in London my colleagues Lawrence J. Youlten and Maurice H. Lessof and I were able to show that in some cases a specific food allergy is responsible. Actually the connection between intestinal distress and one food, milk, was recognized more than two centuries ago by a well-known victim of milk allergy, but apparently not by physicians. A diarist named Sullivan recorded the fact that Charles Stuart ("Bonnie Prince Charlie"), pretender to the British throne, suffered from "a bloody flux" (hemorrhagic diarrhea) and came to associate his symptoms with drinking milk. "The Prince drank no more milk, lived upon watter & was parfectly well."

Our finding at Guy's Hospital began with an observation involving patients with ulcerative colitis. This serious condition has been treated for some 30 years with the drug sulfasalazine, a combination of sulfapyridine and 5-aminosalicylic acid, an aspirin analogue. In 1977 A. K. Khan, J. Piris and S. C. Truelove of the University of Oxford Faculty of Clinical Medicine found that patients who responded to this rather toxic drug would respond equally well to the salicylate (aspirinlike) component alone. The finding interested my colleagues and me because a disorder that responds favorably to aspirin and its analogues is likely to be mediated by prostaglandins, whose synthesis they inhibit. We decided to see if inhibition of prostaglandins would help allergic patients who develop intestinal symptoms after eating certain foods.

A group of food-allergic volunteers found that taking ordinary aspirin before ingesting whatever provoked their diarrhea and colic allowed them to eat or drink the food with impunity. We then looked at a group of patients whose ulcerative colitis had been responding to sulfasalazine but whose symptoms had not been ascribed to allergy. We tested them on a simple diet that eliminated some of the foods most often implicated in intestinal allergies. In a number of these patients the colitis responded well to withdrawal of the suspect foods and recurred when the foods were again included in the diet; the recurrence was prevented when aspirin was taken before eating or drinking. The trial diet and aspirin treatment also worked well in some patients with the less acute intestinal syndrome referred to as spastic colitis or irritable colon.

The evidence that prostaglandins have a role in intestinal allergies was strengthened by Youlten and David Heinzelmann. They showed that ingestion of the allergenic foods was followed by an outpouring of prostaglandins into the intestines and a rise in the level of prostaglandin components in the blood, accompanied by colic and diarrhea. Both the prostaglandin secretion and the symptoms were abolished by treatment with aspirin or one of its analogues. Subsequent studies suggested that several other diarrheal illnesses—both acute and chronic conditions and both allergic and nonallergic ones—involve the release of prostaglandins and can be helped by aspirin. Here, however, a word of caution is required. There is no evidence that allergic reactions in parts of the body other than the intestines are mitigated by aspirin or aspirin analogues. Indeed, these drugs may actually provoke asthma in susceptible individuals.

Treatment

The efficacy of aspirin in treating intestinal allergic symptoms nonetheless suggests it should be effective in other allergic reactions mediated by prostaglandins. So far that has not been found to be the case. In some allergic reactions prostaglandins may have only a peripheral role, augmenting the effects of the mediators. For example, when an experimental wheal-and-flare skin reaction is induced by the instillation of histamine, the concurrent application of prostaglandin E_2 increases the degree of itching tenfold. Yet under clinical conditions aspirin does not alleviate allergic itching. Antihistamine drugs are occasionally helpful in treating hives but not in treating the eczema that is a major symptom of contact dermatitis. Contact dermatitis is a cellular, delayed-hypersensitivity reaction and its itching and pain are induced primarily by lymphokines or some other product of the *T* lymphocytes rather than by IgE and the mast-cell mediators. The only effective drugs for treating such a reaction are the corticosteroid hormones, which (in high doses over a long time) may have side effects as hazardous as the condition they are meant to treat.

In hay fever the repertory of responses is limited to itching of the eyes and nasal passages, sneezing and the secretion of water (tears and runny nose). These responses are largely mediated by histamine, and indeed antihistamine drugs are usually effective in relieving most hay-fever symptoms. Antihistamines are ineffective against asthma, which is a disorder of the lower airways. Some prostaglandins have a constrictive effect on the larger bronchi but in most asthma patients aspirin either has no effect or makes the asthma worse. It is the leukotrienes, with their powerful constricting effect on the smooth-muscle cells of the small bronchioles, that are now thought likely to be the most important mediators in asthma. So far no antileukotriene drug is available. Now that the structure of these mediators is known intensive efforts are under way to find agents that can block their synthesis or their activity. The discovery of such a drug would clearly be a major advance in the treatment of asthma.

The drug sodium cromoglycate is often prescribed for the treatment of hay fever and allergic asthma. It is thought possibly to work by somehow preventing the inflow of calcium ions into the mast cell after the interaction of IgE and an allergen, thus forestalling the chain of events leading to degranulation. The drug must be given before the antigen binds to the mast cell, however; once degranulation begins nothing can stop it. Cromoglycate must therefore be taken in advance as a preventive measure when an encounter with an allergen is expected. The same thing is true of some other drugs that can keep mast cells from degranulating, for example corticosteroid hormones (which stabilize the membrane of granules and of the cell) or agents that enhance the supply of the regulatory molecule cyclic adenosine monophosphate (cyclic AMP). Once the granules have released their contents treatment must be directed at blocking the effects of the various mediators on their target tissues or counteracting the mediators' actions. It is at this stage that the antihistamines have their effect. Corticosteroids, epinephrine and its analogues and (in some special situations) aspirin may also be helpful.

The ideal treatment for allergy, of course, is avoidance of contact with the allergen, but that is not always possible. The next-best hope is to treat a patient so that the contact with a potential allergen does not generate an allergic response. The usual technique is desensitization, an approach developed at St. Mary's Hospital Medical School in London soon after the turn of the century. Small amounts of the offending antigen are administered, usually by a course of injections, in an attempt to

PROSTAGLANDINS AND LEUKOTRIENES are mediators of the allergic reaction that are not stored in mast cells. They are synthesized, primarily in various leukocytes (white blood cells), when arachidonic acid is formed after the cell membrane is disrupted as a result of trauma, infection, hormonal stimulus or some event associated with the allergic response. Phospholipid molecules of the cell membrane are converted by the enzyme phospholipase into arachidonic acid. The enzyme cyclooxygenase (whose activity is inhibited by aspirin and its analogues) initiates the prostaglandin pathway (*left*). The enzyme lipoxygenase initiates the leukotriene pathway (*right*). The leukotrienes C_4, D_4 and E_4 together constitute the constrictor of small bronchioles that is known as the slow-reacting substance of anaphylaxis (*SRS-A*).

stimulate the patient's immune system to make IgG (not IgE) antibodies against the antigen; the rationale is that when the patient later encounters the antigen spontaneously, the IgG will combine with the antigen, whose binding to mast-cell IgE will thereby be blocked.

Desensitization seems to work primarily for some of the pollens associated with hay fever and for bee and wasp venom. It has limited success in asthma patients and is not effective against food allergies or delayed-hypersensitivity (cellular) allergic reactions. Even when "allergy shots" are successful, the mode of action is really not known. The level of IgG—the supposed blocking antibody—in a patient does not seem to be correlated with the degree of desensitization. Some patients who show a favorable response have little allergen-specific IgG in their blood and some in whom the injections are ineffective have large amounts of it. Hence the synthesis of "blocking antibody" cannot satisfactorily explain successful instances of desensitization. There is undoubtedly a substantial placebo effect.

There are other drawbacks to allergy shots. It is not easy to prepare desensitization vaccines in a sufficiently pure state and with standardized potency on a commercial scale. Most important, the injections can generate severe local or systemic allergic reactions, which in rare instances can be fatal. For all these reasons many physicians have come to regard desensitization treatment with considerable reserve.

The Allergic State

Given that the drugs and other treatments available for dealing with particular allergic reactions are all more or less imperfect, a more fundamental approach to therapy would be to prevent the development in an individual of the state of atopy itself. This calls for better understanding of the factors predisposing to the allergic state, so that they can be avoided or suppressed. Allergies tend to run in families; an allergic person is likely to have more relatives with allergies than would be expected on the basis of chance. Either genetic factors or family-environment factors must therefore be at work. The mode of inheritance of susceptibility to allergy is still in dispute. Some investigators find evidence that a single gene is responsible, but they disagree on whether the gene is dominant or recessive. Others think the allergic state is polygenic: transmitted by a number of genes whose combined expression is necessary to make a person allergic.

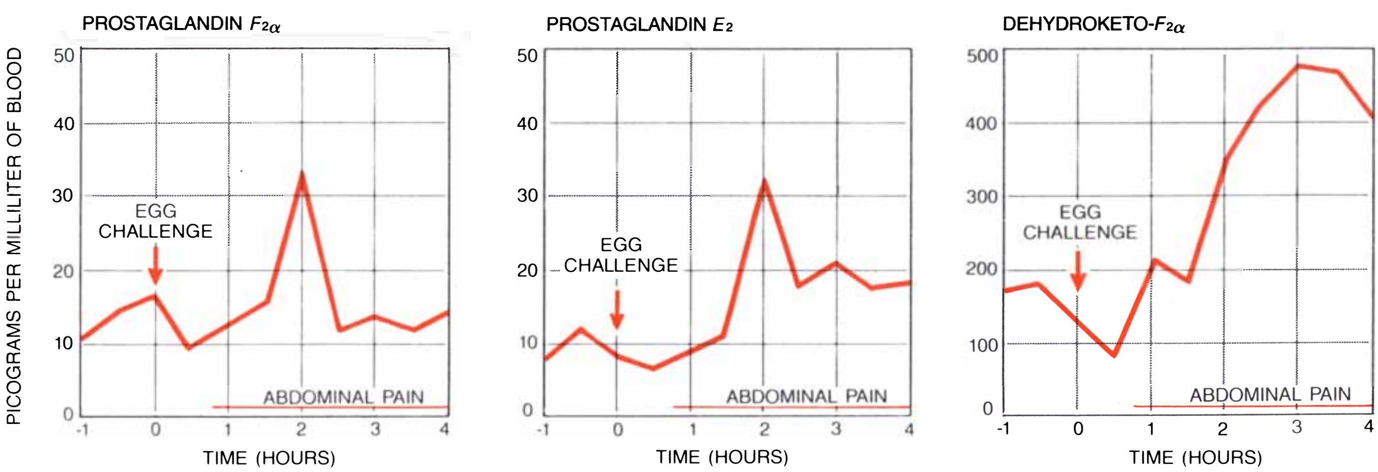
The genetics is complicated by the fact that allergy, like many clearly established heritable conditions, exhibits "incomplete penetrance," or variable genetic expression. Two people may both have the gene or genes specifying the allergic state and their environments may be similar, and yet only one of them may show clinical evidence of allergy. Data from a Swedish study of twins suggest that some 18 percent of the population carry the gene or genes for allergy and that the penetrance is about 40 percent: fewer than half of those who are genotypically allergic manifest the condition clinically. Another study, in the U.S., showed that two allergic parents (a combination seen in less than 1 percent of the study group) had a 58 percent chance of having an allergic child; where one parent was clinically allergic the risk of having an allergic child was 38 percent; where neither parent was allergic (although one or both could have been genotypically allergic if the trait is recessive) the risk was 12.5 percent.

The gap between the inheritance of allergy genes and their expression is generated not only by some inherent degree of penetrance but also by environmental factors, particularly those affecting the newborn infant. A most important factor seems to be the mode of feeding: breast or bottle. For some decades it has been known that infants fed exclusively on their mother's milk are less likely to develop gastrointestinal infections than infants given cow's milk or a formula based on cow's milk. Breast-fed infants are also less likely to develop allergies. There may be a connection between the two observations.

As early as 1934 Clifford G. Grulee of Rush Medical College in Chicago showed that among some 20,000 infants 36 percent of those fed only mother's milk developed neonatal respiratory or gastrointestinal infections, whereas 63 percent of those fed on formula did so. This was before antibiotics were available, and whereas only .13 percent of the babies who were exclusively breast-fed died of infectious diseases in the neonatal period, 7.56 percent of the bottle-fed infants died. Mother's milk presumably confers protection against infection by supplying antibodies to pathogenic bacteria and viruses during a time when the baby's intestinal immune defenses are not fully developed.

Mother's Milk and Allergy

In 1970 John A. Walker-Smith of the Queen Elizabeth Hospital for Children in London showed that gastrointestinal infections in infants are somehow directly related to the subsequent development of allergic illness. In 1976 (by which time the incidence of neonatal infection was much reduced) I studied the neonatal feeding history and the genetic predisposition to allergy of 72 children aged five and older who were allergic to cow's milk. The largest group (60 percent of the children) had at least one allergic parent and also had been bottle-fed in infancy; 26 percent had a family history of allergy but had been breast-fed; 14 percent had nonallergic parents but had been bottle-fed; there were none who had been breast-fed and who had no allergic parent. In other words, it is the combination of a family history of



FOOD ALLERGY was demonstrated in a patient who developed abdominal pain and diarrhea soon after eating eggs. There was a concomitant rise in the blood level of the prostaglandins $F_{2\alpha}$ and E_2 and of the prostaglandin breakdown product dehydroketo-$F_{2\alpha}$. When aspirin was taken before the egg challenge, there was no increase in prostaglandin concentrations and there were no intestinal symptoms.

allergy and early exposure to cow's milk that leads to a high risk of subsequent allergy to cow's milk.

How is one to explain these various observations? Under normal circumstances in mature individuals very little ingested foreign protein gets through the intestinal wall into the bloodstream, where it can encounter *B* cells and elicit an immune response. Most food protein is broken down into peptides and individual amino acids in the stomach and the small intestine; it is these breakdown products that enter the bloodstream, and they are not antigenic. What little protein ordinarily gets through is apparently tolerated by the immune system. If too much protein passes through the epithelial lining of the intestine, however, the immune system can be sensitized and an allergic response can ensue.

Perhaps gastrointestinal infection can damage the lining of the intestine and thereby promote the passage of proteins through the wall. That would provide one possible link between bottle feeding (which increases susceptibility to infection) and subsequent intestinal allergies. Even in the absence of infection it may be that the intestine of a newborn baby is hyperpermeable to food proteins, as it is known to be in many newborn animals. If that is the case, neonatal exposure to foreign (cow's milk) proteins may be allergenic, whereas exposure to the mother's milk proteins may not be. It is likely too that there are factors in mother's milk that help to seal the infant's intestinal epithelium, making it impermeable to foreign proteins; this is a well-documented phenomenon in many other mammals. For lack of those sealing factors a bottle-fed infant may be exposed even later in life to proteins it would not otherwise encounter. Such effects may be all the more important in an individual genetically predisposed to allergy. Whatever the connections between bottle-feeding and allergy may be, many pediatricians have come to believe a newborn infant, and in particular one with parents or siblings who are clinically allergic, should receive absolutely no milk other than mother's milk for the first three to six months of life; it is likely that the most vulnerable period is the first three to seven days.

Perhaps other factors predisposing to allergy will be discovered and will be as avoidable as bottle-feeding. Perhaps eventually ways will be found to switch off the production of IgE once it has started, to prevent the attachment of IgE to the mast cell or even to remedy the defect in *B*-cell regulation that leads to the manufacture of IgE. More immediate objectives should be the development of a real capability for desensitization and the discovery of drugs that can neutralize the leukotrienes and cope more effectively with the chemical mediators liberated by mast-cell granules.

JOIN

THE PLANETARY SOCIETY

EARLY MORNING ON A MARTIAN DESERT—VIKING 1

AND SEE THE WORLDS

IN THE LAST TWO DECADES the human species has undertaken an unparalleled series of voyages of exploration and discovery. For the first time in history we have studied our neighboring worlds close-up and in detail. We have examined the enigmatic pyramids and ancient river valleys of Mars; the broiling, asphyxiating hellhole called Venus; the active volcanos on Io; the clouds of organic matter on Titan; the exquisite system of 10,000 Saturnian rings. Our generation is privileged to be the first to make these discoveries and to calibrate our own tiny and lovely little world.

WE HAVE DISCOVERED forty new worlds that are plausible objectives for future missions. The United States, the Soviet Union, the European Space Agency and Japan are planning missions that are likely to include landers, a balloon observatory and a radar orbiter at Venus; encounters with Halley's Comet; and an orbiter around Jupiter to examine its moons and radiation belts and to drop, for the first time, a probe into its atmosphere. Such ventures will be remembered by our descendants thousands of years from now when most other aspects of our epoch will be obscure and forgotten.

THE PLANETARY SOCIETY is the largest space interest group in the world and the fastest-growing membership organization in America in the last decade. In just one year its membership has grown to more than 100,000 people from all walks of life, from every state and every continent. Members receive a bi-monthly color publication, The Planetary Report, filled with timely pictures and articles. The Society holds lectures for members and major exhibits such as Planetfest '81, the largest privately-sponsored event in the history of solar system exploration. Members can purchase pictures, slides, and the latest books at special rates. The success of the Society has also permitted us to support a few relatively low-cost but critical scientific projects, including the construction of a quarter-million channel receiver and computer system to be attached to existing radio telescopes in the search for signals from advanced civilizations in space. Membership in The Planetary Society provides an opportunity to participate in the greatest adventure of our times—the exploration of the planets and the search for extraterrestrial life. Join The Planetary Society and see the worlds.

OFFICERS: CARL SAGAN, President, BRUCE MURRAY, Vice President, LOUIS FRIEDMAN, Executive Director
BOARD OF ADVISORS: DIANE ACKERMAN, JAMES VAN ALLEN, HARRY ASHMORE, ISAAC ASIMOV, RICHARD BERENDZEN, JACQUES BLAMONT, RAY BRADBURY, JOHNNY CARSON, NORMAN COUSINS, FRANK DRAKE, LEE A. DUBRIDGE, JOHN GARDNER, SHIRLEY M. HUFSTEDLER, JAMES MICHENER, PHILIP MORRISON, PAUL NEWMAN, BERNARD M. OLIVER, HARRISON H. SCHMITT, ADLAI E. STEVENSON, III, LEWIS THOMAS

*I accept your invitation to join **THE PLANETARY SOCIETY***

NAME ______________________________

ADDRESS ______________________________

CITY ______________ STATE ______________ ZIP ______________

☐ *My $15 dues payment is enclosed.* ☐ *I prefer to be billed.*

International members please add $5.00

*Mail checks to: **THE PLANETA… …adena, CA 91103***

Surface Diffusion

Atoms or molecules are strongly bound on a clean metal surface, but they can be surprisingly mobile in two dimensions. At low temperature they apparently "tunnel" from one site to the next

by Robert Gomer

On the atomic scale even the smoothest surface of a metal is a rough and undulating terrain. Because the atoms at the surface cannot be subdivided the surface looks somewhat like a layer of billiard balls, made by cutting a regular, three-dimensional array of balls along a plane. When metal atoms are exposed at a surface, some of the bonds that would ordinarily connect them with other atoms in the bulk of the solid are broken, and so the surface atoms have a bonding capacity that is not completely satisfied. A clean metallic surface is therefore a highly reactive region, capable of forming strong bonds with most atoms. Such a surface can even break a molecule into its constituent atoms or into molecular fragments.

In spite of the roughness of the surface, atoms or molecules that are adsorbed on it can in many instances move about. The motion of the adsorbed particles is called surface diffusion; the study of the motion reveals features of the binding of adsorbates that cannot be obtained in any other way. As the adsorbed particles diffuse across a surface the strength of the bond that holds them to the surface varies from place to place. The reason is that the bond strength depends on the geometric arrangement of the atoms making up various facets, or surface planes, of a particular substance. Moreover, the adsorbed atoms or molecules can interact with one another, and the consequences of the interactions for diffusion lead to still more fascinating observations. Atoms diffusing on a surface are the best approximation to a strictly two-dimensional system currently available. Two- and three-dimensional systems are predicted to differ in certain fundamental properties, such as the way they can become ordered and the way they transport heat and matter. Through surface diffusion the differences can be explored.

Surface diffusion is also an important step in many chemical reactions that are catalyzed by surfaces. On a surface the dissociation of a molecule requires much less energy than it does in a liquid or a gas because the molecular fragments can bind to the surface with a net release of energy. In many such reactions at least some of the molecular fragments must diffuse across the surface in order to encounter other adsorbed particles with which they combine to form new molecules. If the new molecules are not strongly bound to the surface, they can leave it and make room for more reactants. For example, in the catalytic converter of an automobile exhaust system a thin film of platinum or another metal dissociates oxygen molecules and adsorbs both the oxygen atoms and the molecules of carbon monoxide. The mobility of the oxygen atoms and the carbon monoxide molecules is sufficiently great for the two kinds of particles to come into contact and react to form carbon dioxide. The carbon dioxide molecules then escape from the metallic surface.

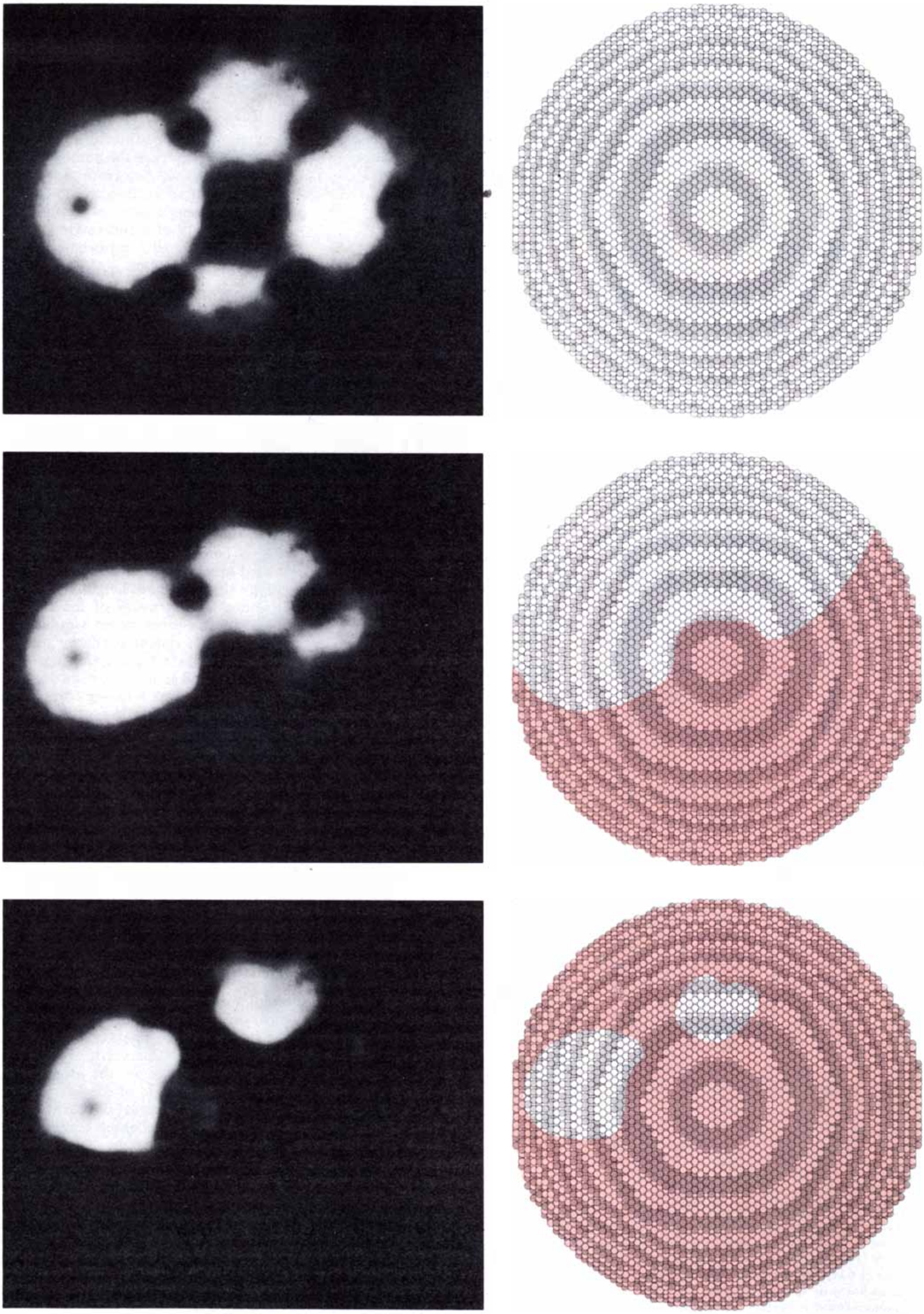
An ordinary metallic solid is crystalline, that is, its atoms are arranged in regular, repeating units. Many kinds of repeating structural unit are possible; one of the simplest is the body-centered-cubic structure, found in tungsten and a number of other metals. The basic structural unit consists of atoms that lie at the eight corners and in the center of a cube. Stacks of unit cubes form the regular crystalline lattice. A planar cut through the lattice can be parallel to one of the faces of the cube or it can make any angle to the face. Different arrays of atoms are thereby exposed on the surface. The cuts are designated by a system of indexes. For the purpose of discussing surface diffusion, however, it is sufficient to recognize that the density and the arrangement of the metal atoms on an exposed surface plane depend on the orientation of the cut with respect to the lattice.

The various surface planes differ in the way they interact with the molecules of an adsorbed substance. Consider the diatomic molecule of hydrogen, H_2. In the H_2 molecule two hydrogen nuclei share two electrons, reducing the total energy of the atoms and forming a chemical bond. A bond between atoms can form only if a nucleus can share electrons in an exchange with its bonding partner. Some molecules bond to atoms, solids or other molecules by the same mechanism. In molecular hydrogen, however, the bonding capacity of the two atomic nuclei is completely satisfied by the two shared electrons, and so the molecule cannot form strong bonds with anything.

A hydrogen atom, on the other hand, can establish a strong bond with most metallic surfaces because it can share its single electron with the atoms of the metal. If more energy is released when the surface bonds are formed than is spent in breaking the bond between the two hydrogen atoms, diatomic hydrogen molecules tend to dissociate on the surface.

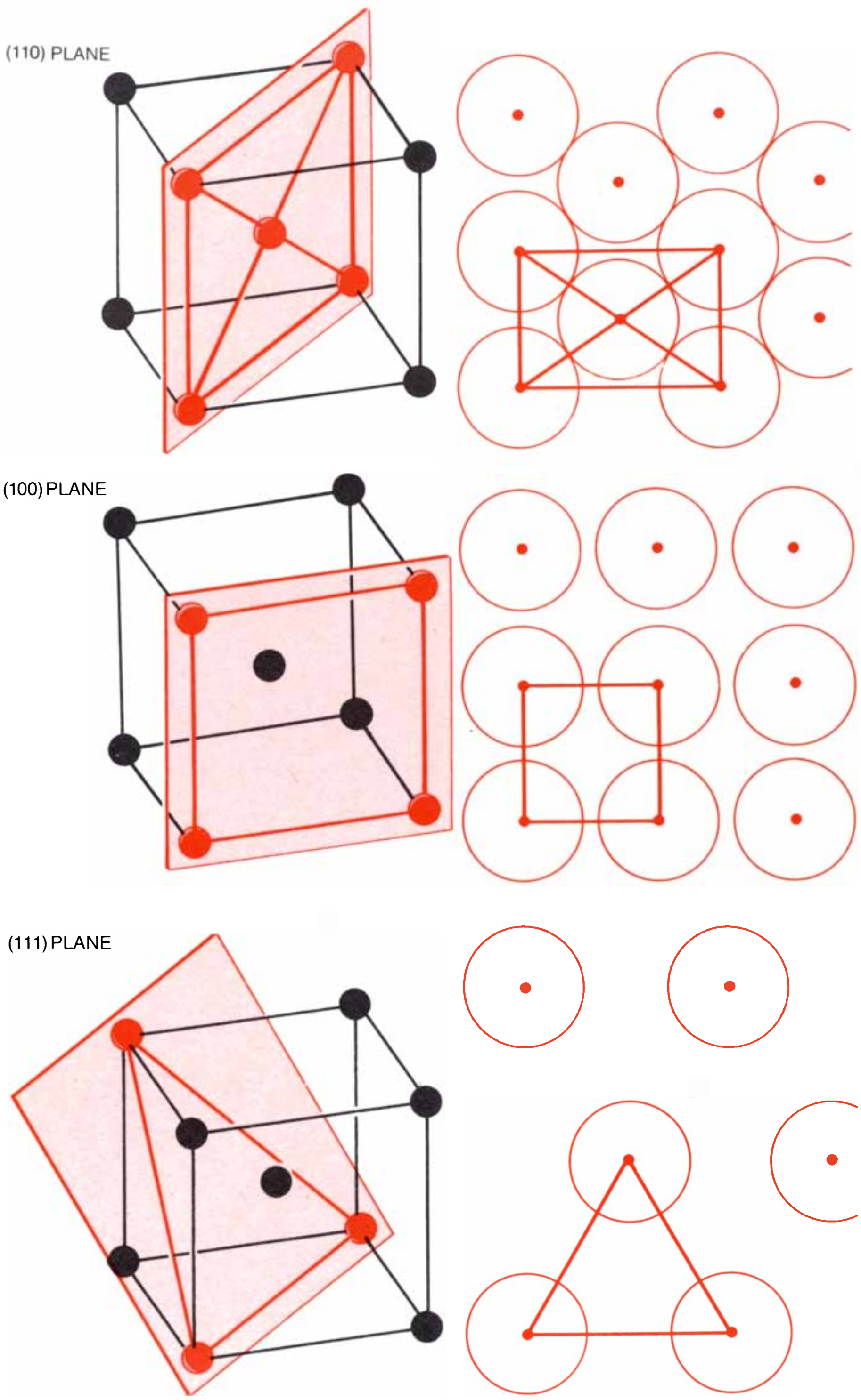
Energy is required to initiate the process of dissociation, but it can be drawn from the thermal energy of the metal atoms. Most surfaces on which dissociation is favorable can pull apart a diatomic molecule such as H_2 even at low

DIFFUSION OF HYDROGEN ATOMS on the hemispherical tip of a tungsten needle is shown in successive stages in the three photomicrographs at the left, made with the instrument called the field-emission microscope. At the right the diffusion is shown schematically; each tungsten atom is represented by a ball and the diffusing hydrogen atoms are represented by the areas shown in color. The photomicrograph at the top shows the clean surface of the single crystal of tungsten that forms the tip of the needle. The dark square in the center and the other dark spots correspond to smooth facets of the crystal, where the atoms exposed at the surface are packed close together. In the middle photomicrograph the lower half of the needle tip is covered with hydrogen atoms at 20 degrees Kelvin and appears dark. In the bottom photomicrograph the emitter is heated to 190 degrees K. and the hydrogen atoms diffuse over the surface. The regions invaded by hydrogen appear dark. The diffusion boundary moves almost radially outward from the central, closely packed facet. The magnification is 300,000 diameters.



temperature. Some crystal surfaces that are relatively smooth on the atomic scale, however, bind H_2 so weakly that the temperature required to make it desorb, or leave the surface, is lower than the temperature needed for dissociation. Thus hydrogen molecules do not dissociate on such a surface even though hydrogen atoms can bind quite strongly to it. One way the atoms can reach the surface is by diffusion from neighboring surfaces that can pull apart the diatomic molecules.

The adsorption of atoms or molecules onto a surface through the formation of chemical bonds is called chemisorption. In many respects the bonds between a surface and a chemisorbed substance resemble the bonds in a small molecule; for example, the energies required to break bonds of both kinds are comparable. In a small molecule, however, the number of atoms is constant and the atoms remain fixed in space with respect to one another, except for small vibrations about their equilibrium positions. On a metallic surface the number of adsorbed atoms can vary from zero to the number required to cover the entire surface with a single layer of atoms. The surface density of a full single layer is approximately 10^{15} adsorbate atoms per square centimeter.



PLANES INTERSECTING a regular atomic lattice can cut the lattice so as to expose a variety of atomic arrangements at the surface. The atoms at the surface can be packed close together, as they are on the so-called (110) plane (*top*), or they can be more widely separated, as they are on the (100) plane (*middle*) and the (111) plane (*bottom*). Such planes of atoms make up the facets of a crystal; each plane has a characteristic symmetry. The lattice structure shown is the one assumed by tungsten and certain other metals. It is called body-centered-cubic. One metal atom is found at each corner and in the center of a cube; the cubes repeat to form the crystal.

The regularity of the atoms exposed on the surface of a metallic substrate creates many energetically equivalent sites for adsorption. The exact positions of the optimum sites vary with the composition of the substrate and the adsorbate. Some adsorbates are bound most strongly on top of a substrate atom, whereas others favor the gaps between the atoms. In spite of these variations adsorbed atoms are still bound quite tightly everywhere on a metal surface. For this reason surface diffusion is possible without desorption.

It is convenient to picture the adsorbed atoms on an abstract potential-energy surface whose height at any given place is a measure of the binding energy an adsorbed atom would have there. On the potential-energy surface the regions where the bonding is strongest correspond to energy wells, and the regions of weakest bonding correspond to energy peaks. The alternating pattern of wells and peaks has the same regularity or symmetry as the atomic surface itself. In moving from one energy well to the next an atom need not climb all the way to an energy peak; instead it can take the path requiring the least climbing by moving over the passes, or saddle points, between the wells.

How do the adsorbate atoms acquire the energy needed for climbing over a pass? At any temperature above absolute zero the atoms of the substrate and the adsorbate vibrate randomly about their equilibrium positions. Most of the vibrations of the adsorbate atoms have a small amplitude, but occasionally a big push or several concerted pushes from the substrate can impart enough energy to an adsorbed atom for it to move up and over the pass. As the atom slides down the other side of the barrier into the next well it generally loses its energy rapidly and becomes trapped; the buildup of energy must then start all over again. On each successive jump the atom is equally likely to jump into any accessible well, including the one it came from. Because the energy imparted to an adsorbed atom depends on the temperature the number of successful jumps increases rapidly with increasing temperature.

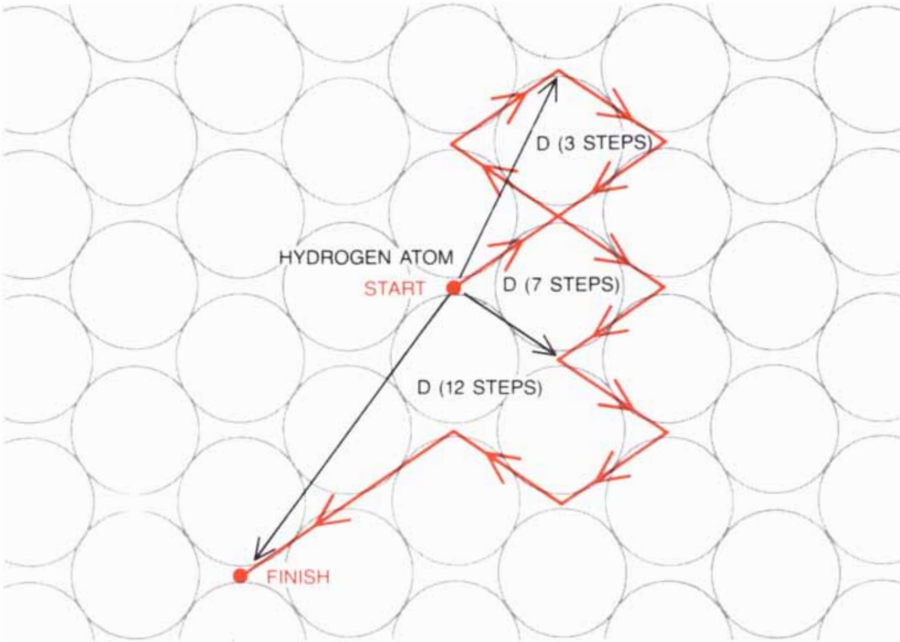
The motion of an adsorbed atom over

the surface is said to be diffusive, and the resulting path is called a random walk. The path can be compared to the aimless wanderings of a drunk, who begins his walk at a lamppost and with each step is equally likely to proceed in any direction. If the lamppost is in a large open place, the drunk can be expected to take as many steps to the East as to the West and as many to the North as to the South, and so he is "not going anywhere." Nevertheless, he covers more and more territory as his meanderings continue; on the average, the area he covers is a circle with the lamppost at its center. The square of the radius of the circle, which is called the mean square displacement, is equal to the total number of steps he has taken multiplied by the square of the length of each step. Hence after an infinite amount of time the drunk can be found with equal probability anywhere on the surface available to him.

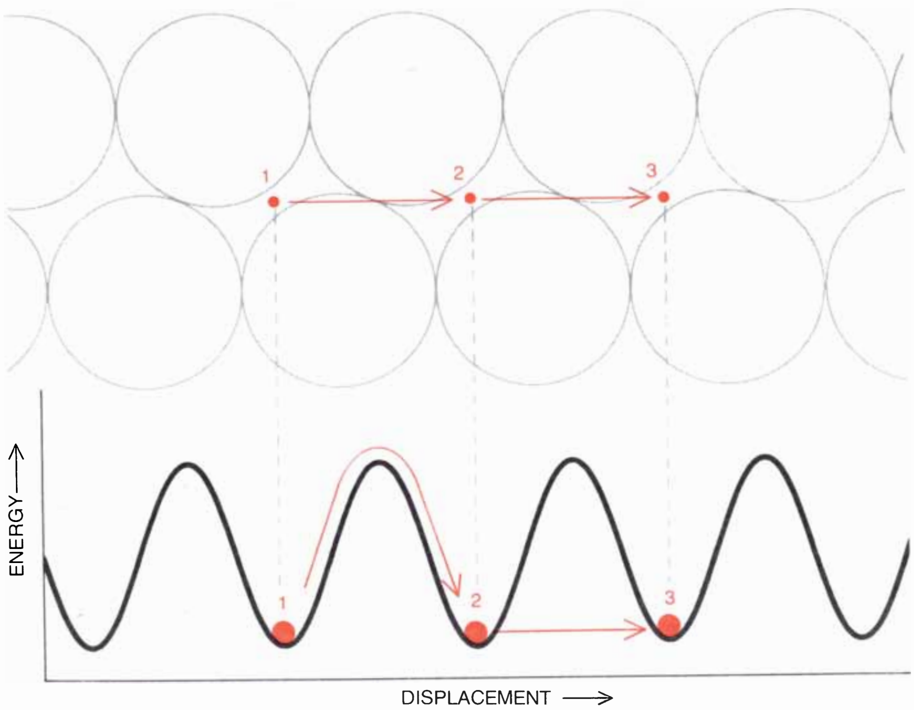
For the adsorbed atom the number of jumps made in a given time is the average frequency of its jumps multiplied by the time. Thus the mean square displacement is calculated by multiplying the square of the jump length by the jump frequency and the time. Both the jump length and the jump frequency are properties of the particular adsorbate-substrate system, whereas time is not. It is therefore convenient to define a quantity called the diffusion coefficient that describes diffusion but does not depend on time. The diffusion coefficient is equal to one-fourth of the jump frequency multiplied by the square of the jump length.

The diffusion coefficient is a measure of the intrinsic rate at which an adsorbed particle is most likely to be displaced from its initial position. If the adsorbed particle is not isolated but also interacts with other particles adsorbed on the surface, diffusion becomes more complex. As the coverage is increased many of the stable bonding sites become occupied; in general an adsorbate particle cannot enter a filled site. Moreover, the interaction of one adsorbate particle with the substrate can affect the motion of other adsorbate particles. In spite of these effects the diffusion coefficient for a set of adsorbate particles is similar to the coefficient for an isolated particle. In both cases the coefficient is determined by the adsorbate and substrate materials, by the substrate topography and by the temperature at which the process takes place.

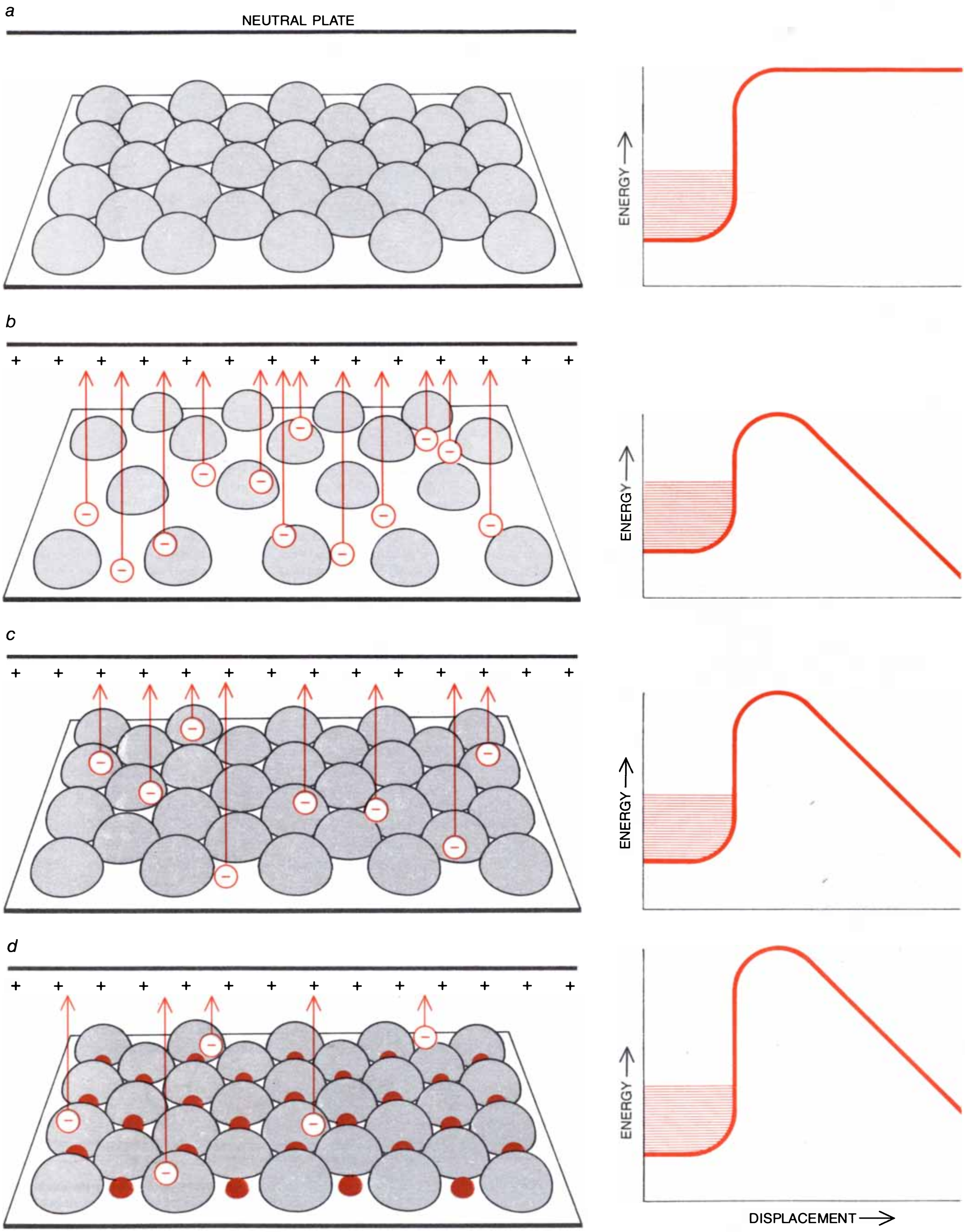
How can surface diffusion be studied? The most straightforward approach is to cut a large, single crystal of a metal to expose the surface plane one wants to study, polish and clean the surface, mask half of it and cover the other half with adsorbate atoms at a temperature too low for the atoms to jump about. When the surface is heated, the ad-



RANDOM WALK of an adsorbed atom proceeds from site to site across an array of surface atoms much as a drunk might stumble aimlessly away from a lamppost. The adsorbed atom can minimize its energy at many equivalent sites in the array where it is most tightly bound to the surface. To move from site to site the atom must gain enough energy from the thermal vibrations of the underlying lattice to jump over an energy barrier. Once the jump is made, however, the adsorbed atom loses its energy to the lattice. Thus each jump is independent of the previous one, and so the direction in which the atom jumps is random at each new site (just as the drunk may set out in a new direction with each step). The square of the most probable distance the atom travels in any direction from its start is proportional to the number of jumps made. A quantity called the diffusion coefficient, which is a measure of the mean-square distance covered by an adsorbed atom per unit time, is defined with respect to the random-walk model.



POTENTIAL-ENERGY DIAGRAM shows the amount of energy associated with each position of an adsorbed hydrogen atom at rest on a surface of closely packed tungsten atoms in the (110) configuration. According to classical, or non-quantum-mechanical, principles, the only way an adsorbed particle can move from one energy well to another is to acquire enough kinetic energy to overcome the pass, or potential-energy saddle, separating the wells (*arrows from point* 1 *to point* 2). According to quantum mechanics, however, the adsorbed particle cannot be precisely localized: there is some probability that it will be found anywhere. In particular there is some probability that a particle in one energy well will be found in an adjacent well, without having gained sufficient energy to jump over the energy barrier (*arrows from point* 2 *to point* 3). The particle is said to have tunneled through the barrier. Tunneling can be an important mechanism of diffusion at low temperature for atoms having a low mass.



ELECTRON EMISSION from a metallic surface in a strong electric field varies with the density of the metal atoms at the surface and with the coverage of the surface by an adsorbate. Electrons in a metal occupy many discrete but closely spaced energy levels (*horizontal lines in graphs at right*). In the absence of an electric field (*a*) no electrons leave the surface because they cannot surmount the potential-energy barrier at the surface and because the barrier is infinitely thick. When a strong electric field is applied, however, the potential energy outside the surface of the metal decreases, so that the energy barrier confronting the electrons in the metal has a finite height and thickness (*b*). As a result electrons tunnel out of the metal. When the surface atoms are closely packed (*c*), the energy barrier is higher and the electron flux smaller. When the surface is covered with an adsorbate (*d*), the barrier is higher still and the flux becomes even smaller.

sorbed atoms diffuse onto the bare surface. From an experiment of this kind, at least in principle, one could determine the rate of spreading, that is, the change in the adsorbate concentration over the surface with time.

Such experiments have been done, but they are difficult and have many drawbacks. Even the best macroscopic surfaces have steps and other imperfections that can affect diffusion. Moreover, the coverage, or percentage of the adsorption sites occupied by the adsorbate, varies from its initial value on the covered half of the substrate to zero on the other half. Because the diffusion coefficient can be strongly dependent on the coverage the interpretation of such experiments is complex.

There is another difficulty. Since clean surfaces are so reactive, a surface-diffusion experiment must be carried out in an extremely high vacuum, in order to avoid contaminating them with unwanted atoms from residual gas in the apparatus. Even in the best vacuum systems surfaces become contaminated in from 10 to 100 minutes, and so the duration of the experiment is limited. Because diffusion is slow it is difficult to measure enough movement in such a short time.

What is needed, then, is a method for studying surface diffusion over short distances on single crystal surfaces free from imperfections. Ideally it is desirable to study diffusion on a surface with a constant or nearly constant coverage of adsorbate. The experimental technique should also enable the constant coverage to be varied over as wide a range as possible.

Many of these requirements, although not all of them, are met by an instrument called the field-ion microscope, invented by Erwin W. Müller in 1950. The device can resolve single metal atoms adsorbed on the planes of a microscopic metallic crystal; the planes are free of defects even on the atomic scale. Gert Ehrlich of the University of Illinois at Urbana-Champaign, Michael Bassett of the Imperial College of Science and Technology of the University of London and Tien Tzou Tsong of Pennsylvania State University have employed the field-ion microscope to make extremely detailed studies of the diffusive motion of metal atoms adsorbed on metallic substrates. Unfortunately only metallic adsorbates can be observed because adsorbates such as hydrogen, oxygen and nitrogen do not form images in the field-ion microscope. Another device, however, enables the diffusion of almost any adsorbate to be studied. It is called the field-emission microscope, and it was invented by Müller in 1937. In many cases field-emission studies meet all the desired specifications for the investigation of surface diffusion.

The field-emission microscope is not a microscope in the conventional sense of the term. Its essential component is a very sharp needle with an approximately hemispherical tip. Electrons are emitted by the tip and their trajectories diverge almost radially outward from it. If they impinge on a fluorescent screen, they create a highly magnified map of the surface of the tip. The magnification of the instrument is approximately equal to the tip-to-screen distance divided by the radius of the tip [*see illustration below*].

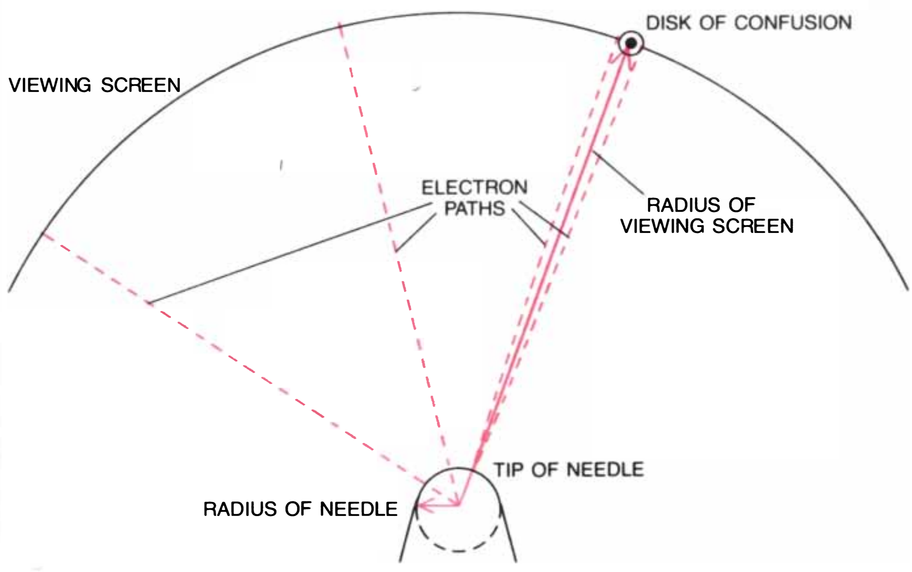
It is possible to prepare an emitter needle by etching a fine wire; a needle made of tungsten or some other metal with a high melting point can then be cleaned and smoothed by heating it to a high temperature. By this means the needle can be given a tip with a radius smaller than 1,000 angstrom units (one angstrom unit is equal to a ten-billionth of a meter). A magnification of a million or more can easily be achieved.

Still further magnification of the needle tip could be attained simply by constructing a bigger screen having a larger distance from the needle, but such modifications would not improve the resolution of the microscope. Electrons leaving the surface of the needle do not reach the screen along perfectly straight lines; they emerge with some random velocity perpendicular to the direction of emission, and so the source of any electron detected on the screen cannot be determined to within less than 20 or 30 angstroms.

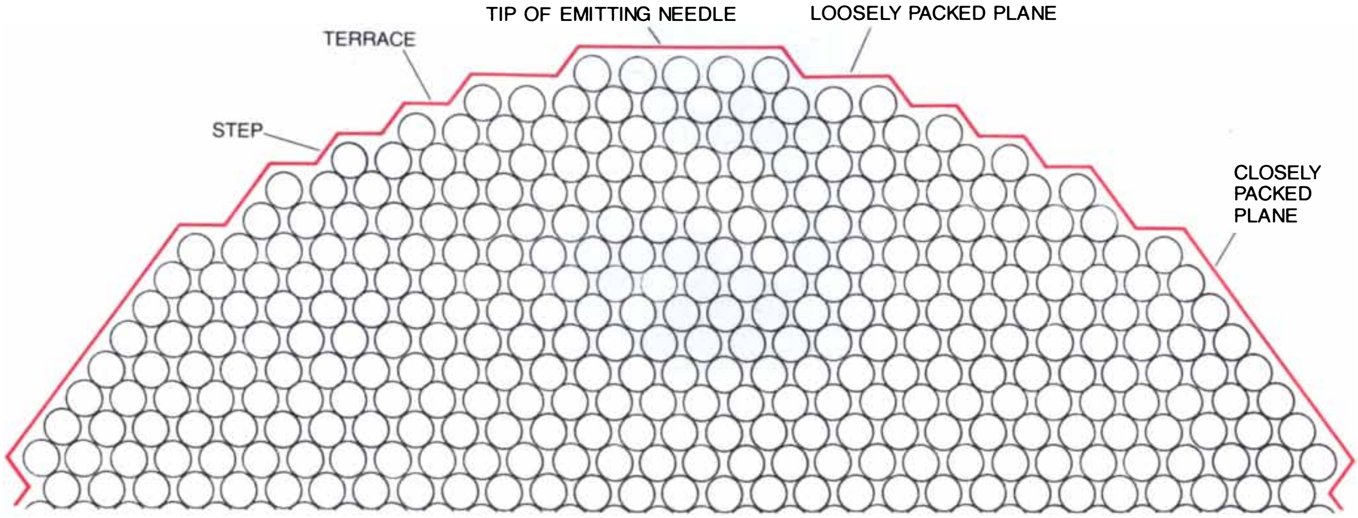
An image generated by the field-emission microscope (that is, any pattern of light and dark regions on the fluorescent screen) is an electron-emission map of the tip surface. Field emission comes about through the quantum-mechanical phenomenon called tunneling. Electrons can move freely in a metal, but they are ordinarily confined within its boundaries: the potential energy of an electron rises steeply at the surface of the metal to a plateau significantly higher than the energy of any electron in the metal. When the surface of the metal is placed in a strong external electric field, however, the potential energy outside the metal decreases with distance from the surface. The electrons are still confronted by a potential-energy barrier at the surface, but the barrier now has a finite height and thickness. According to the laws of quantum mechanics, the position of an electron cannot be fixed with absolute precision, and so there is some chance that an electron originally inside the metal will be found outside it.

The electric field required for appreciable electron emission is high, from about 30 to 70 million volts per centimeter. When a metallic object is electrically charged, the electric field is highest where the curvature of the metal surface is greatest. It turns out that fields high enough for electron emission can be generated at the tip of a needle by applying a potential difference of only a few thousand volts between the needle and the fluorescent screen.

The emitting tip of the needle is much smaller than a single crystal. Hence the



FIELD-EMISSION MICROSCOPE projects a magnified image of the electron-emission pattern at the surface of the tip of a metal needle. Field emission is caused by a strong electric field at the surface of a metal. When a voltage is applied between the needle and a fluorescent viewing screen, the sharp curvature of the needle tip concentrates the electric lines of force and so establishes the necessary electric field at the tip. Electrons tunneling out of the tip follow the lines of force, which diverge radially outward. A highly magnified emission pattern is therefore seen on the screen. The magnification is approximately equal to the ratio of the radius of the viewing screen to the radius of the tip. (The magnification is not exactly equal to the ratio because of the influence of the shank of the needle.) The resolution of the instrument is limited to roughly 30 angstroms because an electron leaving the tip has a small random velocity perpendicular to a radial line. Hence the electron can arrive anywhere within a disk of confusion on the screen. Its origin can be determined only to within a corresponding region on the tip.



ATOMIC STRUCTURE of the surface of the tip of the needle in the field-emission microscope is a multifaceted series of planes and terraces approximating a hemisphere, shown schematically in cross section. The flat plane at the top of the hemisphere merges in single atomic steps with terraces of atoms whose surface configuration is the same as that of the top plane. As the hemisphere curves away from the top the terraces become narrower, and new planes, steps and terraces having different surface configurations are exposed.

tip is almost always part of one crystal, even if the needle itself is fabricated from an ordinary wire made up of many crystals. Although the tip is approximately hemispherical, it has flat facets that correspond to the stablest crystal planes. The regions between the facets are made up of terraces and steps that correspond to the variety of cuts through the crystal needed to create the overall hemispherical shape.

The field emission of electrons is lowest from planes on which the metal atoms are most closely packed; these are the same stable planes that form facets. Emission is greatest from regions where the surface atoms are loosely packed, or in other words where the surface is atomically rough. Thus spatial variations in the intensity of the emission on the fluorescent screen have the same symmetry as the crystal that forms the emitter tip. It is therefore possible to identify the crystallographic directions that correspond to various portions of the emission pattern. What makes the field-emission microscope suitable for studying surface diffusion, however, is the fact that field emission is also sensitive to the presence of adsorbed atoms.

The simplest method of studying surface diffusion, developed primarily in my laboratory at the University of Chicago in the 1950's, is to cover only one side of an emitter needle with an adsorbed gas. Ordinarily, however, gas from a source placed to one side of the needle will quickly cover its entire surface: molecules going past the needle rebound from the walls of the glass tube that encloses the field-emission apparatus and stick to the "shadow" side of the needle. The rebounding can be prevented by cooling the field-emission tube in a bath of liquid helium to within a few degrees of absolute zero. Molecules that hit the wall of the tube are then frozen to the glass on impact, and only the side of the needle facing the gas source receives a deposit. Because most gases have a low vapor pressure at the temperature of liquid helium, immersion in helium also ensures an extremely good vacuum in the tube and so keeps the surface of the needle from becoming contaminated. The method is called shadowing.

When the needle is heated, adsorbed atoms diffuse over the tip and the emission pattern changes. Because of the high magnification a migration rate of a few angstroms per second on the tip corresponds to a rate of almost a millimeter per second on the fluorescent screen. The diffusion patterns can readily be photographed or recorded in a motion picture.

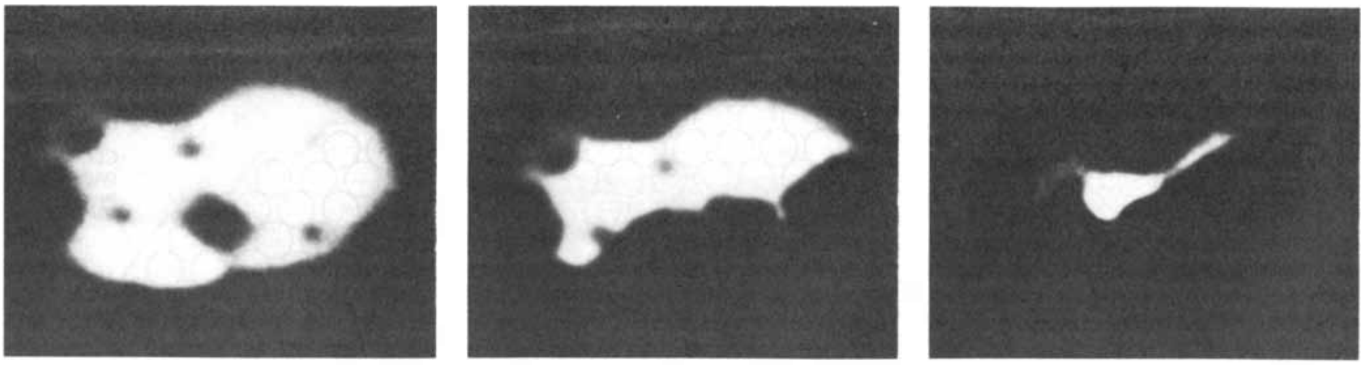
The diffusion patterns observed on a shadowed needle can be the source of much interesting information and are also quite beautiful. For example, when more than a single layer of oxygen molecules (O_2) is deposited on one side of a tungsten emitter needle a sharp boundary is seen moving over the needle even at a temperature as low as 25 degrees K.

The explanation is simple. The first layer of oxygen molecules dissociates on the surface and the resulting oxygen atoms are adsorbed; they become mobile only at 400 to 500 degrees K. Although a tungsten surface can form strong bonds only with a single layer of atoms, molecular oxygen can be weakly adsorbed on top of the chemisorbed atoms at low temperature. Such weak adsorption is called physisorption; the physisorbed oxygen molecules are able to diffuse even at 25 degrees. When they reach the edge of the chemisorbed layer, they "fall off" and are dissociated by the clean surface. The oxygen atoms are then chemisorbed, thereby extending the chemisorbed layer and allowing more second-layer diffusion. The sharp boundary observed visually corresponds to the edge of the chemisorbed layer [*see illustration on page 106*].

Ordinarily diffusion does not show a sharp boundary because in a random walk some diffusing atoms move appreciably ahead of others. Diffusion with a boundary, however, can sometimes be seen in chemisorbed layers as well as in physisorbed layers. If one side of a tungsten needle is covered with hydrogen atoms and the emitter is heated to between 180 and 220 degrees K., a sharp boundary moves almost radially outward from the central, closely packed plane. After the boundary has penetrated some distance into the atomically rougher regions it stops; its advance depends on the size of the initial deposit, and its motion will resume if more hydrogen is deposited from the side.

The phenomena reflect variations in the potential energy of the adsorbed atoms on different regions of the tungsten surface. At 180 to 220 degrees K. hydrogen atoms can readily diffuse only over the smooth facets of the surface; the atoms become trapped on the rough regions. Once the deep energy wells on the rough regions of the surface are saturated with hydrogen, however, the regions become artificially smoothed. The diffusion of additional hydrogen atoms can then proceed as easily as it does on the intrinsically smooth regions.

Since the smoothing of the rough regions is caused by the precipitation of hydrogen into the deep wells, there is a discontinuity between the bare tungsten surface and the part of the surface on which all the deep wells are filled. The discontinuity moves forward as more hydrogen moves across the newly smoothed surface and is trapped in turn [*see illustration on page 99*]. If all the available hydrogen is trapped in energy wells, the diffusion stops. It can be restarted, however, by increasing the tem-



DIFFUSION OF OXYGEN MOLECULES over a tightly bound layer of oxygen atoms adsorbed on a tungsten crystal is shown in three photomicrographs made with the field-emission microscope. The photograph at the left shows a field-emission pattern from a clean tungsten crystal; the central dark square and the dark spots surrounding it correspond to closely packed facets of the crystal. In the middle photograph the bottom half of the crystal is covered with oxygen. The oxygen deposit consists of a layer of atoms, which are chemisorbed, or tightly bound, to the tungsten, and excess molecular oxygen, which is weakly adsorbed on top of the atomic layer. When the crystal is heated to only 25 to 30 degrees K., the molecular oxygen becomes mobile on the chemisorbed layer. When the molecules reach the edge of the layer, they "fall off" and dissociate into their constituent atoms on the bare surface, thereby extending the chemisorbed layer. More oxygen molecules can then migrate over the newly formed chemisorbed layer and become trapped in turn at its edge. The photomicrograph at the right shows the dark advancing limits of the chemisorbed layer of atoms, magnified approximately 145,000 diameters.

perature of the needle so that the atoms can jump randomly from trap to trap.

Although the shadowing method provides a good overview of diffusion on a surface, it cannot give an accurate estimate of the diffusion coefficient on a single crystal plane uniformly covered with an adsorbate. It is possible to modify the field-emission microscope, however, in order for such measurements to be made. A small hole, called a probe hole, is made in the fluorescent screen, and steering electrodes are installed so that the emission from a single, extremely small region of the surface of the needle can be directed through the hole. In this way the electron emission from a region with a radius of only 50 to 100 angstroms can be measured after suitable amplification of the electron signal. Such a region is small enough to lie entirely within one of the larger facets on the tip of the emitter needle. The facets in turn are small enough to be virtually free of imperfections.

If the needle is uniformly covered with an adsorbate and heated until the adsorbed atoms become mobile, the number of atoms in the probed region fluctuates randomly by small amounts. Because the probed region is so small the fluctuations are significant, and they cause measurable changes in the emission current from the probed area. The average time required for a fluctuation to build up or decay is the time required for an atom to move from the center to the edge of the probed area; it is called the mean relaxation time for a fluctuation. By sampling the small fluctuations in emission current from the probed region many thousands of times it is possible to determine their mean relaxation time; from this one can calculate the diffusion coefficient on the probed plane at a specified temperature and coverage.

Since the fluctuations are random, they can provide useful information only through their average properties. In order to sample these properties the so-called time-autocorrelation function of the fluctuations is determined. The function gives a measure of how rapidly a fluctuation "forgets" earlier fluctuations; accurate values of the mean relaxation time and the mean square size of the fluctuations can thereby be obtained. The procedure is called the fluctuation method.

The fluctuation method has made it possible for the first time to determine diffusion coefficients under precisely controlled experimental conditions. Its main disadvantage is that the fluctuation measurements must be made while the surface of the needle tip is subjected to the electric field required for emission. Certain adsorbed atoms, such as those of cesium and sodium, are easily polarized by a strong electric field, that is, their charge distribution is significantly altered. Polarization can affect both the adsorption energy of the atoms and the height of the diffusion barriers, and so the results cannot easily be related to diffusion in the absence of the strong field.

Other adsorbates are only slightly polarized by the electric field; they include atomic hydrogen and oxygen and molecules of carbon monoxide and other gases. For such substances the electric field does not significantly alter the diffusion coefficients, and the fluctuation method has enabled my colleagues and me to investigate in considerable detail how the diffusion coefficient varies with temperature and with the surface coverage of the adsorbate. The results indicate that the phenomena are much more complicated than had been suspected.

The most intriguing systems we have investigated so far by the fluctuation method are the diffusion of hydrogen and deuterium on the most closely packed atomic plane of tungsten, the plane designated (110). This system has been studied by my associate Rocco DiFoggio and me for the past three years.

It is often convenient to graph the relation between the diffusion coefficient and the temperature so that the graph is a straight line whenever diffusion corresponds to barrier climbing by thermally activated adsorbate atoms. This can be done by plotting the logarithm of the diffusion coefficient against the reciprocal of the absolute temperature. In other words, the value of the diffusion coefficient should increase exponentially with a decrease in the reciprocal of the temperature if thermal agitation is the sole cause of diffusion. (Remember that the diffusion coefficient is in effect a measure of the rate at which adsorbate particles climb from one energy well to another.) The slope of the line in such a graph is proportional to the height of the barrier. For atomic hydrogen at temperatures above 110 to 170 degrees K. the diffusion coefficient varies in just this way with the temperature [*see top illustration on page 109*].

Below temperatures of 110 to 170 degrees we found that the graphs changed from steeply sloping lines to horizontal ones. The diffusion coefficient for atomic hydrogen remains almost constant and so becomes independent of temperature to the lowest temperatures we investigated, namely 30 degrees. We immediately suspected that tunneling through the energy barriers between energy wells had come to be more important to surface diffusion than climbing over the barriers.

The tunneling of a hydrogen atom through an energy barrier on a surface is analogous to the tunneling of an electron through the energy barrier at the surface of a metal when a strong electric field is applied. Tunneling is a common phenomenon for electrons, but it is not for atoms and molecules because the tunneling probability decreases rapidly with an increase in the mass of the tun-

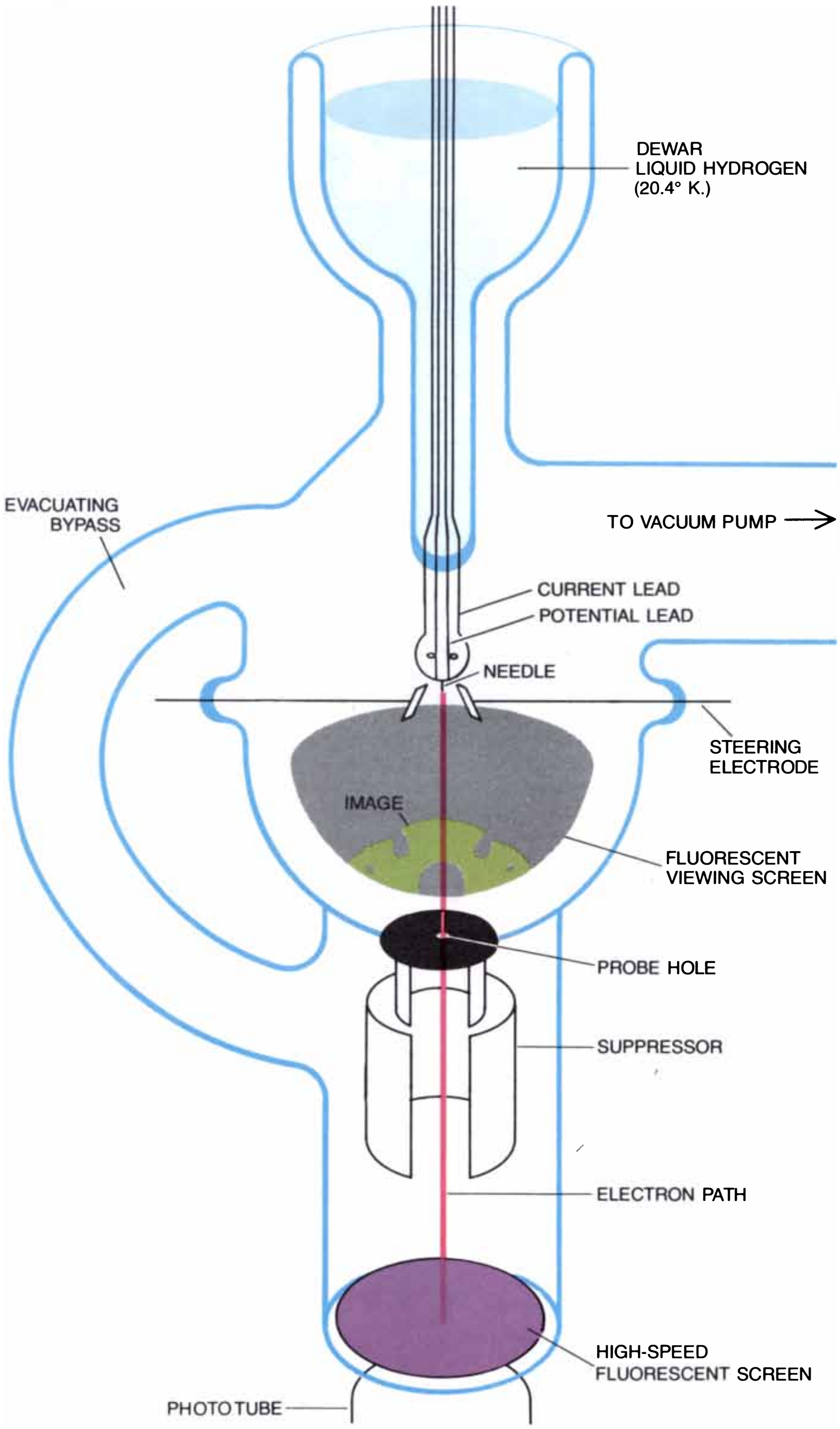
neling particle. Nevertheless, for relatively light atoms such as hydrogen temperature-independent tunneling remains a distinct possibility, particularly in surface diffusion, where the energy barriers are small.

At higher temperatures barrier climbing is much faster than tunneling, and tunneling makes no significant contribution to diffusion. As the temperature decreases, however, and barrier climbing becomes less likely, tunneling becomes more important and the contribution of barrier climbing to diffusion becomes negligible. In this way diffusion can become independent of temperature. To test the hypothesis we decided to study the diffusion of deuterium.

Deuterium is an isotope of hydrogen: its atomic nucleus includes a proton and a neutron rather than just a proton, so that it is about twice as massive as hydrogen. Chemically, however, the two isotopes are almost identical. Because of the greater mass of deuterium we expected that if the temperature-independent diffusion were caused by tunneling, the diffusion of deuterium would also be independent of temperature below a certain threshold but the diffusion coefficient would be significantly smaller than it is for hydrogen. Our data for deuterium at low temperature confirm these predictions; hence it is fairly certain that the tunneling hypothesis is correct.

The demonstration that the tunneling of the adsorbate particles can account for surface diffusion at low temperature may have significance for astrophysics. The formation of diatomic hydrogen molecules in space is thought to take place on cold interstellar dust grains, and the tunneling of hydrogen atoms on the surface of the dust grains is one mechanism that has been proposed to account for it. Although such grains are not made of tungsten, the possibility of tunneling by hydrogen on cold interstellar grains is bolstered by our demonstration that tunneling diffusion can be detected in the laboratory.

Our data for the diffusion coefficients of hydrogen and deuterium suggest that quantum-mechanical effects play a role not only in tunneling but also in barrier-climbing diffusion. One would expect, for example, that at temperatures where barrier-climbing diffusion is dominant the diffusion coefficients of hydrogen and deuterium would be quite similar. Our data indicate otherwise, and the reasons may be closely related to fundamental quantum-mechanical differences between hydrogen and deuterium. The slopes of the graphs of the diffusion coefficients for hydrogen and deuterium are nearly equal, and so the barrier heights are nearly identical for both isotopes. In most cases, however, the curves are displaced from each other, so that the diffusion coefficients are quite different.

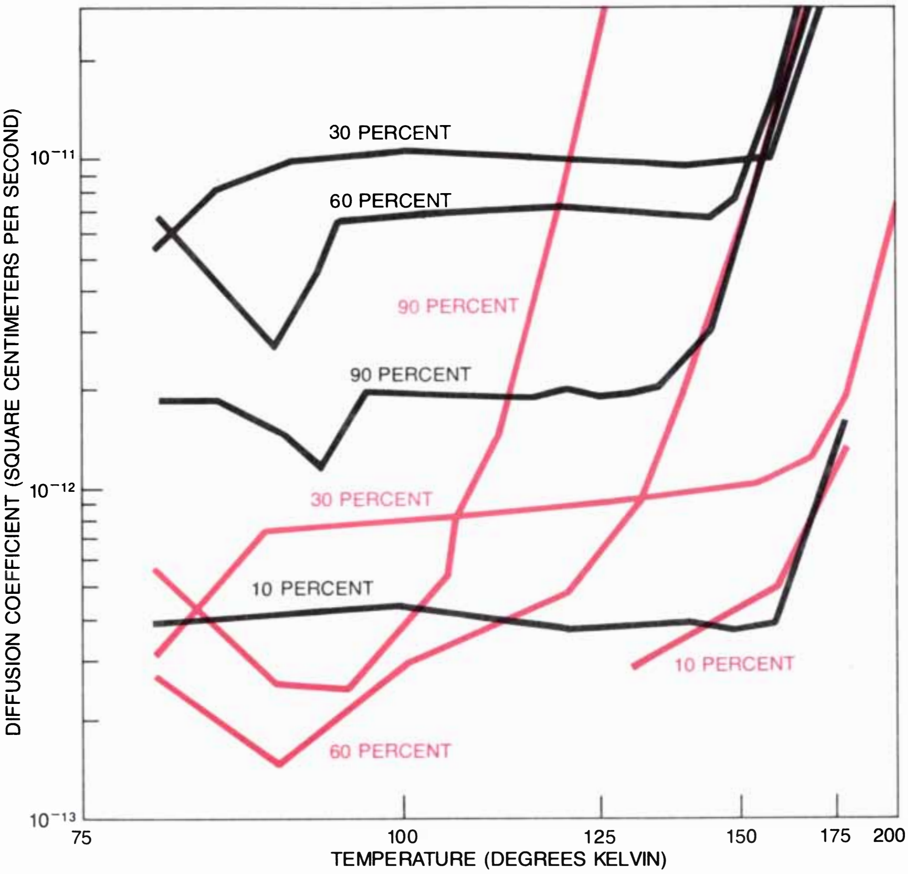


MODIFIED FIELD-EMISSION MICROSCOPE, shown here schematically, can determine the diffusion coefficient of an adsorbate on individual planes of a crystal at constant coverage. Steering electrodes move the pattern on the viewing screen until electrons from a region in the center of the plane being studied pass through a probe hole and onto a fast-reacting fluorescent screen. There the current of electrons from the probed region is converted into light, which is measured with a phototube. The amplified output of the phototube is again an electric current, and the fluctuations in this current are an amplified version of the fluctuations in the electron emissions from the probed region. The fluctuations in emission are in turn caused by the movement of adsorbed atoms into and out of the probed region. The diffusion coefficient of the atoms adsorbed on a single plane of the crystal can be calculated from the rate at which the fluctuations in the emission current build up and decay. The needle is cooled by heat-conducting wires that connect it thermally to a bath of liquid hydrogen. The temperature at the tip can be increased by passing an electric current through a loop of tungsten wire; the temperature of the needle can be monitored by measuring the electrical resistance of the loop. To prevent contamination of the surface the entire system is kept in an extremely high vacuum.

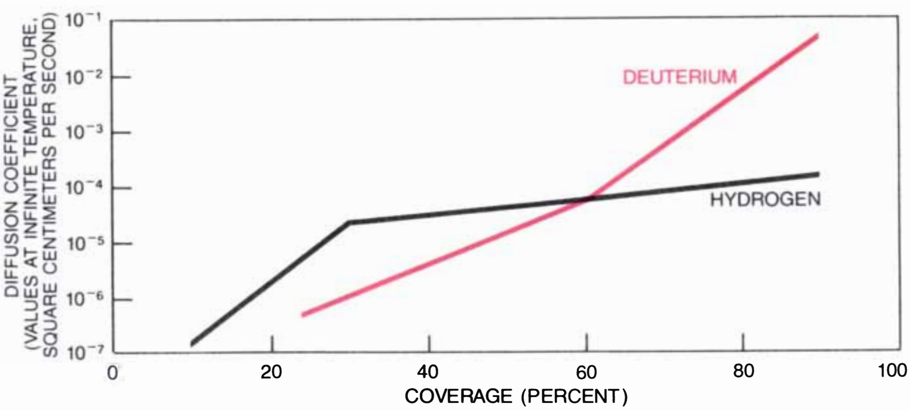
Suppose a hydrogen or a deuterium atom that had cleared an energy barrier finds the energy well on the other side of the barrier already occupied by another atom. It is possible in such a situation for the occupant of the well to be pushed out and over another energy barrier, where it might scatter a third atom. Such a chain of scatterings could continue until one of the scattered atoms lands in an empty site.

For the purpose of determining the length of a jump, a chain of scattering events is indistinguishable from a single long jump by the first adsorbate particle into the final empty slot in the chain. Since the magnitude of the diffusion coefficient varies with the square of the jump length, it is evident that the forward scattering of particles can lead to chains of scattering events that can significantly increase the diffusion coefficient. The probability that long chains of scattering events will take place, however, depends on the quantum properties of the particles forming the chains. It turns out that quantum principles prevent hydrogen atoms from approaching each other closely enough for forward scattering in most encounters, and so long scattering chains do not develop. Deuterium atoms, on the other hand, may be heavy enough to interact according to classical, or non-quantum-mechanical, principles. Hence deuterium atoms may approach closely enough to each other to cause long scattering chains. This may account for the observation that the diffusion coefficients of deuterium are larger than those of hydrogen when the surface is densely covered with adsorbate. (It is under conditions of dense coverage that the scattering chains should be most important.)

Although hydrogen is the simplest and best-understood atom, studies of its diffusion on a single smooth plane with a rather simple atomic structure have revealed some remarkably complex and unexpected phenomena. The unanticipated complexity serves to emphasize that surface diffusion in general is not yet well understood. Secure information has begun to develop about chemisorbed atoms on single crystal surfaces of metals, but almost nothing is known so far about diffusion on semiconductors, on insulators, on amorphous metals or on organic materials. There are virtually no quantitative data on diffusion in a mixture of adsorbates, although such information is important for the understanding of catalysis. Finally, theoretical and computer calculations of diffusion coefficients with models that treat adsorbate-adsorbate and adsorbate-substrate interactions realistically are just beginning to be done, and a great deal remains to be learned about these topics as well. The experimental and theoretical challenges await new ideas and new talent.



DIFFUSION COEFFICIENTS for hydrogen (*black*) and deuterium (*color*) for various coverages on the closely packed (110) surface of tungsten are plotted as a function of temperature. Equal intervals on the horizontal axis correspond to equal units of the reciprocal of the temperature rather than to equal units of the temperature itself. When diffusion is activated by thermal energy, enabling the adsorbed atoms to jump over energy barriers, the logarithm of the diffusion coefficient should increase linearly with the reciprocal of the temperature. The graph shows that at temperatures above 110 to 170 degrees K., depending on the percentage of the adsorption sites occupied, thermal activation is primarily responsible for diffusion. At temperatures below this range the diffusion coefficient varies little with temperature. Diffusion on cold surfaces is caused mainly by tunneling, a process almost independent of temperature.



VARIATIONS IN COVERAGE have different effects on the value of the diffusion coefficient when the adsorbate is ordinary hydrogen and when it is deuterium. When only a small fraction of the available sites are covered by the adsorbate, the diffusion coefficient is lower for deuterium than it is for hydrogen, but as the coverage increases, the diffusion coefficient for deuterium surpasses that of hydrogen. The higher diffusion coefficient for deuterium at higher coverages is probably caused by chains of scattering events as the deuterium atoms diffuse over the surface. A deuterium atom that jumps into a site already occupied by another atom may be able to knock the second atom out of its place. The second atom in turn can displace a third atom and so forth, until the last atom in the chain reaches an unoccupied site. The series of displacements is equivalent to one long jump by the first atom to the last site in the chain; hence the average length of a jump is multiplied many times and the diffusion coefficient is increased. Because of quantum-mechanical constraints hydrogen atoms cannot approach one another closely enough for such displacements in most encounters, but the constraints do not apply to the heavier deuterium atoms. The diffusion coefficients are plotted for the values they would have at infinite temperature so as to eliminate the effects of barrier height.

The Coevolution of a Butterfly and a Vine

Heliconius butterflies lay their eggs only on Passiflora vines. In defense the vines seem to have evolved fake eggs that make it look to the butterflies as if eggs have already been laid on them

by Lawrence E. Gilbert

A major focus of modern evolutionary ecology is investigating the details of interactions between organisms in nature and analyzing the consequences of those interactions. Perhaps the most significant category of ecological interactions in terms of the net transfer of energy in the global food web is the interactions between plants and animals. Such interactions encompass a wide variety of ecological phenomena and the majority of known species on the earth.

There is much current interest in the question of how interacting pairs of plants and animals have influenced each other in the course of evolution. Among the best-known examples of such coevolution are the obligatory relations between figs and fig wasps and between yuccas and yucca moths. In both instances the plants have come to need the services of the insects as pollen carriers and the insects in turn have come to call on the plant to sacrifice some of its ovules as larval feeding sites to promote the insect's reproduction. In such instances plants and animals have taken turns acting as agents of natural selection with respect to each other.

Not all coevolution, however, is mutualistic. In many instances one of the interacting organisms is parasitic on the other. One of the most remarkable interactions for the study of animal-plant coevolution of the host-parasite type is the interaction between certain brightly colored butterflies of the New World Tropics and certain vines. The butterflies are members of the genus *Heliconius;* the vines are passion-flower vines, members of the genus *Passiflora.* The passion-flower vines have evolved effective chemical defenses against insects, but a few insects, the *Heliconius* butterflies among them, have evolved the ability to circumvent these defenses. That ability apparently precludes the *Heliconius* butterflies' being parasitic on other plants. *Heliconius* butterflies thus deposit their eggs only on *Passiflora* vines, where the eggs hatch into larvae that feed voraciously on the leaves of the vine. The remarkable thing is that some species of the vine have features that appear to mimic the distinctive bright yellow eggs of the butterflies.

What accounts for this mimicry, if that is what it is? One possibility is the coevolution of *Passiflora* with *Heliconius.* At first this explanation may seem farfetched, particularly to anyone familiar with the long list of insects, mites and fungi that attack a single species of oak or maple in Temperate Zone regions. How could any one trait of a plant be causally attributed to natural selection imposed by one species or genus of insects among so many? The answer is that such a selective effect is almost impossible. That is why plants such as the passion-flower vines, with their simplified, specialized populations of animal parasites, are of such interest. With only a few major herbivores such as *Heliconius* to account for, interpreting the defensive traits of passion-flower vines is relatively free of ambiguity. It also helps that the *Heliconius* butterflies themselves are highly visible, are easily studied in the field and in artificial culture, and are prime experimental animals for investigations ranging from behavior to population biology.

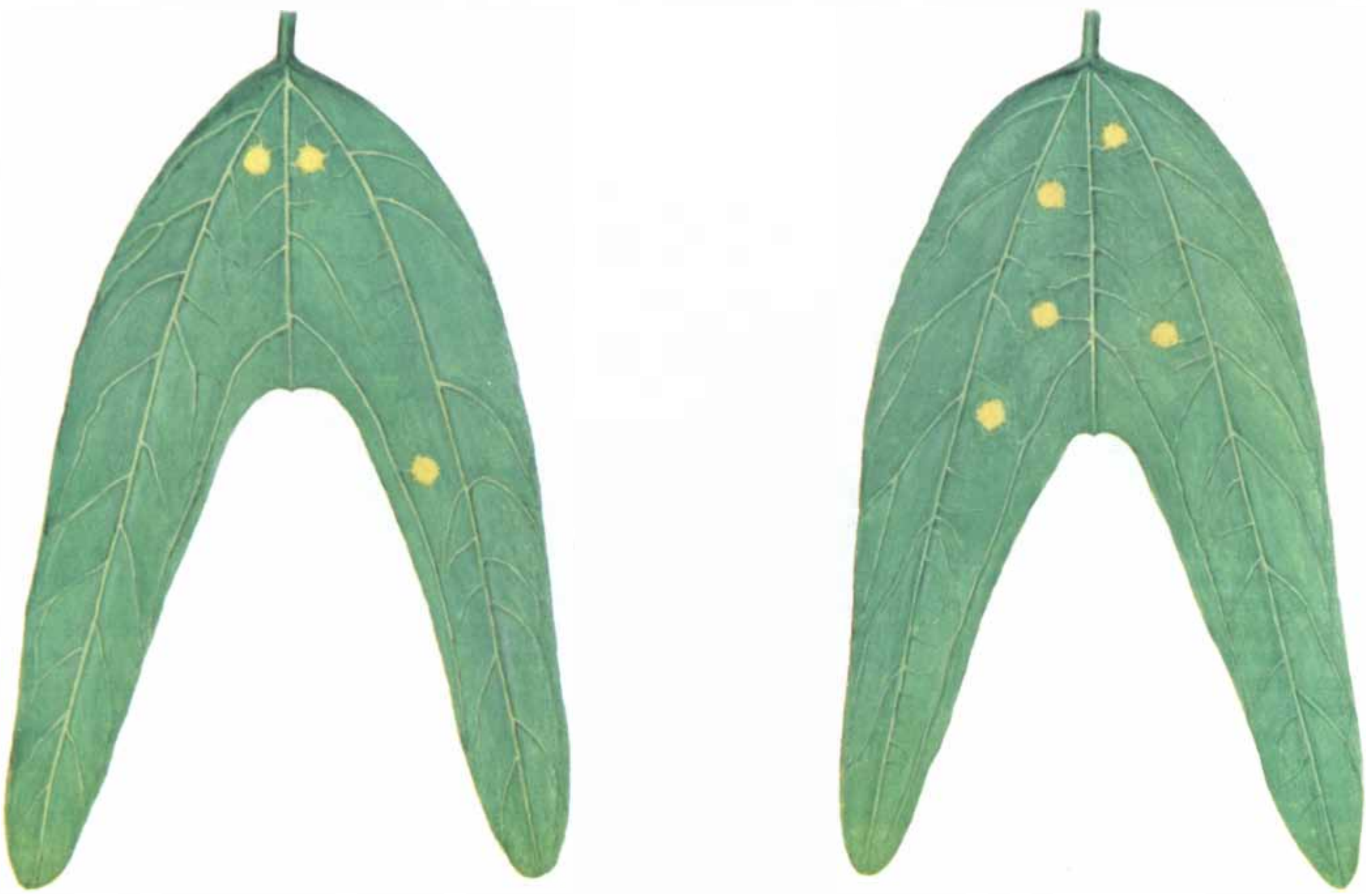
Other features of the *Passiflora-Heliconius* system also make it ideal for the demonstration and study of coevolution. For one thing, the interaction between the vines and their butterfly parasites is highly detrimental to the plant: the leaves on which it depends for its own nourishment are eaten by the larvae, and occasionally the defoliation is fatal. In other words, the coevolutionary interaction is a strong one. Equally useful for the investigator is the fact that the same basic interaction is found over a wide geographic area, so that considerable evolutionary divergence can be observed in both the plants and the insects.

Some 500 species of passion-flower vine are found in the New World Tropics. Most can be identified only by botanists who work in the Tropics and by the butterflies whose emerging larvae may feed on a single species of *Passiflora.* For example, in Costa Rica only the *Passiflora* vine species *P. pittieri* is parasitized by the *Heliconius* species *H. hewitsoni.* The female butterfly deposits its eggs selectively on the vine's freshest shoots. On the other hand, the females of the butterfly species *H. cydno* lay their eggs on many different species of passion-flower vine. In another respect, however, the *H. cydno* females are more fastidious than the *H. hewitsoni* ones; whichever passion-flower vine *H. cydno* parasitizes, it lays its eggs only on the freshest shoots close to the forest floor. Nevertheless, *H. cydno* is still relatively specialized, being unable to feed on any other family of plants.

As far as egg laying is concerned, preferences of this kind mean the butterflies' opportunities are limited. The extent of this limitation is indicated by a 10-year survey of the flora of Barro Colorado Island in Panama recently completed by Thomas B. Croat of the Missouri Botanical Garden. He determined that the island has 133 families of plants, represented by a total of 1,369 species. Of this total the number of species belonging to the family of passion-flower vines was only 11. If one can accept this proportion as an average for the New World Tropics, *Heliconius* parasites have available to them less than 1 percent of the plant life there. This sparse pasture is made sparser still by the butterflies' preference for depositing their eggs on fresh shoots. It is clear that in the mere discovery of an appropriate egg-laying site the egg-laying female butterfly faces a difficult and time-consuming task. This, however, is only part of the story.

Observation in the field reveals that even after the female has found an appropriate host plant and the appropriate

EGGLIKE YELLOW STRUCTURES appear on the three species of *Passiflora* (passion flower) vines shown in this painting. They mimic the yellow eggs of *Heliconius* butterflies that lay their eggs on the vines. The larvae of the butterfly then feed on the vine. At the left is a stem of the *Passiflora* species *P. cyanea;* the main modified egglike structures are the swollen ends of stipules: paired leaflike appendages. In the middle is a stem of the species *P. auriculata;* the main modified structures are nectar glands of the leaf stem. At the right is a stem of an undetermined species of passion-flower vine of northeastern Peru; the main modified structures are nectar glands of the leaf near the point where the leaf is attached to the leaf stem. In this species delayed expansion of stem-developing leaves keeps growth points hidden behind leaf displaying fake eggs. Growth points are vulnerable to being fed on by caterpillars that hatch out of real eggs.



FAKE EGGS ARE RANDOMLY DISTRIBUTED on two leaves of the species *P. candollei*. Here the egglike structures are slightly raised nectar glands. Shoots of *P. candollei* seem to have eggs haphazardly laid around the growth point, simulating shoots laden with eggs.

HELICONIUS BUTTERFLY is one of a large genus in the New World Tropics. At the left a female of the species *H. ismenius* is shown depositing an egg on a vine of the species *P. serratifolia*. At the right a caterpillar of the species *H. pachinus* is shown on a shoot of another vine.

fresh growth on the host it may not deposit an egg. Apparently the site not only must suit the female but also must be both suitable and safe for the progeny that will emerge from the eggs. What are the criteria of suitability and safety? Competition with insects other than *Heliconius* butterflies for egg-laying sites is apparently not among them. Few other insects deposit their eggs on passion-flower vines. Flea beetles and flag-legged coreid bugs are among those that do, but neither of these insects is as skilled as the butterflies at finding new shoots on mature vines or at finding isolated juvenile vines.

What are the consequences for a passion-flower vine of being parasitized by a *Heliconius* butterfly? Much depends on the vine's age and its position within its habitat. The seed of a passion-flower vine can germinate only on open ground where the soil is exposed to sunlight. Such gaps are found where the fall of a tree or the erosion of the land has created a break in the leafy canopy that normally shades the forest floor. The vegetation grows rapidly to fill such a gap, and so the vine seedlings are in a race to rise and remain in the light of the sun. A single *Heliconius* caterpillar can easily consume all the leaves of a juvenile vine in the interval between the caterpillar's hatching and its going into the pupal stage. Such defoliation may not kill the vine, but it will certainly slow its rate of growth at a critical time in its life cycle. If the seedling is thereby locked into a permanent juvenile condition, it may never reach sexual maturity and produce seed.

A mature vine that has managed to reach the canopy or the subcanopy of the forest is rarely subject to complete defoliation. The reason is that glands on parts of the vine other than the flowers secrete an "extrafloral" nectar that attracts predaceous ants, wasps and flies, all of them capable of destroying young *Heliconius* caterpillars. Representatives of one ant genus, *Ectatomma*, can colonize the base of a large passion-flower vine and then forage over the entire plant for months if not years.

The presence of such "guards" does not, however, prevent the *Heliconius* caterpillars from injuring or destroying the new vine shoots before they themselves are destroyed. One of my students, John T. Longino, and I have observed that the fate of the eggs and caterpillars on a parasitized passion-flower vine may be determined by whichever of several ant species patrols the vine. Some ants, including *Ectatomma* species, are primarily interested in the nectar secreted by the vine. They ignore both the eggs and the caterpillars until damage to the tip of a shoot blocks the development of a new branch.

On balance, whereas *Heliconius* caterpillars may keep juvenile passion-flower vines from attaining sexual maturity, their actions in preventing the growth of new branches and flowers on mature passion-flower vines can be equally damaging. Plants that do not flower, like plants that do not mature, will produce no seed. Neither juvenile nor mature passion-flower vines need be totally defoliated in order for their reproductive fitness to be adversely affected.

These observations, made both in the field and in the greenhouse, show how *Heliconius* parasitism can exert a selective pressure sufficient to affect the evolution of the host plants. Within any particular population of passion-flower vines a chance genetic variant that is at a lower risk of infestation but is otherwise normal has a reproductive advantage over susceptible individuals. The mutant individual will make a larger contribution to the population of seedlings on the forest floor than individuals that are at a higher risk of reduced seed production. Such fitness gradients due to the activities of the butterflies can in time change the genetic makeup of host-plant populations and species.

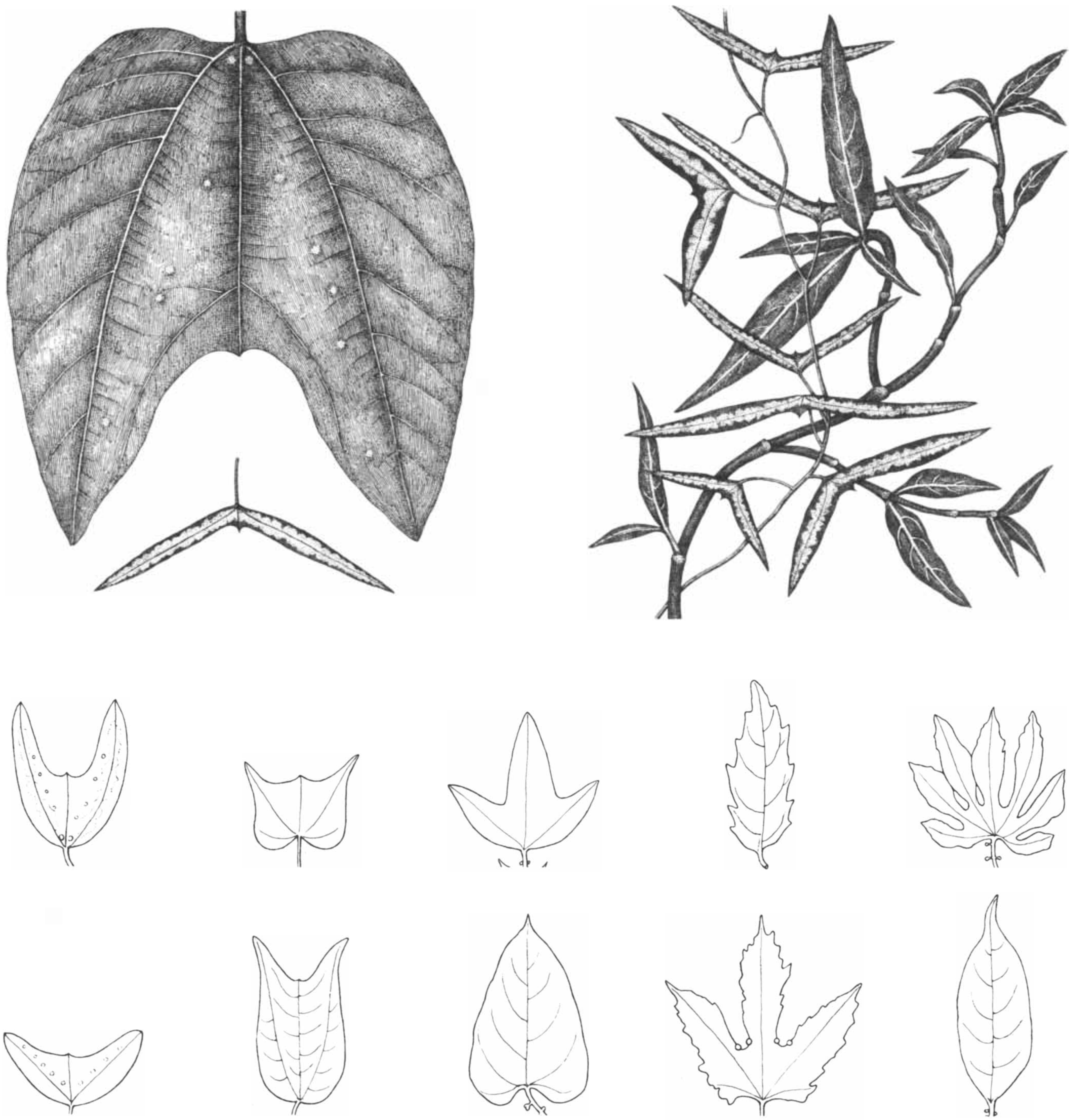
That being the case, what traits would make a mutant vine less subject to parasitism? To answer this question one must understand three aspects of the interaction between the butterfly and the vine. The first aspect is how the female butterfly finds a host plant. The second is how the butterfly makes a choice between depositing its eggs or not depositing them. The third consists of the factors that affect the survival of the eggs and the caterpillars after they are in place on the vine. Natural selection resulting from *Heliconius* parasitism would obviously favor any mutant vine that was harder for the butterfly to find, that was less likely to be selected for egg laying once it was found and that was inhospitable to the butterfly's offspring once they were hatched. (Of course, as the passion-flower vines evolved under these pressures, natural selection would also favor any mutant butterfly that was better able to find a passion-flower vine. I shall not, however, go into this component of coevolution here.)

To take up the first aspect of the interaction, just how does the female butterfly find a suitable host vine? Laboratory experiments by Newton H. Copp and Demorest Davenport of the University of California at Santa Barbara have shown that when a stream of air is passed over the leaves of passion-flower vines, it carries away with it certain volatile substances that elicit a response in butterflies closely related to those of the genus *Heliconius*. In nature, however, this kind of odor orientation would serve only to bring a searching butterfly to the general vicinity of a vine that was one among hundreds of forest plants. More precise localization would call for direct contact with the surface of a leaf. *Heliconius* butterflies, like other members of their family (the Nymphalidae), have a specially modified pair of front legs that can "drum" on the surface of a leaf. This direct means of sampling leaf substances is evidently what enables the butterfly to recognize the appropriate host plant.

Even before this chemical assay is accomplished the searching female evidently responds to cues of a different character: visual perceptions. These aspects of host selection are complex. For example, in my laboratory at the University of Texas at Austin, Thomas Abrams showed that young *Heliconius* females without prior experience were strongly attracted to wire models of tendrils of the passion-flower vine. This behavior suggests that at least some *Heliconius* species instinctively sort plants into two visual categories: vines and nonvines.

Field and greenhouse observations of older females suggest that they respond to even subtler visual cues. For example, I have often seen *Heliconius* females approach and investigate plants that resemble their particular host plant in overall appearance and leaf shape. This suggests that with experience the butterfly progresses from a stage where any likely leaf is drummed to a stage where only leaves with shapes like those of the desired host are drummed. Occasionally in greenhouse cultures eggs are actually deposited on totally inedible plants that only resemble host vines.

Adequate experimental investigation of this possibility of searching on the basis of learned images has not yet been done with *Heliconius* subjects. Mark D. Rausher of Duke University has, however, investigated similar behavior in the butterfly *Battus philenor,* the pipevine swallowtail. Butterflies of this species deposit their eggs only on plants of the birthwort family, specifically the genus *Aristolochia* ("Dutchman's pipe vine"). Working in the field in eastern



VARIETY IN LEAF SHAPE is another characteristic of *Passiflora* vines that is thought to have evolved under the selective pressure of parasitism by *Heliconius* butterflies. In some species of passion-flower vine the shape of the mature leaf (*leaf at top*) differs greatly from the shape of the juvenile leaf (*leaf below leaf at top left*); the leaves shown here are from vines of the Venezuela species *P. cuneata*. Some vines are further protected from *Heliconius* parasitism by having leaves shaped like those of nonhost plants in the setting typical of that stage of the vines' growth. For example, the leaves of a juvenile plant of the species *P. cuneata* (*top right*) generally resemble those of several nonhost plants common in small forest clearings where juvenile passion-flower vines grow; the leaves of one nonhost plant are shown intermingled with the *P. cuneata* leaves. An egg-laying female butterfly whose search for a host plant was guided by leaf shape could easily overlook the leaves of the passion-flower vine. The rows of leaves in the middle and at the bottom illustrate the variety of *Passiflora* leaves in two areas: Trinidad (*middle row*) and Costa Rica (*bottom row*). The two leaves in each vertical pair show the variation between close relatives in the two areas. From left to right are *P. tuberosa* (*top*), *P. biflora* (*bottom*), *P. capsularis* (*top*), *P. costaricensis* (*bottom*), *P. cyanea* (*top*), *P. oerstedii* (*bottom*), *P. quadrangularis* (*top*), *P. vitifolia* (*bottom*), *P. serrato-digitata* (*top*) and *P. ambigua* (*bottom*).

Texas, Rausher found that *Battus* females searching for egg-deposition sites spontaneously approached not only *Aristolochia* leaves but also all other leaves similar in general shape to those of the particular *Aristolochia* species the butterflies had most often encountered in a particular part of their flight season.

This possibility that recognition of leaf shape plays a part in the search for host plants, combined with the likelihood that such recognition will reduce the reproductive potential of the plant, led me to propose that three elements in the morphology of the leaves of passion-flower vines may have evolved under the pressure of parasitism by *Heliconius* butterflies. The first element, which is supported by field observations, is that in many habitats the leaves of the passion-flower vine closely resemble in shape and texture the leaves of certain nonhost plants growing abundantly there. My colleagues and I have noted remarkable resemblances between the leaves of particular *Passiflora* species and those of coexisting plants such as *Philodendron* or *Rubus*. In other habitats a *Passiflora* species has the leaf characteristics of other plants. An example is the lanceolate, leathery leaves shared by *P. ambigua* and many trees and shrubs of the mature rain forest.

The second element is also supported by field observations. In many areas where different species of passion-flower vine coexist they differ from one another in leaf shape over a surprisingly wide range. The third element, also supported by field observations, is that as many species of passion-flower vine grow from seedling to maturity the shape of their leaves changes dramatically. Moreover, the leaves of the juvenile vine tend to resemble the leaves of other plants in its forest-floor habitat, and the leaves of the mature vine tend to resemble the leaves of other plants in its forest-canopy habitat. This suggests that the pressure of *Heliconius* parasitism has favored the evolution of leaves that deceive the egg-laying female butterfly. The evidence supporting the hypothesis, however, remains circumstantial and comparative.

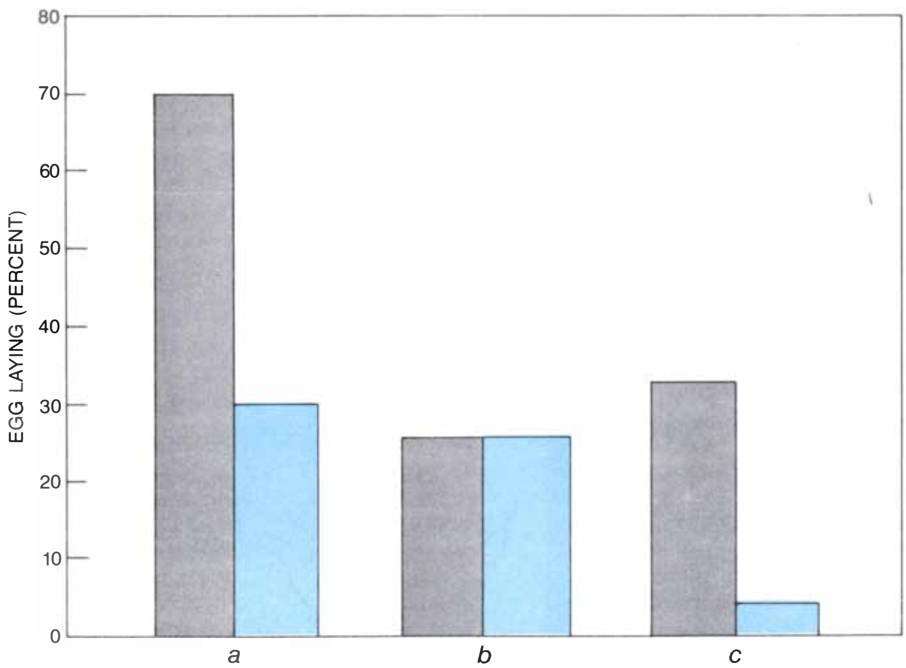
Among the insect-plant interactions I have examined in the hope of elucidating the role of leaf shape is the one between the *Heliconius* species *H. hewitsoni* and its host *Passiflora* species *P. pittieri*. This butterfly deposits its eggs in clusters rather than depositing single eggs at separate sites. Observing marked females, I found that after they had deposited their eggs on a plant they returned to it several times a day. Some of the marked butterflies divided their time between a half-acre patch with four host plants and a patch of cucurbit flowers 150 meters away, where they could feed on pollen and nectar.

I collected lengths of *P. pittieri* vine with newly developed shoots and placed them in the patch of vines that was being regularly revisited. The females did not discover or investigate the potential egg-laying sites I had supplied. Moreover, I had also placed abstract shapes behind the egg-deposition sites and elsewhere in the half-acre patch; the females did not approach them either. Apparently the butterflies had entered a secondary stage of foraging behavior, one in which they were devoting all their searching time to the reinvestigation of a host plant they had already discovered. If my abstract shapes had any influence at all, it was to provide landmarks for keeping track of where not to place additional eggs! How has such a two-stage searching pattern evolved? Evidently it arises from the vine's pattern of growth and from the resulting egg-deposition tactics of *H. hewitsoni*.

The *P. pittieri* vines, it so happens, produce new shoots rather infrequently. Accordingly the *H. hewitsoni* females' foraging behavior can be seen to consist of two equally critical phases. The first one may be termed the discovery phase: the young, inexperienced butterfly searches for plants with the appropriate leaf chemistry. The plant may or may not have appropriate egg-deposition sites at the time of discovery.

The second phase may be termed the decision phase: Exactly when should the female deposit its cluster of eggs on previously discovered and regularly visited host plants? If the eggs are deposited too early, the hatching caterpillars may devour the shoot before its leaves appear, and because they are too small to eat tough old leaves they will starve. If the eggs are deposited too late, the caterpillars may find that the rapidly growing leaves have already become too tough to eat; the consequences are equally fatal. It appears that because the *H. hewitsoni* female has evolved to deposit its eggs only on a single species of host plant (and perhaps also because the host plant is relatively large and therefore easy to find) these butterflies have specialized to devote almost all their effort not to a search for new host plants but to the selection of a proper "time window" for depositing their eggs on host plants they have already discovered. This species therefore tends to exploit cues more general than leaf shape for finding host plants.

The behavior of the *H. hewitsoni* females contrasts sharply with that of *H. cydno* females. The *H. cydno* females, which lay eggs singly but may deposit from six to 10 eggs in a day on any one of a variety of available *Passiflora* species, frequently select host vines that are too small to offer sites for laying egg clusters and are certainly too small to



COLOR DISCRIMINATION in *Heliconius* females was demonstrated in a series of experiments with passion-flower vines. When *H. cydno* females were presented with a choice (*a*) between a vine bearing no eggs (*gray bar*) and a vine bearing an egg (*colored bar*), in a total of 217 inspections the butterflies selected the egg-free site 70 percent of the time. To determine whether color or chemical cues govern this behavior the butterflies' next choice (*b*) was between a vine bearing no eggs (*gray bar*) and a vine bearing an egg that had been dyed green (*colored bar*). In a total of 80 inspections the butterflies showed no greater preference for the egg-free site. Finally the butterflies were offered a choice (*c*) between a vine bearing a green egg (*gray bar*) and a vine bearing a normal yellow egg (*colored bar*). In 66 inspections the butterflies selected the site with the green egg more than 30 percent of the time and the site with the yellow egg less than 5 percent. Where the percentages in bars do not add up to 100 percent, the remaining fraction is accounted for by inspections in which the butterfly did not lay an egg.

If you could reach the shutter release, it would work.

How does Olympus OM-1 work at temperatures like this? Flawlessly. From arctic cold to tropic heat, its torture-tested toughness is the envy of other SLR makers. Take the OM-1 shown above: if you could reach the shutter release, it would work!

Because the world's first SLR compact rejects gimmicks that degrade reliability. While lavishing infinite attention to qualities that count, for you and your photographs.

You see it in its incredibly large, brilliant viewfinder. Free of distracting clutter, to let you concentrate on creativity without compromise.

You sense it in its uncannily-precise **manual** meter that lets you control exposure, every time. With accuracy that's been proven by more than a **million** OM-1 owners.

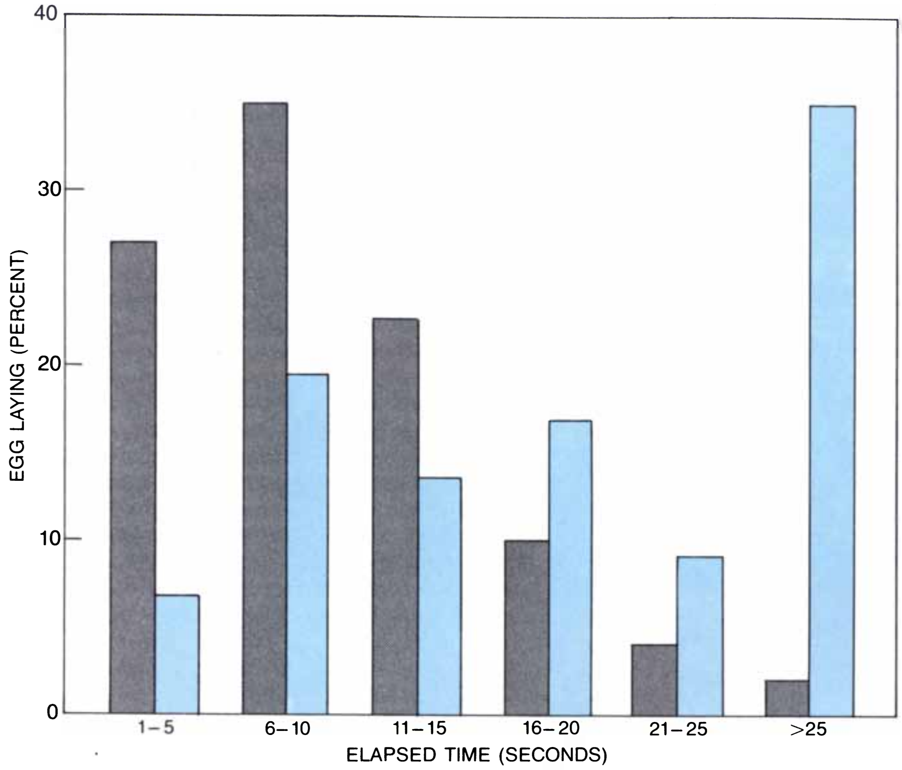
You hear it in its whisper-quiet shutter–working at **all** speeds, with no battery. Of which **Modern Photography** magazine reports, "It's as accurate as it is inventive."

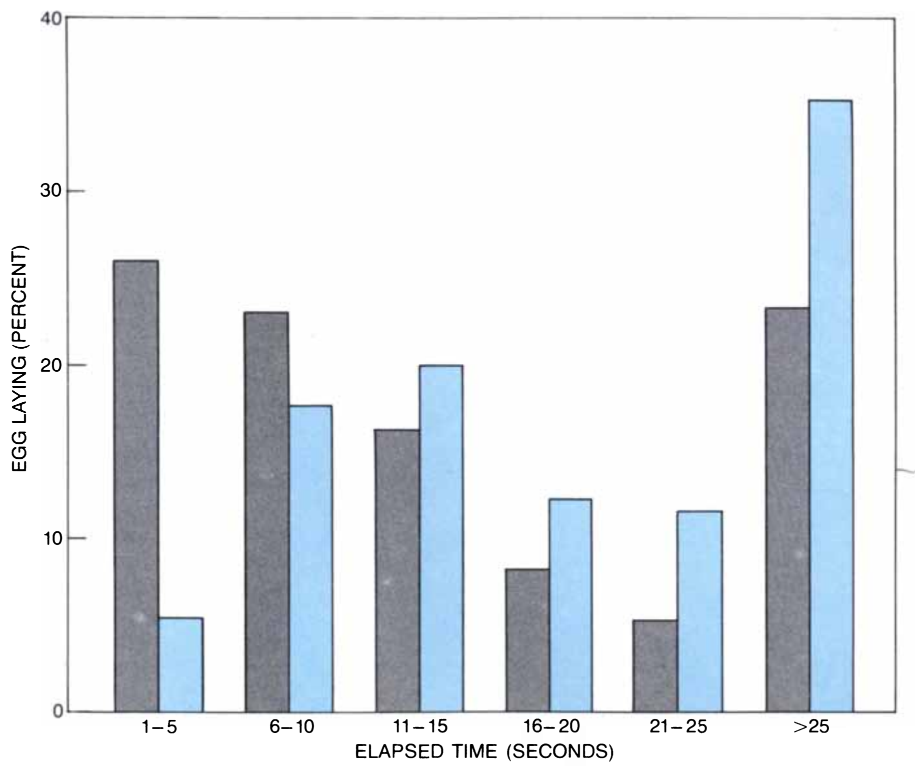
You feel it in its superbly-balanced **metal** body–with a ruggedness to match professional demands. And in its photographs–images of an excellence that is both legend and reality.

More than a camera, OM-1 is the heart of the largest compact system in all 35mm. With more than 300 quality-matched components, including dozens of matchless Olympus lenses.

See the OM-1 now at your Olympus dealer. You'll see the hottest SLR compact of all time. For details, write Olympus, Woodbury, NY 11797. In Canada: W. Carsen Co. Ltd., Toronto.

OLYMPUS® OM-1
THE CLASSIC





INHIBITORY EFFECTS of normal yellow eggs and mimic eggs were compared in a further experiment. As is shown in the top chart, the *H. cydno* females were first offered a choice between a vine bearing no eggs (*gray bar*) and one bearing a normal yellow egg (*colored bar*). After only 25 seconds 113 eggs had been deposited. Most of the eggs deposited in the first 15 seconds were placed on the egg-free vine; the mean time for deposition was 9.7 seconds. Most of the eggs deposited on the occupied vine were emplaced after 15 seconds had passed; the mean time was 21.8 seconds. As is shown in the bottom chart, in the next test the females were offered a choice between two lengths of the egg-mimicking vine *P. cyanea;* one length of vine had its mimic eggs removed (*gray bar*) and the other had the mimics in place (*colored bar*). After 25 seconds 180 eggs had been deposited. Most of the eggs deposited in the first 20 seconds were placed on the mimic-free vine; the mean time was 15.4 seconds. Most of the eggs deposited on the vine with mimic eggs were emplaced after 15 seconds had passed; the mean time was 21.9 seconds. Thus the inhibitory effect of the mimic eggs was equal to the effect of the real eggs.

feed a large number of emerging caterpillars. With as many as 10 eggs to deposit each day, the *H. cydno* females must continually search for new host plants in a forest habitat where the density of passion-flower vines is low. They have evolved a pattern of behavior that probably depends on the visual recognition of leaf shapes similar to those of host plants they have encountered previously. The large number of species exploited and the greater efficiency of visual recognition, as opposed to random drumming, mean that in the course of a day the females will find many egg-deposition sites to choose from.

The corollary is that if a species of passion-flower vine were parasitized by a butterfly that depended on the recognition of leaf shape as its main search stratagem, the target species would be under strong selective pressure to evolve both in the direction of (1) convergence toward plants other than *Passiflora* and (2) divergence away from other *Passiflora* species that can feed caterpillars. (If a plant shares a leaf shape with another plant, it may attract unwanted visitors. This conclusion is in accord with the hypothesis mentioned above, and it is supported by the observed variety of leaf shapes among species in local communities including numerous *Passiflora* species exposed to the same searching insects.)

We come now to the second aspect of the interaction between *Heliconius* butterflies and *Passiflora* vines. After the butterfly has selected a suitable host-vine site it must decide whether to lay its egg (or eggs) or to search further. The criteria for making this decision are obviously important to the success of the parasite, and yet there are only a few hints to their nature. For example, it seems reasonable to suppose one criterion would be the presence or absence of predators or competitors at the chosen site.

To .consider predation, the emerging caterpillars of most *Heliconius* species that deposit single eggs are cannibalistic. One may suppose, then, that a major criterion affecting the decision of a female of these species to deposit eggs or not to deposit them would be the presence or absence of another female's egg at the selected site. A mechanism favoring the avoidance of such sites could easily evolve because mutant butterflies with such a mechanism will have more progeny than butterflies that deposit eggs at occupied and unoccupied sites randomly.

Has such avoidance behavior actually evolved? The answer is yes; it is an observed fact that among *Heliconius* species that deposit single eggs the act of egg deposition is not random. Field studies show that vine shoots bearing only one egg outnumber those bearing two or more eggs in a proportion far

The $1795. Personal Business Computer that is changing the way people go to work.

Henry Ford revolutionized personal transportation.

Adam Osborne has done the same for personal business computing.

Virtually everything you need in a personal business computer to work faster and better with words, numbers, and ideas is included in a portable carrying case you can take anywhere.

The Osborne is easy to learn, fun to use, and quickly becomes indispensable.

One price, $1795, buys it *all.*

The Osborne is available from a computer retailer near you.

And you can buy it in any color you want. As long as it is blue.

Put away your buggywhips.

$1795. Complete. Including Software.

For your nearest dealer, call (in California) 800 772-3545 ext. 905; (outside California) call 800 227-1617 ext. 905.

OSBORNE
COMPUTER CORPORATION

The $1795 suggested retail price for the Osborne 1 (a trademark of Osborne Computer Corporation) includes a full business keyboard, built-in CRT display, two built-in floppy disk drives, CPU and 64 kilobytes of RAM memory, RS-232 and IEEE 488 interfaces, and the following software packages: WORDSTAR® word processing with MAILMERGE® (a trademark and a registered trademark of MicroPro International Corporation of San Rafael, California); SUPERCALC® electronic spreadsheet system (a trademark of Sorcim Corporation); CBASIC® (a registered trademark of Compiler Systems); MBASIC® (a registered trademark of Microsoft); and CP/M® (a registered trademark of Digital Research).

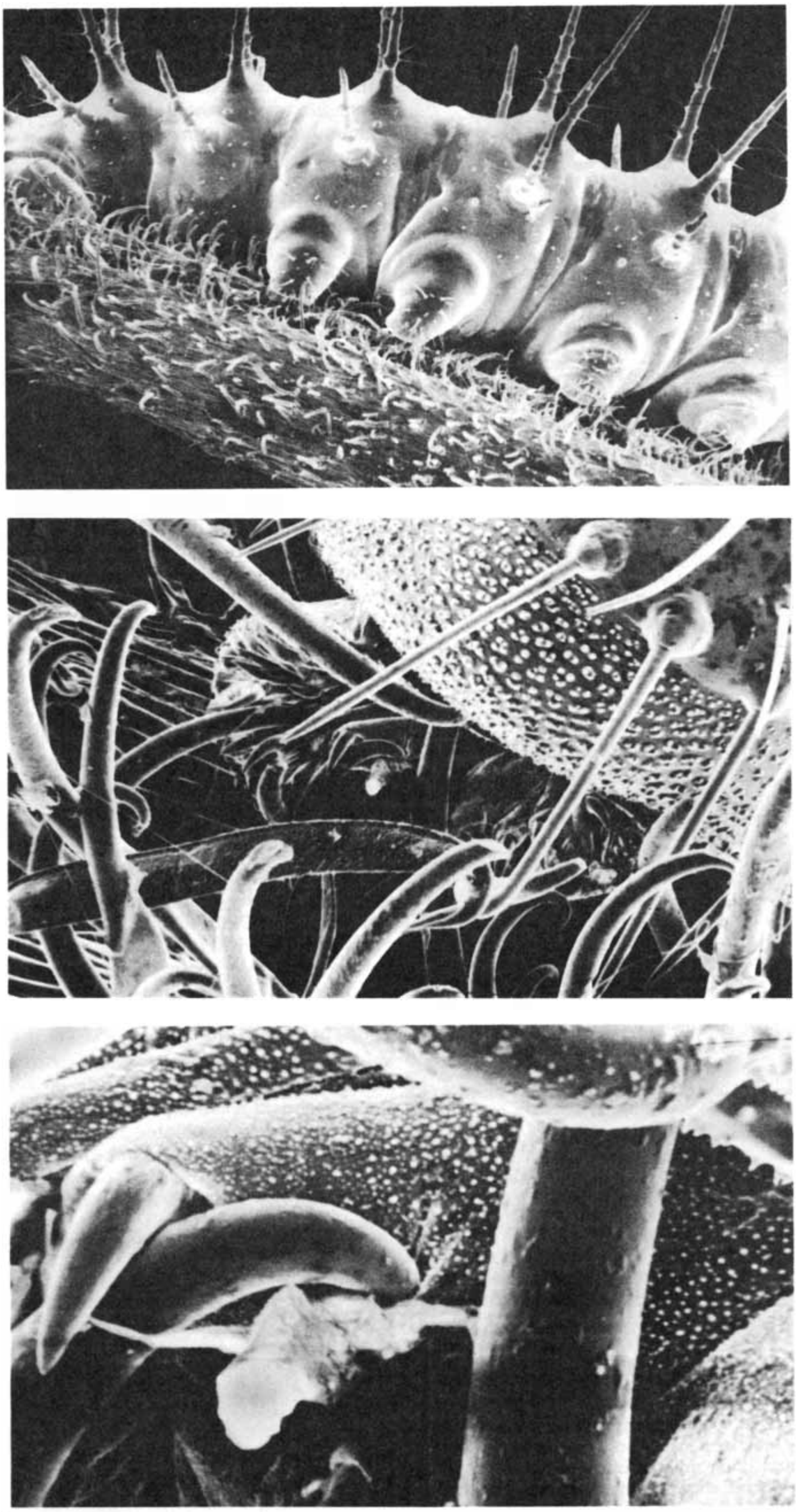
greater than statistics would predict for random events. Indeed, this pattern of behavior, technically termed "egg-load assessment," is a well-documented phenomenon not only among *Heliconius* species but also among other insects, including other families of butterflies.

Assuming that the avoidance behavior exists, what is the cue that gives rise to it? At least among the *Heliconius* species one potential cue stands out above all others: the females' eggs are deposited in conspicuous places such as the tip of a shoot and they are bright yellow. Until recently, however, the possibility that a color cue played a part in inhibiting egg deposition was supported only by circumstantial evidence.

Cues apart, how does this mechanism of egg-load assessment enter into the pattern of coevolutionary interactions between butterflies and passion-flower vines? To answer the question we must turn to a bit of history. In the late 1960's W. W. Benson (who is now at the University of São Paulo) and I were working in different parts of the New World Tropics and came to an identical conclusion. Both of us noticed that certain passion-flower vines (different species in our respective areas) bore yellow markings. Our hypothesis, arrived at independently, was that markings identical in color (and in many instances identical in shape) with *Heliconius* eggs would reduce the probability of any eggs' being deposited on the vine. The reason, of course, would be that it would appear eggs had already been deposited there.

This hypothesis was particularly exciting to me because few evolutionary developments in the plant kingdom can be clearly attributed to selective pressures exerted by clearly identifiable agents. In the years that followed, comparative studies of what became known as "egg mimicry" in various *Passiflora* species offered strong support for the new hypothesis. Although the apparent egg-mimicking features have so far been found in only some 2 percent of the 500 New World species of *Passiflora,* they arise in too many different ways for them to be correlated with some single botanical function. They also seem to have evolved sporadically among *Passiflora* lineages; they are found among some populations of *Passiflora* species or subgenera but not among others of the same species or subgenus. This pattern indicates that egg mimicry evolved recently: since the major radiation of passion-flower vines in the New World.

In 1976 Kathy S. Williams, an undergraduate student working in my laboratory, and I undertook to establish experimentally whether *Heliconius* butterflies were indeed the selective agents responsible for the evolution of mimic eggs on *Passiflora* vines. We isolated *H. cydno* females in a small greenhouse



CAPTURED HELICONIUS CATERPILLAR, one of its prolegs hooked by a sharply pointed hair on the leaf stem of a *Passiflora* vine, is seen in this series of scanning electron micrographs showing another defensive measure of some species of the plant. At the top the entrapment is not apparent; the hooked leg is the third from the right. In the middle the tip of the plant hair has penetrated the surface of the proleg at the center of the micrograph. At the bottom the tear that has been made in the proleg by the plant hair is visible just above the hook's point of entry. The vine in the micrographs is *P. adenopoda;* the caterpillar is *H. melpomene.*

where they could feed but where no passion-flower vines were available for egg deposition. As our first experiment we introduced four shoots of the vine *P. oerstedii,* two of them bearing a conspicuous yellow *Heliconius* egg and two bearing no egg. We then recorded the number of times the females inspected the shoots. Inspection was followed by egg deposition in 70 percent of the visits to egg-free shoots, whereas it was followed by egg deposition in only 30 percent of the visits to egg-bearing shoots. These results strongly supported the reality of egg-load assessment but did not exclude the possibility that the cue to whether or not the shoot was occupied by an egg was not a visual cue but an olfactory one.

To test the reality of color discrimination by the butterflies we repeated the experiment, this time giving them a choice among unoccupied shoots, shoots occupied by yellow eggs and shoots occupied by eggs that had been dyed green. (The eggs of both colors were rinsed in distilled water to eliminate the possibility that dyeing the eggs had washed off a repellent chemical.) The butterflies deposited eggs in equal percentages per inspection on the unoccupied shoots and the shoots with the green eggs. The percentage of egg deposition on the shoots with the rinsed yellow eggs was significantly the lowest: less than 5 percent compared with 25 percent for both the unoccupied shoots and the shoots with the green eggs. The results demonstrated that for the butterflies the major cue to the presence of an egg is its natural color.

Our third experiment was designed to test the related hypothesis that whereas in certain circumstances the butterflies will deposit their eggs on occupied shoots, there is still some residual inhibition against their doing so. The design of the experiment was simple. The *H. cydno* females were again isolated and presented first with egg-free *Passiflora* shoots and then with egg-bearing ones. The deposition of eggs took on the average more than twice as long to accomplish (22 seconds) on egg-bearing shoots as it did on egg-free ones (10 seconds). The extra 12 seconds were spent seeking a site on the egg-bearing shoot some distance from the site already occupied. This kind of discrimination may be rare in nature and may be at least partly attributable to crowded greenhouse conditions. Nevertheless, the existence of time-consuming search behavior leading to eggs' being placed in secondary sites on occupied shoots suggests that crowded shoots have been a factor in the evolution of *H. cydno* egg-deposition behavior.

Williams and I devised further experiments to test the effectiveness of mimic eggs in warding off egg deposition. The vine we worked with, the species *P. cyanea,* develops conspicuous yellow swellings on its stipules: paired leaflike appendages on the stems of the plant. When *H. cydno* females were given a choice between shoots bearing mimic eggs and shoots with the mimic eggs removed, their preference for the egg-free shoots was significant: 40 percent compared with 25 percent. Moreover, the average time that passed before a butterfly deposited an egg on a shoot bearing a mimic egg was almost exactly the same as the time we had measured earlier for shoots bearing actual *Heliconius* eggs. This suggests that whereas mimic eggs do not totally deter egg deposition, their effect in forcing the butterflies to select secondary deposition sites some distance from the tip of the shoot may at least reduce the number of eggs deposited at that vulnerable point.

Certain patterns of growth and development are unique to the *Passiflora* species that bear mimic eggs. For example, in passion-flower vines typical of the group called *Decaloba,* such as *P. biflora,* the growth of shoots is parallel with the development of leaves. One species of this group found in northeastern Peru is an exception. It has mimic eggs; the modified structures are nectar glands on its leaves. In this species the growth and development of the leaves are normal but the growth of the stem between nodes (the points at which leaves and other organs diverge from the stem) proceeds in pulses. The result is that the vulnerable shoot tip is always "protected" by a leaf bearing a prominent display of mimic eggs.

Some populations of the species *P. auriculata* bear mimic eggs; others do not. Among those that display mimic eggs the modified structures are the nectar glands of the leaf stem. These develop precociously, reaching mature dimensions while the developing leaf is still tiny. As a result the vulnerable tip of the shoot gives the appearance of being covered with large golden eggs.

Perhaps the most remarkable of these egg-mimicking modifications is found in the species *P. candollei.* The modified structures are the leaf nectar glands; they have the appearance of egglike yellow spots. Near relatives of this vine typically have their leaf nectar glands arrayed along two diverging lines, and the positions occupied by the nectar glands on one leaf are usually occupied on other leaves of the same plant. On the leaves of the egg-mimicking *P. candollei,* however, the developmental control of the position of the nectar glands on the leaf is apparently lost, so that the egglike glands appear haphazardly. Some leaves have only one yellow spot; others may have as many as eight spots. The numbers and positions vary from leaf to leaf on the same plant. This randomized system simulates the irregular pattern characteristic of egg laying in many species of *Heliconius* butterflies.

All these observations strongly suggest that the hypothesis of protective egg mimicry by passion vines is a valid one. The greenhouse experiments done by Williams and me also provide evidence in support of the hypothesis. Nevertheless, field studies will be necessary to determine whether the butterflies' capacity for egg-load assessment makes for significant differences in the survival rate and reproductive success of mimetic passion vines compared with nonmimetic ones in the same population.

This brings us to the third aspect of the interaction between *Heliconius* butterflies and *Passiflora* vines. What happens when the eggs are laid and what happens when they hatch? Have the passion-flower vines evolved traits to "deal with" the eggs and caterpillars of the parasite? I have identified in certain *Passiflora* species a few traits that are inimical to the parasites in these stages of their life cycle. One antiegg example can be seen in the species *P. ambigua.* The stipules of this vine are long and tendril-like; they are shed soon after they appear. When the stipules are shed, any eggs laid on them will perish.

A second trait that may have evolved as an anticaterpillar measure is found in *P. adenopoda* and its relatives. The leaf surfaces of these passion vines bear an array of fine hooked hairlike trichomes, and most *Heliconius* caterpillars are unable to move across the leaf without being pierced and fatally immobilized.

The observations and experiments I have described here represent only a few of the coevolutionary relations between *Heliconius* butterflies and *Passiflora* vines. They indicate, however, that the dual interaction between the animal parasites and their plant hosts is a dynamic, continuing process. It is certain that many of the differences among various *Passiflora* populations and species have arisen because of selective pressures exerted by the butterflies.

A decade ago I looked on these phenomena as a special and extreme case of the tropical environment in action. Now, however, the same phenomena are being discovered in the interactions between butterflies and plants in the Temperate Zone. One case in point is Rausher's observations of the pipe-vine swallowtail butterflies and *Aristolochia.* Another is the mimicking of butterfly eggs by certain plants of the mustard family recorded by Arthur M. Shapiro of the University of California at Davis. In any event the presence in the New World Tropics of *Passiflora* species numbering in the hundreds and local populations numbering in the millions provides a natural laboratory for the study of coevolution in all its genetic and ecological ramifications.

Ophiolites

The crust of the earth under the oceans is different from the crust of the continents. Ophiolites seem to be fragments of oceanic crust on land. They are thus clues to how oceanic crust forms and spreads

by Ian G. Gass

Plate-tectonic theory proposes that oceanic crust, which forms some 70 percent of the earth's solid surface, has been and is being constantly created at the axes of oceanic ridges and rises. Thereafter, by the processes of sea-floor spreading, it moves away from these axes and ultimately plunges into the earth's interior along subduction zones. Since the oceanic crust has this built-in self-destruct system, no part of the present oceanic crust is much more than 200 million years old. In contrast, the continents, being lighter than oceanic crust, are not easily subducted and move passively over the face of the earth in response to sea-floor spreading and plate-tectonic processes. Indeed, continental rocks preserve evidence of earth history going back almost four billion years.

Occasionally a fragment of oceanic crust, instead of being subducted, is preserved at the leading edge of a plate that rides over a subduction zone. To describe this process Robert G. Coleman of the U.S. Geological Survey has coined the term obduction: the reverse of subduction. Coleman has also calculated that less than .001 percent of all oceanic crust has been obducted and remains on dry land. Small though these remnants are, they provide unique information on the processes currently operating under the axes of oceanic ridges and rises. They also yield clues to the evolution of ancient oceans, the mechanisms of collisions between plates and the position of ancient destructive plate margins, and they strongly suggest that plate-tectonic processes have been active for at least the past billion years. These on-land fragments of oceanic crust are known as ophiolites.

Like many other geological terms "ophiolite" has since the advent of plate tectonics taken on a new meaning and significance. Even before this latest revolution in earth science, however, the meaning of the term evolved with changes in the understanding of geological processes. The term first appeared in the geological literature in the 1820's, when Alexandre Brongniart of France coined it to describe rocks also known as serpentinite or serpentinized peridotite: a type of igneous rock usually found in areas deformed by tectonic processes. "Ophiolite" is derived from the Greek *ophis,* meaning snake or serpent; therefore it has the same meaning as serpentinite. Both terms are appropriate only in that both the rocks and some reptiles have a mottled green appearance. Other than demonstrating the addiction of European geologists to the classical languages, "ophiolite" had little significance and was used in an ad hoc way throughout the 19th and early 20th centuries to describe serpentinized peridotites and the rocks commonly associated with them.

In the 20th century there were two major changes in the usage of the term. In 1906 Gustav Steinmann of Germany noted the close association of serpentinized peridotite with other igneous rocks and deep-water sediments (such as radiolarite) in the Alpine fold mountains around the Mediterranean. Later, in honor of this outstanding geologist, the association of serpentinite, radiolarite and pillow lavas (lavas erupted under water) became known as the Steinmann Trinity. Then, through common usage, "Steinmann Trinity" and "ophiolite" became synonymous. In other words, "ophiolite" no longer referred to one kind of rock but to an assemblage of related rocks.

Steinmann particularly emphasized the association between deep-water sediments and the serpentinite and the pillow lavas. Others developed this lead and proposed that ophiolites were masses of igneous rock emplaced in geosynclines: huge linear depressions in the earth's crust that become filled with sediments. Some workers believed the igneous rocks were intruded into the layers of sedimentary rock as sills (horizontal intrusions); others saw them as immense balloons of magma (molten rock) erupted onto the surface of the sediments along the flanks of the geosyncline.

It was visualized that once the balloon had erupted its skin chilled and fractured and was invaded by dikes (vertical intrusions) from the balloon's still-molten interior. Thereafter pillow lavas erupted onto its surface. As the molten interior crystallized, the heavier minerals settled to produce layered peridotites and overlying layered gabbros.

Here I must pause briefly to define what is meant by terms such as "peridotite" and "gabbro." In this context they apply to the rocks of the oceanic crust and the underlying uppermost part of the earth's mantle. The uppermost part of the mantle is thought to be formed of peridotite, a rock that consists almost entirely of the magnesian minerals olivine $[(Mg,Fe)_2SiO_4]$ and pyroxene $[Ca(Mg,Fe)Si_2O_6]$. Also present, however, is dunite, a rock that consists almost entirely of olivine.

The oceanic crust is formed of rocks, such as basalt and gabbro, that are somewhat richer in silica (silicon oxides). Basalts are fine-grained; gabbros, having crystallized more slowly, are coarse-grained. Both, however, consist

OPHIOLITE ON THE ARABIAN PENINSULA is represented by the dark-colored mountains running from the upper left to the lower right in this Landsat picture of northern Oman. The body of water at the upper right is the Gulf of Oman, which connects the Persian Gulf and the Indian Ocean. The mountainous area is known as the Samail nappe. Light-colored area to the left of the dark-colored ophiolite is continental rocks, mainly limestones; they are underlain by granitic rocks of the Arabian continental plate. The dark color of the ophiolite results from the abundance of basalts, gabbros and peridotites, rocks characteristic of the oceanic crust and the underlying mantle of the earth. The small patches of red along the coast and in other areas are vegetation, which appears red because of the arbitrary color coding of the wavelengths detected by the sensors carried on the Landsat satellite. The Landsat image data have also been further processed by computer in order to enhance the image. The processing was done by the Earth Satellite Corporation. The field of view in the picture is 130 kilometers across.

of the same minerals: olivine, pyroxene and plagioclase ($NaAlSi_3O_8$–$CaAl_2Si_2O_8$). Basalts form the upper one to 2.5 kilometers of the oceanic crust (exclusive of the overlying sediments), gabbros the lower 3.5 to 6 kilometers.

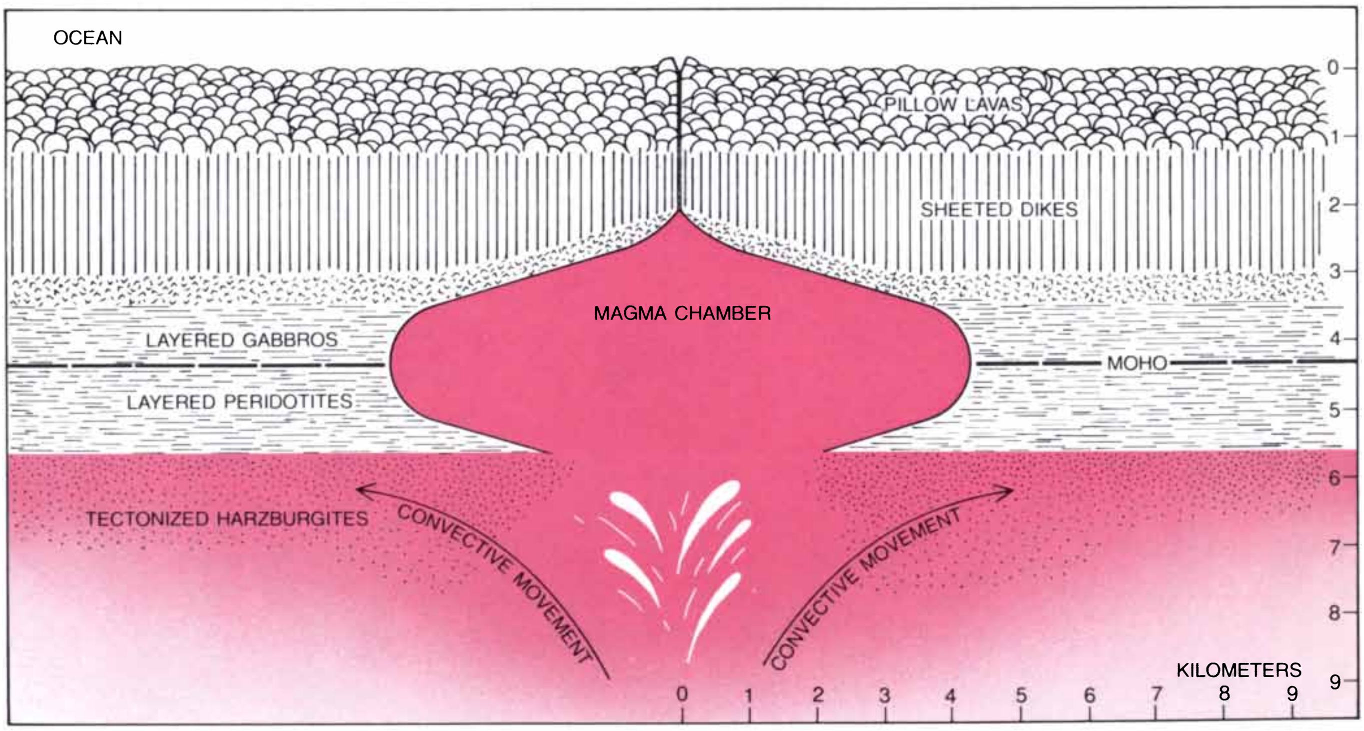
To return to the geosyncline model of ophiolites, it developed in the 1930's, 1940's and early 1950's and remained the consensus among geologists until the mid-1960's. The essence of the model is that ophiolites were interpreted as being the result of magmatism during the initial stage in geosyncline development. They were therefore in situ (autochthonous) igneous rocks and, whether they were intruded as sills or erupted, they were interleaved with the sedimentary rocks of the geosyncline. It is this basic concept that has changed. Today almost all ophiolite masses are regarded as being allochthonous, that is, they were formed elsewhere and were transported tectonically to their present position.

The model now widely accepted is that an ophiolite is a fragment of oceanic crust formed at the axis of an oceanic ridge or rise, moved across the ocean floor by sea-floor spreading and finally lifted above sea level. Concurrently with this switch from an in situ to a transported origin for ophiolites the concept of the geosyncline, which once dominated the geological literature, has been virtually abandoned.

The complete change in ideas on the origin and significance of ophiolite complexes was brought about by a variety of investigations. First it was demonstrated that virtually all ophiolites are allochthonous, having been brought in contact with the adjacent rock formations by tectonic processes. Second, detailed studies of ophiolites in the eastern Mediterranean, particularly the Troodos massif on the island of Cyprus, showed that their internal structure was not compatible with their having been formed in situ in a geosyncline; it could be realistically explained only on the basis of magmatic processes operating at oceanic ridges and rises, where new oceanic crust is generated. Third, it was noted that in each of the eastern Mediterranean ophiolites the same rock sequence could be recognized and could be compared with layers in the sea floor deduced from geophysical evidence. Fourth, it was shown that ophiolite rock types were similar to those of rocks that had been dredged from the deep-sea floor.

The similarity of ophiolite sequences to one another and to oceanic sequences and the comparability of oceanic and ophiolitic rocks (demonstrated by Nikolas I. Christensen of the University of Washington and Matthew H. Salisbury, now at the Scripps Institution of Oceanography) led to the general acceptance of ophiolites as fragments of oceanic crust formed at oceanic ridges and rises. Since then it has been common practice to use ophiolite data to clothe the necessarily meager skeleton of oceanographic data on oceanic-ridge-and-rise structures. This view became widely accepted in the 1970's, but even so when those attending the Geological Society of America's Penrose Conference on ophiolites undertook to redefine the term in 1972, they stressed that it should be applied only to a particular rock association and should not have any connotations of origin.

A complete ophiolite, such as the relatively undeformed ophiolite of the Troodos massif of Cyprus, is an orderly sequence of rock types. In most instances the processes of obduction have disrupted the originally coherent mass and have spread parts of the ophiolite over a large area. The component parts of such a dismembered ophiolite can often be fitted back together by unraveling the tectonic jigsaw puzzle. It is easier, however, to study relatively undeformed ophiolites such as the Troodos



SCHEMATIC CROSS SECTION OF OCEANIC CRUST is based on ophiolite studies. In the middle is the axis of an oceanic ridge or rise. At the top of the crust (except for a thin layer of sediment that is not shown) are pillow lavas: lavas that have flowed out onto the ocean floor from a high-level magma chamber. Below the pillow lavas are sheeted dikes: vertical slabs that were originally intruded at the axis of the ridge or rise and were then moved outward along with the spreading sea floor. Below the sheeted dikes are coarse-grained, unstructured gabbros representing melts that crystallized against the roof of the magma chamber. Crystallizing within the chamber are layered gabbros and peridotites; the "Moho" (for Mohorovičić discontinuity) is the boundary between them. Below the layered peridotites are tectonized (deformed) harzburgites. As the pressure decreases upward the peridotite of the mantle begins to melt and the basaltic liquid so formed collects into balloonlike masses called diapirs. As the diapirs rise with the convecting mantle the mineral olivine precipitates out of them, but the liquid remaining in the melt escapes upward to replenish the magma in the high-level chamber.

massif and the Samail nappe in the Sultanate of Oman on the Arabian peninsula. Therefore it is on these studies that much of the following generalized description is based.

Starting from the top and going down, most ophiolites are overlain by marine sediments. In some instances, as with the Cyprus and Oman ophiolites, the sediments are deep-water ferromanganoan mudstones or radiolarites. In others the sediments are similar to those on continental shelves or adjacent to arcs of volcanic islands. Therefore they indicate that the oceanic crust, from which the ophiolite was derived, must have been near the margin of a continent or of an island arc at some stage of its history. More than anything else these sediments provide evidence on the environment of the ophiolite within an ocean basin before it was lifted above sea level; in this regard studies such as those of Alistair Robertson of the University of Edinburgh and Alan Gilbert Smith of the University of Cambridge have proved particularly valuable.

The top of the ophiolite proper consists of extruded basalts. Many of these rocks have the pillow forms characteristic of lavas that have erupted under water. Indeed, such lavas have been filmed in the act of formation by divers during a submarine eruption off Hawaii. Pillow lavas have also been repeatedly identified by manned and unmanned submersibles investigating the floor of the axial rift of the Mid-Atlantic Ridge and the axial zone of the East Pacific Rise.

Some lava pillows are nearly spherical; others are more elongated and might better be termed bolsters. The form of the pillows probably depends on the configuration of the sea floor onto which the lava is erupted. On steep slopes the melt forms globules that roll down the slope and accumulate at the base. The elongated pillows probably form on gentler slopes, where the lava flows downslope for some distance before solidifying. When the lava is erupted in hollows or on flat surfaces, it is unlikely to be pillowed, and unstructured layers of varying thickness seem to form.

Pillow lavas are known to form in very deep water, in shallow water and even where lavas erupted on land flow into water. How much water was above them at the time of their formation is indicated by their content of gas vesicles, or bubbles. Such vesicles are formed when the gas dissolved in a molten rock separates from the melt as the pressure is lowered. If the pressure of the overlying water is sufficiently high, however, the gas does not separate and no vesicles form. By measuring the vesicularity of basalts from known depths James G. Moore of the U.S. Geological Survey has shown that there is a crude

PILLOW LAVAS are seen in the Wadi Jizzi of the Samail nappe in Oman. These elongated pillows, which are more like bolsters, were probably erupted over a gentle slope on the ocean bottom and flowed for a short distance from the top right to the bottom left before solidifying.

SHEETED-DIKE COMPLEX is seen in the Samail nappe. Each dike, originally an intrusion of molten rock, is a vertical sheet of rock. Each was intruded upward along the axis of an oceanic ridge or rise. The outward convective movement of the underlying mantle (*see illustration on opposite page*) produces tension, and when a tension crack opens, magma from the chamber escapes upward and another dike is formed. The dikes are on the average one meter thick. Therefore the sea floor spreads in one-meter jumps every 50 to 100 years rather than in the smooth one to two centimeters per year visualized in the models of theoretical geophysics.

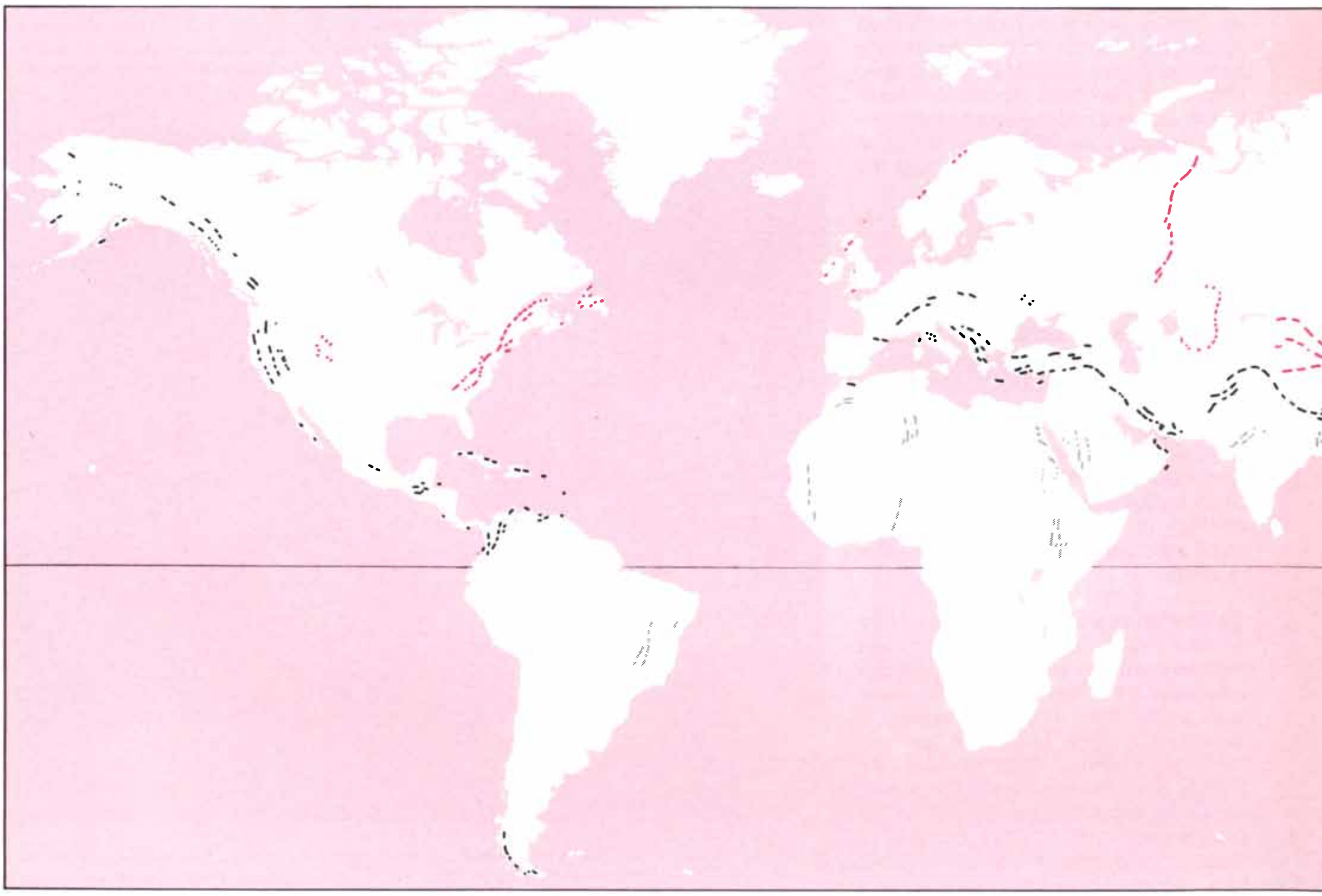
correlation between the depth below the surface of the water at which a lava was erupted and the vesicularity of the lava; with increasing depth (pressure) the vesicularity decreases. This can give only an approximate estimate of the depth because the abundance of the gas in the magma varies. Moreover, sea-floor spreading can move a lava to a depth other than the one at which it was erupted. Still, vesicularity is a further clue to the origin of ophiolites at oceanic ridges and rises.

Most of the lavas in ophiolites are basalts, similar to those dredged from present-day oceanic ridges. In Oman, however, it is only the lower part of the pile of lavas and the dikes under it that resemble the oceanic-ridge basalts in chemical composition. The upper part of the lava sequence differs geochemically from the lower, and Julian A. Pearce and Tony Alabaster of the Open University in England have convincingly demonstrated that these upper lavas were erupted in an island-arc environment later than the main event of sea-floor spreading. In Cyprus all the lavas have a composition comparable to that of volcanic rock erupted above a subduction zone.

The top of the lava pile consists entirely of extruded rocks. Farther down dikes become increasingly numerous. Many of the dikes are sinuous and seemingly had to wriggle their way upward through the lava pile. The thickness of the lava pile varies, but in the Troodos massif it is about one kilometer and in Oman it ranges between .5 kilometer and 1.5 kilometers. Near the base of the sequence the ratio of dikes to lavas is about 1 : 1; then, within a downward distance of 50 to 100 meters, the abundance of dikes increases from 50 to 100 percent and there is little or no lava between them. Moreover, the disposition of the dikes is not sinuous but a regular and often almost vertical array. Indeed, this part of the ophiolite has been likened to a pack of cards standing on edge. It has been termed the sheeted-dike complex.

In the 1950's, before the advent of plate tectonics, it was the sheeted-dike complex that presented the main conceptual problem in ophiolite studies. Nowhere else had a rock complex consisting entirely of dikes been found. In classic fossil volcanic areas such as the Hebridean province of northwestern Scotland the abundance of dikes with respect to the total outcrop is less than 10 percent. Therefore the search was always on for the host rock in which the dikes had been emplaced. Any rock that looked slightly different was studied in detail; the commonest contenders as host to the dikes were structureless flows of basalt. No one, as far as I am aware, then thought of the possibility that there was no host rock at all. Only after the 100 percent dike structure of the sheeted-dike complex had been demonstrated in the undeformed and well-exposed Troodos massif was it sought and identified in more deformed masses.

When the plate-tectonic model was proposed, it was quickly realized that the axis of an oceanic ridge or rise was just the place where a 100 percent dike complex would form. The current concept is that most of the dikes are injected along a narrow zone, no more than 50 meters wide, at the axis of the ridge or rise and that the dike material is moved away from the axis by sea-floor spread-

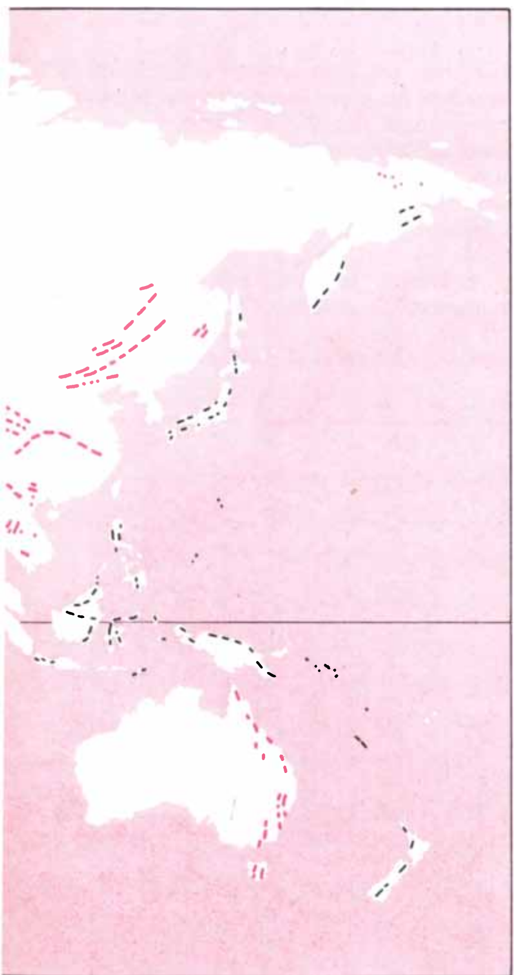


WORLDWIDE DISTRIBUTION OF OPHIOLITES is given on this map. The lines in black are ophiolites less than 200 million years old. The lines in color are ophiolites between 200 and 540 million years old. The lines in gray are ophiolites between 540 million and 1.2 billion years old. The younger ophiolites are those related to the present cycle of plate tectonics. They include the ophiolites emplaced around the Pacific, and all are close to the sites where oceanic crust is being subducted. The next-oldest ophiolites (running along the Appalachians, north into Nova Scotia and Newfoundland and continuing into Ireland, Scotland and Norway) mark the closing of the Iapetus Ocean

ing. Recent studies of sheeted-dike complexes, unencumbered by preconceived geological ideas, show that this part of an ophiolite complex does consist of 100 percent dike.

Johnson R. Cann and Rupert G. W. Kidd, working at the University of East Anglia, identified within the dikes the phenomenon of one-way chilling, which can be explained as follows. Magma being injected into a fissure in cold rock will form a dikelike body. The melt will cool most rapidly, and will therefore form the finest-grained rock, where it is in contact with the cold host. Cann and Kidd recognized that in the Troodos massif many dikes, instead of having two chilled margins, have only one. In addition, more of the chilled margins were on one side of the dikes than the other; this is the phenomenon of one-way chilling.

Cann and Kidd proposed that at an oceanic ridge magma would tend to be injected along the still liquid or softer axis of an earlier dike and so would split the two halves of the earlier dike, with one half moving in one direction and the other half in the opposite direction. On each side of the ridge the dikes would show a preferential one-way chilling on the side farther from the ridge axis. On this basis it has been inferred that the oceanic ridge where the north-south dikes of the Troodos massif formed lay to the west of the present-day outcrop. It should be emphasized that the statistical excess of one-way chills in one direction over those in the other is small and also that continuous outcrops with a statistically valid number of dikes are found only rarely. One-way chilling is an attractive concept and one that seems intuitively acceptable, but it is by no means proved.



in the Paleozoic era. The ophiolites in the U.S.S.R. also mark a Paleozoic plate-collision suture. Not all the ophiolites marked have been positively identified. For example, most ophiolites in China are not well described.

Although the sheeted-dike complexes are the best evidence that ophiolites formed along zones of tension in oceanic ridges and rises, such complexes are not always present. For example, in the Voúrinos ophiolite in Greece, as it has been described by Eldridge M. Moores of the University of California at Davis, sheeted dikes are poorly developed. It has been suggested that sheeted-dike complexes form only when sea-floor spreading is slow enough to allow a dike to solidify before the next one is injected. Since oceanic crust cools rapidly and seawater is percolating through all this material, the explanation is questionable. Moreover, the Oman ophiolite, the fossil of an oceanic ridge where the sea floor spread in both directions at the high rate of two centimeters per year, has a superb sheeted-dike complex. Although the absence of sheeted-dike complexes from some ophiolites remains a geological problem, it is worth noting that many ophiolites originally described as being without a sheeted-dike complex have been found on closer examination to have one.

The dikes of a sheeted complex must have come from an underlying source of magma. It is therefore not surprising to find that with depth the basaltic dikes give way, over a distance of between 10 and 100 meters, to gabbroic rocks that have a similar composition but a markedly coarser-grained texture. Detailed mapping of these plutonic complexes (igneous rocks formed well below the surface) suggests that the uppermost gabbros, which form a layer between 10 and 300 meters thick, were formed by the melt cooling and crystallizing against the roof of a chamber from which the magma flowed toward the surface. Below this layer the gabbros and the underlying peridotites, if there are any, are markedly layered. Until recently it had been widely accepted that the layering was the result of the minerals' crystallizing out of the melt and then settling to the bottom of the magma chamber.

Alexander R. McBirney and Richard M. Noyes of the University of Oregon have proposed an alternative mechanism as a result of studying the classic layered gabbros of Skaergaard in eastern Greenland. They suggest that gradients of chemical composition and heat cause minerals to crystallize out of the melt along horizontal planes within it; no movement of the crystals is needed to account for the layering. Whether the crystals settle or form layers in place, however, the inference of the layered plutonic rocks is that there is a large body of magma below the surface along the axis of an oceanic ridge or rise. Work on the Troodos massif by Cameron R. Allen of the University of Cambridge has suggested that along this slow-spreading (one centimeter per year) axis there were numerous magma chambers four or five kilometers in diameter. For the faster-spreading (two centimeters per year) Oman ophiolite the magma chambers seem to have been some 20 kilometers in diameter.

The evidence of these layered plutonic rocks is that below the axis of the oceanic ridges there were magma chambers whose dimensions depended essentially on the input of heat into the ridge from the underlying mantle of the earth and on the rate of cooling brought about by sea-floor spreading. But do these layered rocks represent a single body of magma crystallizing completely or, as seems intuitively more likely, was the magma body fed periodically from below? Studies by E. Dale Jackson of the U.S. Geological Survey on the Voúrinos mass in Greece, by Cameron Allen on the Troodos massif and by John D. Smewing and Paul Browning of the Open University on the Samail nappe in Oman show that in all these ophiolites the melt was periodically replenished by new batches of magma.

These workers, analyzing separate ophiolites, have shown that the layered sequences consist of repetitive cycles starting with dunite and giving way upward to peridotites and then to gabbros. The cycle starts again and is repeated many times throughout the layered sequence. The inescapable conclusion is that the composition of the melt in the main chamber was reset periodically by a fresh influx of magma.

At the base of all complete ophiolite sequences is a tectonized (deformed) peridotite consisting almost entirely of the minerals olivine and orthopyroxene [$(Mg,Fe)SiO_3$]. Peridotites of this composition are called harzburgites, and so they are described as tectonized harzburgites. The chemical and mineralogical homogeneity of the tectonized harzburgites suggests that they represent material of the uppermost part of the mantle from which basaltic liquid has been extracted. This proposal is supported both by the presence within the harzburgites of gabbro pods that are best explained as batches of basaltic melt that crystallized before they could escape from the mantle and by localized patches of peridotite of a different composition (lherzolites) that probably represent remnants of mantle from which little or no basaltic magma had been extracted.

Most workers accept that the tectonized harzburgite represents depleted mantle, the residue from which basaltic melts have been extracted to form the overlying rocks: from the bottom up the layered plutonic rocks, the sheeted-dike complex and the sequences of lavas. The model proposed to explain these features is that at depths of 25 or 30 kilometers basaltic magma first separates from the mantle. The two components, the magma and the harzburgite, move upward under the influence of convection in the mantle.

The rising magma collects into balloonlike bodies called diapirs, one kilometer to five kilometers in diameter. The diapirs move upward with the ascending mantle material, and as they do so olivine crystallizes out of the melt to form a crystal layer at the bottom of the diapir. The melt escapes into the main magma chamber but the olivine remains in the mantle as lenses of dunite within the tectonized harzburgite. Although the harzburgite is hot, it remains solid. Therefore it is deformed as it moves upward and then outward under the axis of the oceanic ridge or rise.

The available evidence suggests that the newly formed oceanic crust is transported by this convective movement in the underlying mantle. Detailed investigations, particularly those of Adolphe Nicolas and his colleagues at the University of Nantes, support that view. They also demonstrate that the deformation took place at about 1,000 degrees Celsius and imposed a linear fabric, perpendicular to the orientation of the overlying ridge axis, on the tectonized harzburgite.

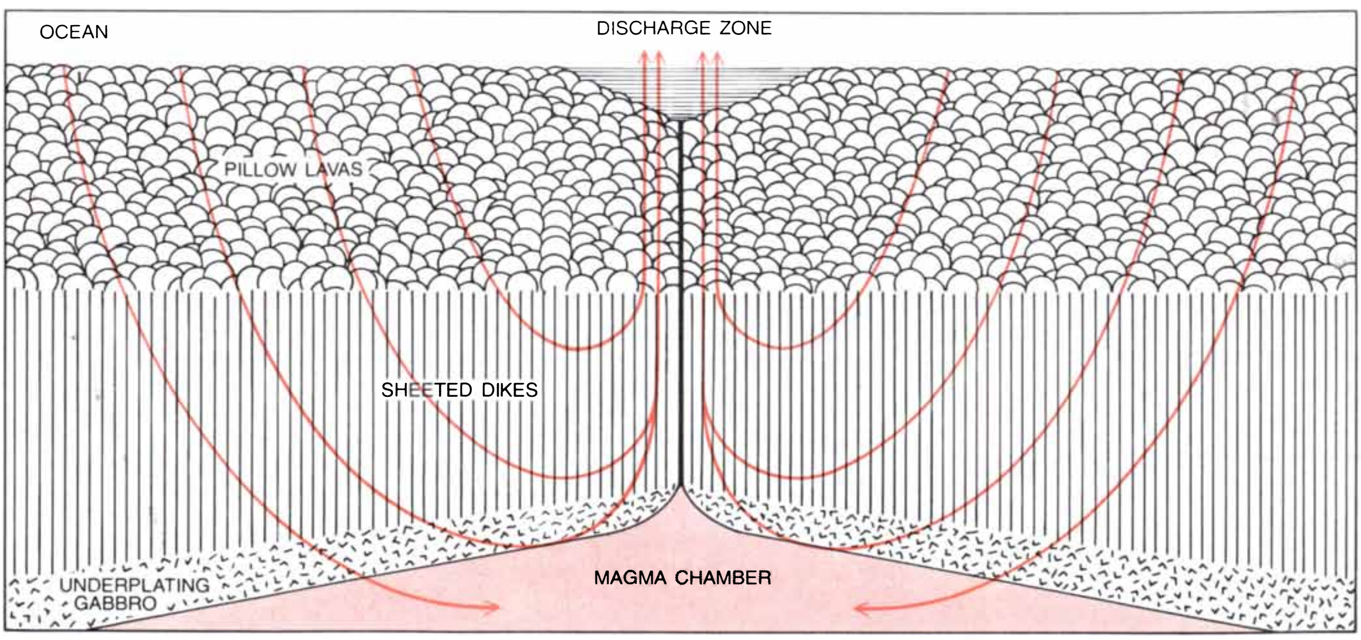
Ophiolite studies have thus revealed much about the structural and magmatic features at fossil oceanic ridges and rises. What else do they reveal about the oceanic crust? Oceanic ridges and rises are repeatedly (on the average of every 30 kilometers) offset by fractures. These fractures, known as transform faults, make it geometrically possible for rigid plates to move over the face of a nearly spherical earth. Even if ophiolites represent only .001 percent of the subducted oceanic crust, one of them, if it preserves more than 30 kilometers of oceanic ridge, is also likely to preserve a transform fault. Indeed, transform-fault structures have been investigated on Cyprus by my former colleague Kapo Simonian, on Masirah Island in Oman by Ian Abbott and Frank Moseley of the University of Birmingham and in western Newfoundland by John F. Dewey and his colleagues at the State University of New York at Albany. The features displayed by these structures at various levels revealed by erosion enhance the understanding of present-day transform faults.

Another feature of ophiolites also relates them to oceanic crust: their metamorphism, that is, the fact their rocks have been greatly altered since they originally formed. Virtually all igneous rocks dredged or drilled from oceanic crust away from the axes of oceanic ridges and rises have been metamorphosed. These oceanic metamorphic rocks, unlike most continental metamorphic rocks, show no directional fabric. Therefore they were altered without the deformation that commonly accompanies continental metamorphism.

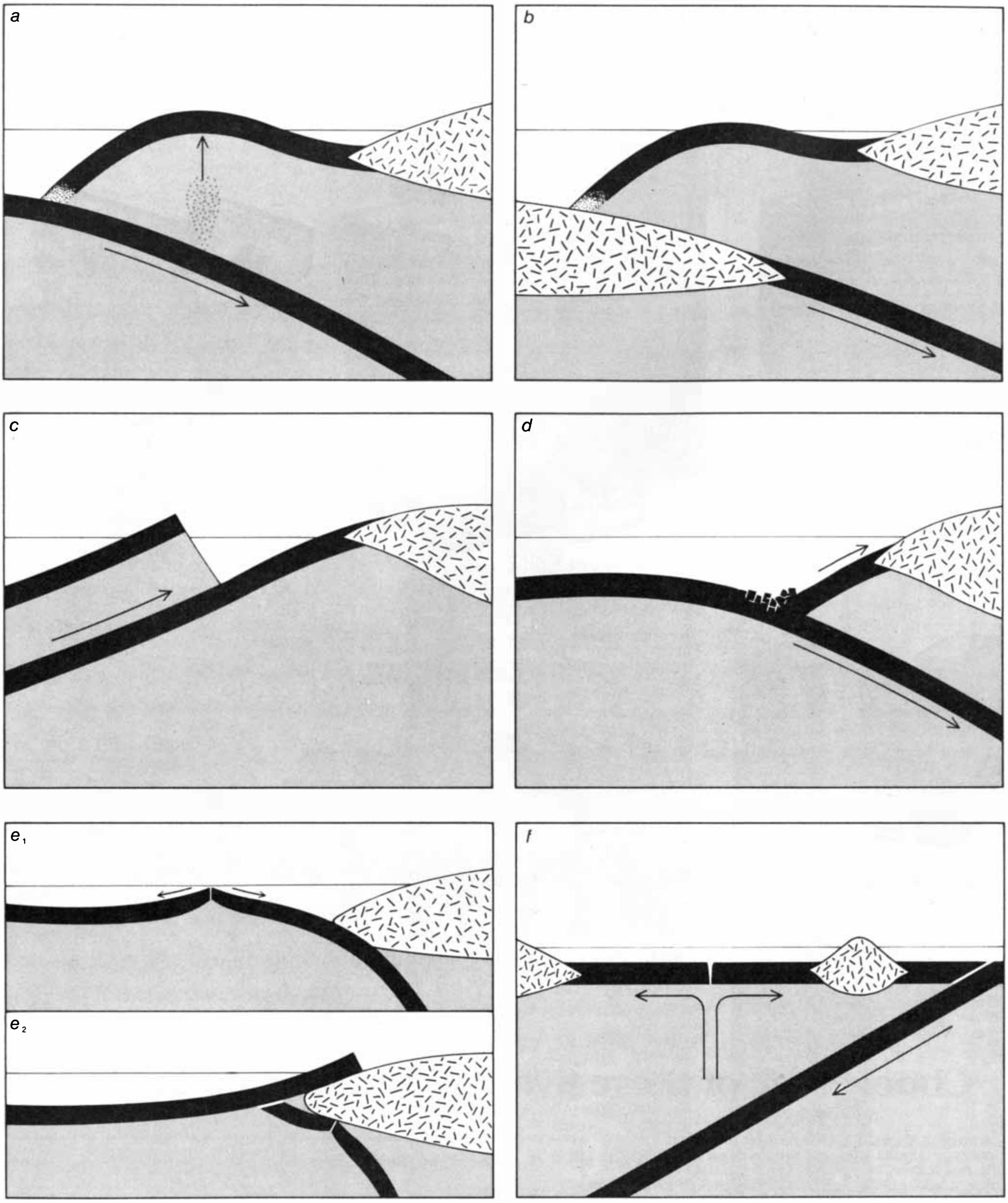
Since the metamorphic processes operated while the ophiolites were part of the oceanic crust, and since there are no directional fabrics, the main agent of metamorphism must have been heat. The main source of heat is the underlying magma. The metamorphic processes involved the circulation of seawater through newly formed, still-hot oceanic crust with a thermal gradient in excess of 150 degrees C. per vertical kilometer. The water is believed to circulate by convection in a single-pass cycle, with the permeable rocks of the oceanic crust being recharged with water continuously over a wide area and discharging water above thermal highs.

The convecting seawater leaches metals out of the oceanic crust and also redistributes silicon and other elements. In returning to the surface the metal-enriched fluids are channeled along faults, and so the emission of the fluids into seawater is localized. In favorable sheltered locations, such as depressions in the sea floor, chemical reactions between the hot, metal-enriched brines and the seawater lead to the precipitation of sulfides of iron and copper and the formation of massive bodies of sulfide ores, and also to the precipitation of iron with manganese to form ferromanganoan sediments. Hydrothermal vents emitting metal-rich brines have now been observed along the East Pacific Rise from manned submersibles.

Interesting developments in the understanding of plate-tectonic and related processes have resulted from these metamorphic studies. For example, it has long been held, particularly by Nikolas Christensen and Matthew Salis-



SEAWATER INFILTRATES HOT SEA FLOOR in the immediate vicinity of the spreading axis. The water is heated as it moves through the newly created oceanic crust and incorporates those elements that go readily into solution. The hot and now enriched brines return to the surface of the ocean bottom along fracture zones. The dissolved elements are precipitated out where the hot water reenters the ocean.



OPHIOLITES MAY BE EMPLACED on dry land by various processes, none of which is fully understood. Six possible models are depicted here. In model *a* oceanic crust (*arched band in middle*) attached to a continent or to an island arc (*lens-shaped body at right*) has been uplifted when other oceanic crust is subducted under it. Water from the subducted oceanic crust alters the mantle material (peridotite) above it into one that is less dense and therefore lifts up the overlying oceanic crust. In model *b* continental crust is subducted under oceanic crust; since the continental crust is lighter than the overlying material, it tends to rise and in so doing lifts anything above it. In model *c* oceanic crust moves upward (*arrow*) along a zone inclined away from the continental mass. In model *d* oceanic crust fractures and breaks up so that blocks of it fall into a nearby oceanic trench; such blocks of oceanic crust are found in mélanges with a matrix of serpentinite or muddy sediment. In model *e* the oceanic rise moves toward a subduction zone (e_1). Since the crust is thin over the magma chamber (*see illustration on page 124*), the part of the crust away from the subduction zone (to the left of the spreading axis), instead of being subducted, is sufficiently buoyant to ride up over the continental margin (e_2). In model *f* a more regional picture is depicted. It is believed most ophiolites represent oceanic crust produced by seafloor spreading in a marginal sea above a subduction zone. Emplacement of the ophiolite at the margin of a continent or an island arc could result from any of these processes or a combination of them.

bury, that the horizontal layering in the oceanic crust, identified through the velocity of earthquake waves, is a metamorphic phenomenon and that the change from the velocity characteristic of the second layer from the top (5.07 kilometers per second) to the velocity characteristic of the layer below it (6.69 kilometers per second) represents the change from one type of metamorphism to another. Certainly in ophiolites the change from the dike complex to gabbros occurs at about the same level as a change in the type of metamorphism. Both changes could influence the earthquake-wave velocities of the oceanic crust.

So far we have been concerned with the structure, composition and metamorphic features of ophiolites and what they tell us about oceanic geology. There is a separate question raised by ophiolites. If they represent oceanic crust formed at oceanic ridges and rises, how are they emplaced near the destructive plate margins adjacent to continents and island arcs? It is to this process that the term obduction is applied. The actual processes of obduction are poorly understood, and further consideration of the term and its implications is needed.

Almost as soon as the processes of plate tectonics were proposed the term subduction came into the geological literature. The word is from the Latin *sub,* under, and *ducere,* to lead. Hence "subduct" literally means "to lead under." "Ob-" rather than "sub-" implies that the movement is in a direction or manner contrary to the usual one. Therefore obduction implies a movement contrary to subduction, a movement over rather than under. It also implies an upward movement of the oceanic crust, and just how this movement comes about is a puzzle.

The simplest explanation of the on-land presence of ophiolites is to regard them as fragments of oceanic crust attached to a continent or to an island arc. As a result of the subduction of oceanic crust seaward of the eventual ophiolite, both the continent or island arc and the oceanic crust attached to it are underlain by a subduction zone. For the oceanic crust to be exposed as an ophiolite it must be lifted above sea level. This can happen in any one of several ways.

If the subducting plate is oceanic, it will take water down into the mantle in water-bearing metamorphic minerals such as zeolites and amphiboles. On being subducted these minerals will be heated and will release water, which will convert some of the peridotite in the overlying mantle into serpentinite. The serpentinization of the peridotite increases its volume and makes it lighter, so that it tends to rise. The process contributes to the uplift of the overlying crust and mantle.

Alternatively, if the potential ophiolite is underthrust by continental crust, the presence of lower-density continental rocks at depth will upset the normal equilibrium and cause uplift. In 1963 David Masson-Smith of the Institute of Geological Sciences and I, investigating variations in gravity over the Troodos massif, proposed that this ophiolite was lifted above sea level by the underthrusting of continental crust. The process was intensified by the rise of a mass of serpentinite under what is now the center of the massif. Similar mechanisms involving the underthrusting of an eventual ophiolite by an oceanic plate are favored by my Open University colleagues working in Oman to explain that ophiolite and by Daniel E. Karig of Cornell University to explain the Zimbales ophiolite in the Philippines.

Support for the underthrusting mechanism comes from the study of the metamorphic rocks under ophiolites. Many ophiolites, notably those of Newfoundland and Oman, are underlain by a thin layer of metamorphosed rocks separating the ophiolite from the underthrusting material. These metamorphic rocks were formed in a zone where the temperature was highest (about 600 degrees C.) immediately adjacent to the overlying ophiolite. The temperature diminished rapidly with depth, so that only a few hundred meters below the top of the zone the rocks are unmetamorphosed. The metamorphic agent here is probably a combination of heat emanating from the overlying slab of oceanic crust and frictional heat generated by the downward movement of the subducting plate. In these instances the assemblages of metamorphic minerals and their disposition are best explained by the continuous subduction of an oceanic plate.

In many of the less deformed ophiolites the rock layers that are correlated with the oceanic layers identified by earthquake-wave velocities are markedly thinner than the oceanic layers. This has led to the proposal that ophiolites represent oceanic crust that is thinner than normal, formed at minor ridges in small, marginal seas. Such thin crust, it has been argued, could be more easily obducted. Indeed, Julian Pearce and others have demonstrated that ophiolitic basalts show geochemical features most compatible with their being derived from a water-bearing melt whose water content had come from an underlying subduction zone. Moreover, the age of an ophiolite and the time of its emplacement are commonly very close together, and so it has been suggested that the oceanic crust formed first gets emplaced as an ophiolite before the conveyor belt of the subduction process gets well under way. This too could explain the thinness of an ophiolite layer.

Other factors, however, are not easily compatible with this simple model. There is no doubt that most ophiolites are in contact with underlying rocks of continental origin. In some instances it can be proved that it is the fragments of oceanic crust that have moved over the continental rocks. A classic case is the emplacement of the Papua–New Guinea ophiolite, which, as it has been described by Hugh L. Davies of the Australian Bureau of Mineral Resources, Geology and Geophysics has been emplaced southward along a thrust zone inclined to the north. This zone, unlike all others around the Pacific, is inclined seaward; it is one of the few cases where the term obduction may be appropriate.

In other instances blocks of oceanic crust, often many kilometers across, are found in a mélange, embedded in a matrix of serpentinite or muddy sediment. In these instances it seems most likely the blocks of oceanic crust were detached as the oceanic plate buckled and fractured before it was subducted. The blocks detached by the process fell into an adjacent deep oceanic trench. Such trenches are the main surface indication of a subduction zone.

The term "obduction" is therefore a troublesome one. It is nonetheless useful as long as the complexities involved are kept in mind. What may be more to the point is that in the present cycle of plate tectonics (there have been other cycles in the past) ophiolites are found mostly at or near subduction zones. Hence their presence can be taken to indicate the proximity of fossil destructive margins. This association has been applied successfully in geological reconstructions of Mesozoic and Lower Palaeozoic terrains by Robert Coleman, John Dewey and Alan Gilbert Smith, among others.

"Ophiolite" too is a somewhat troublesome term for on-land fragments of oceanic crust. Certainly the term has changed its meaning. In geology before the emergence of plate tectonics an ophiolite was associated with the initial stages in the development of a geosyncline; it consisted largely of serpentinite and had been metamorphosed in the course of a cycle of mountain building. The Troodos massif shows none of these features, and in the 1950's, when it was first being studied in detail, it was not even considered an ophiolite. Then by the early 1970's it became apparent that there were enough structures like the Troodos massif for them to need a collective label. Many of the complexes that were already called ophiolites turned out on reexamination to have the same structure; hence it was perhaps inevitable that the label ophiolite was retained. The term was, however, given a new and far more precise meaning. Today most earth scientists accept that ophiolites are on-land fragments of oceanic crust formed at oceanic ridges or rises, and with this consensus the time seems ripe to accept the term ophiolite, however inappropriate it may seem.


DIALOGO
di
GALILEO GALILEI LINCEO
AL SER.mo FERD. II. GRAN. DVCA DI
TOSCANA
Stefan Della Bella F.

The Galileo Affair

In arguing that the earth circles the sun Galileo adopted a mode of reasoning that led not only to his prosecution by the church but also to the new scientific methodology of hypothesis testing

by Owen Gingerich

Galileo's difficulties with the Roman Catholic church, which ultimately led to his trial and humiliation, have often been described as a confrontation between empirical science and blind dogmatism. Notwithstanding his abjuration, Galileo clearly believed in the truth of the heliocentric Copernican system. Today, with the sun-centered arrangement of the planets firmly established, it is easy to see Galileo as right and the church as wrong. In Galileo's time, however, the issues were by no means obvious or clear-cut.

Galileo defended the Copernican system by a series of ingenious arguments, many of them based on his new telescopic observations. From a modern point of view Galileo's defense seems immediately compelling, but when he presented his ideas, there was as yet no observational proof of the new cosmology, and even he remarked that he could not admire enough those who had adopted the heliocentric system in spite of the evidence of their senses. By the standards of his time his reasoning was not only contrary to traditional church doctrine but also flawed in its logic. Indeed, I would contend that Galileo was breaking the accepted rules of science, but by doing so he created new rules that have been accepted ever since.

The outcome of the Galileo affair, in which the church won the battle but lost the war, had important historical consequences, most notably a shift of scientific enterprise northward into Protestant countries. Three hundred and fifty years later, at a time when some individuals are again asserting a religious claim on cosmology, Galileo's experience still has much to say about the practice and the philosophy of science. What was at issue was both the truth of nature and the nature of truth.

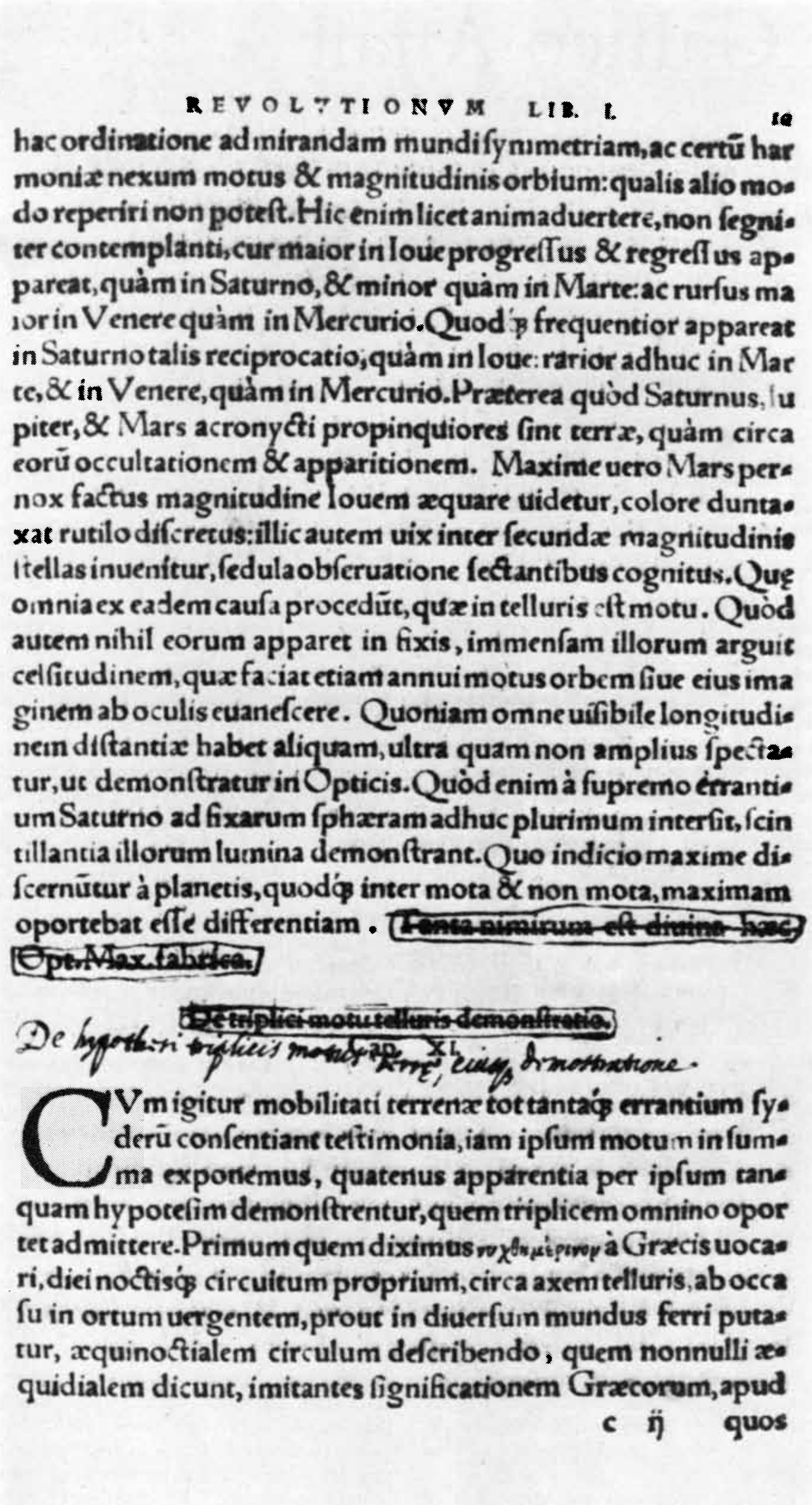
To understand the Galileo affair it is necessary to know something about the introduction of the Copernican cosmology some decades earlier. In 1543, when Copernicus' magnum opus, *De revolutionibus orbium coelestium* (*On the Revolutions of the Heavenly Spheres*) was finally published, there was not a single item of unambiguous observational evidence in its favor. Copernicus' achievement had been in the mind's eye. What he had noted was that by rearranging the planetary orbs so that the sun was near their center a wonderful regularity emerged. The fastest planet, Mercury, had the orbit closest to the sun; the slowest planet, Saturn, was at the outside, and the planets in between also were placed in the order of their periods. Furthermore, the scheme gave a natural explanation to several previously unrelated observational facts, such as the geometry of each planet's retrograde arc (the segment of the orbit in which the planet seems to reverse direction across the sky). This explanatory power came at a high cost, however: it threw the earth into a dizzying flight around the sun, and the earth somehow had to bring the moon along with it. In the framework of the accepted Aristotelian physics the entire scheme was ridiculous. "'Tis all in pieces, all coherence gone," John Donne was to lament a few generations later. And coherency is cherished above all else in science; it is the touchstone by which crank theories can be rejected.

In order to convey the viewpoint of the astronomical community about 50 years after the publication of *De revolutionibus,* I should like to describe an imaginary congress of the International Astronomical Union in 1592. The vice-president, Christoph Clavius of Rome, has risen to praise the remarks of the president, Tycho Brahe of Denmark. Tycho has lately introduced still another cosmological system, in which the planets orbit the sun but the sun itself and the accompanying planets are in orbit around a motionless earth. Clavius remarks that Tycho's system beautifully preserves the relations found by Copernicus in the harmonious spacing of the planets and gives a fully natural explanation of retrograde motions, just as Copernicus' hypothesis does. As Tycho himself has said, the Copernican arrangement nowhere offends the principles of mathematics, but it gives to the earth—this lazy, sluggish body, unfit for motion—a movement as swift as that of the ethereal planets. The Tychonic system brilliantly saves physics by keeping the earth at rest, and it is consistent with the scriptures, such as Psalm 104: "O Lord my God... who laid the foundations of the earth, that it should not be removed for ever."

An informal poll I have taken among the delegates indicates a somewhat mixed reaction: about half accept Tycho's view, but the rest say the choice of systems does not matter since all such geometric schemes are only hypothetical anyway. Some of those who adopt the latter attitude cite the anonymous preface to Copernicus' book: "Beware if you expect truth from astronomy lest you leave this field a greater fool than when you entered." Fewer than 10 percent agree with the Sicilian astronomer Franciscus Maurolycus that Copernicus deserved whips and lashes. The great majority find the Copernican tables preferable for calculating planetary positions, but that does not require a commitment to the heliocentric cosmology because the tables are set up independent of any particular arrangement of the planets.

Although gracious amity prevails be-

COSMOLOGICAL DISPUTE is represented in the frontispiece of Galileo's *Dialogue concerning the Two Chief World Systems,* printed in Florence in 1632. The three figures are Aristotle (*left*), Ptolemy (*middle*), who carries a model of nested geocentric spheres, and Nicolaus Copernicus (*right*), who bears an emblem of his own heliocentric theory. In discussions with Pope Urban VIII, Galileo had agreed to write a neutral account of the Ptolemaic and the Copernican systems, but the *Dialogue* was far from impartial. Galileo's advocacy of the heliocentric cosmology led to his prosecution by the Congregation of the Inquisition. The banner carries the dedication of the *Dialogue* to Ferdinand II de' Medici, the Grand Duke of Tuscany. "Linceo" identifies Galileo as a member of the Academy of Lynxes, a scientific society formed in 1603.


REVOLVTIONVM LIB. I. 10
hac ordinatione admirandam mundi ſymmetriam, ac certũ har
moniæ nexum motus & magnitudinis orbium: qualis alio mo
do reperiri non poteſt. Hic enim licet animaduertere, non ſegni
ter contemplanti, cur maior in Ioue progreſſus & regreſſus ap
pareat, quàm in Saturno, & minor quàm in Marte: ac rurſus ma
ior in Venere quàm in Mercurio. Quodq́ frequentior appareat
in Saturno talis reciprocatio, quàm in Ioue: rarior adhuc in Mar
te, & in Venere, quàm in Mercurio. Præterea quòd Saturnus, Iu
piter, & Mars acronycti propinquiores ſint terræ, quàm circa
eorũ occultationem & apparitionem. Maxime uero Mars per
nox factus magnitudine Iouem æquare uidetur, colore dunta
xat rutilo diſcretus: illic autem uix inter ſecundæ magnitudinis
ſtellas inuenitur, ſedula obſeruatione ſectantibus cognitus. Quę
omnia ex eadem cauſa procedũt, quæ in telluris eſt motu. Quòd
autem nihil eorum apparet in fixis, immenſam illorum arguit
celſitudinem, quæ faciat etiam annui motus orbem ſiue eius ima
ginem ab oculis euaneſcere. Quoniam omne uiſibile longitudi
nem diſtantiæ habet aliquam, ultra quam non amplius ſpecta
tur, ut demonſtratur in Opticis. Quòd enim à ſupremo erranti
um Saturno ad fixarum ſphæram adhuc plurimum interſit, ſcin
tillantia illorum lumina demonſtrant. Quo indicio maxime di
ſcernũtur à planetis, quodq́ inter mota & non mota, maximam
oportebat eſſe differentiam. ~~Tanta nimirum eſt diuina hæc
Opt. Max. fabrica.~~

~~De triplici motu telluris demonſtratio.~~
De hypothesi triplicis motus terræ, eiusq; demonstratione.

CVm igitur mobilitati terrenæ tot tantaq́ errantium ſy
derũ conſentiant teſtimonia, iam ipſum motum in ſum
ma exponemus, quatenus apparentia per ipſum tan
quam hypoteſim demonſtrentur, quem triplicem omnino opor
tet admittere. Primum quem diximus νυχθημερινὸν à Græcis uoca
ri, diei noctiſq́ circuitum proprium, circa axem telluris, ab occa
ſu in ortum uergentem, prout in diuerſum mundus ferri puta
tur, æquinoctialem circulum deſcribendo, quem nonnulli æ
quidialem dicunt, imitantes ſignificationem Græcorum, apud
c ij quos

CENSORED PASSAGES of Copernicus' *De revolutionibus orbium coelestium* were altered to make the heliocentric cosmology acceptable to the church by making it strictly hypothetical. The emendations shown are in Galileo's copy and were entered in his own hand. Following the instructions of the Holy Congregation of the Index, Galileo struck out the last sentence of chapter 10, which had read: "So vast, without any question, is the divine handiwork of the most excellent Almighty." He also changed the title of the next chapter from "On the Explication of the Threefold Motion of the Earth" to "On the Hypothesis of the Threefold Motion of the Earth and its Explication." The decision to censor *De revolutionibus* rather than ban it was made in 1616, partly at the urging of Maffeo Cardinal Barberini, who later became Pope Urban VIII. At the same time Galileo was warned against speaking too forcefully in favor of the Copernican cosmological system, although he was not officially enjoined against teaching it.

tween the Jesuit Clavius and the Lutheran Tycho, international tensions are evident. Michael Maestlin of Tübingen is vociferous in his criticism of Clavius' new calendar. Maestlin is also nettled that his graduate student Johannes Kepler, who has come along on a young astronomer's grant, does not agree that the Gregorian calendar is the work of the devil. As for the 27-year-old Galileo Galilei, an untenured mathematics professor at Pisa, no one at the congress has heard of him.

Nearly four centuries later millions of people who could not identify Tycho, Clavius, Maestlin or Kepler know the name of Galileo. One reason for Galileo's prominence is the importance of his contributions to both physics and astronomy, but his trial at the hands of the Inquisition has surely added to his fame. In the quaint words of the 19th-century physicist David Brewster, Galileo became a "martyr of science." Now, almost 350 years after his trial and abjuration, the Vatican itself has moved to reopen his case.

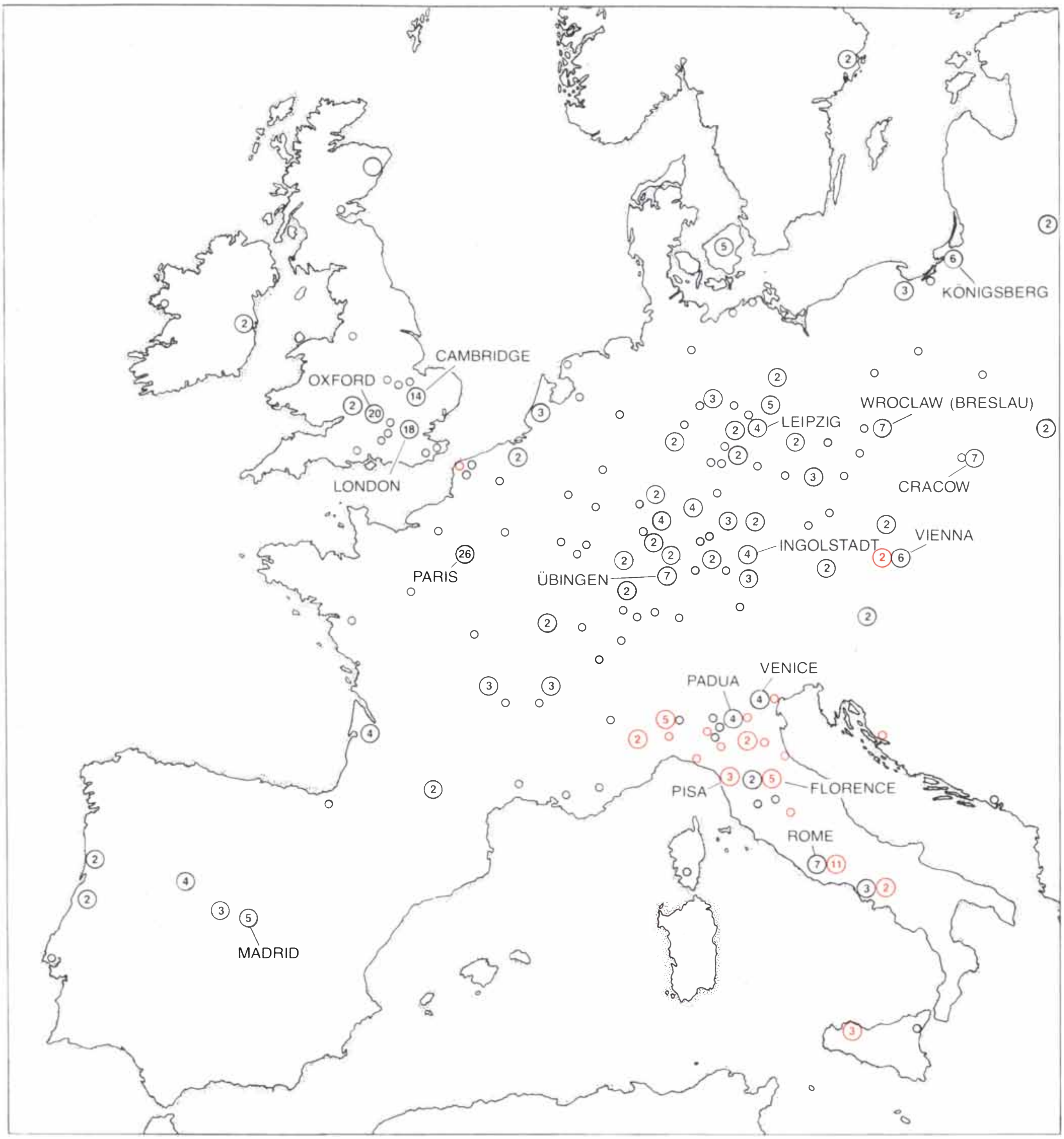
Galileo's ordeal is often referred to as a trial for heresy. Strictly speaking the Copernican system was never officially declared heretical, nor was Galileo condemned for heresy. The judge, who was out to get Galileo, raised a charge of "a vehement suspicion of heresy." He also found that an unofficial panel of theologians had agreed the Copernican system ought to be considered heresy, but this opinion never became the official position of the church. To understand these points it is necessary to examine both the historical circumstances of the trial and what was at stake philosophically.

At the end of the 16th century there was still no compelling reason to accept the Copernican doctrine as a physical picture of the universe. Astronomers were all well aware of its general idea, yet few believed it described the real world. There was widespread agreement that truth resided not in astronomy but in the Bible. Since the Book of Scripture had been literally dictated by God, it had a unique status. Even Galileo accepted this doctrine without hesitation. He did not necessarily agree, however, that the intellectual road to truth lay solely within the territory of the theologians. The Book of Scripture could be ambiguous, he argued, whereas God's Book of Nature could be probed and tested. He conceded that the Bible had its place, but he also believed the Bible told how to go to heaven, not how the heavens go.

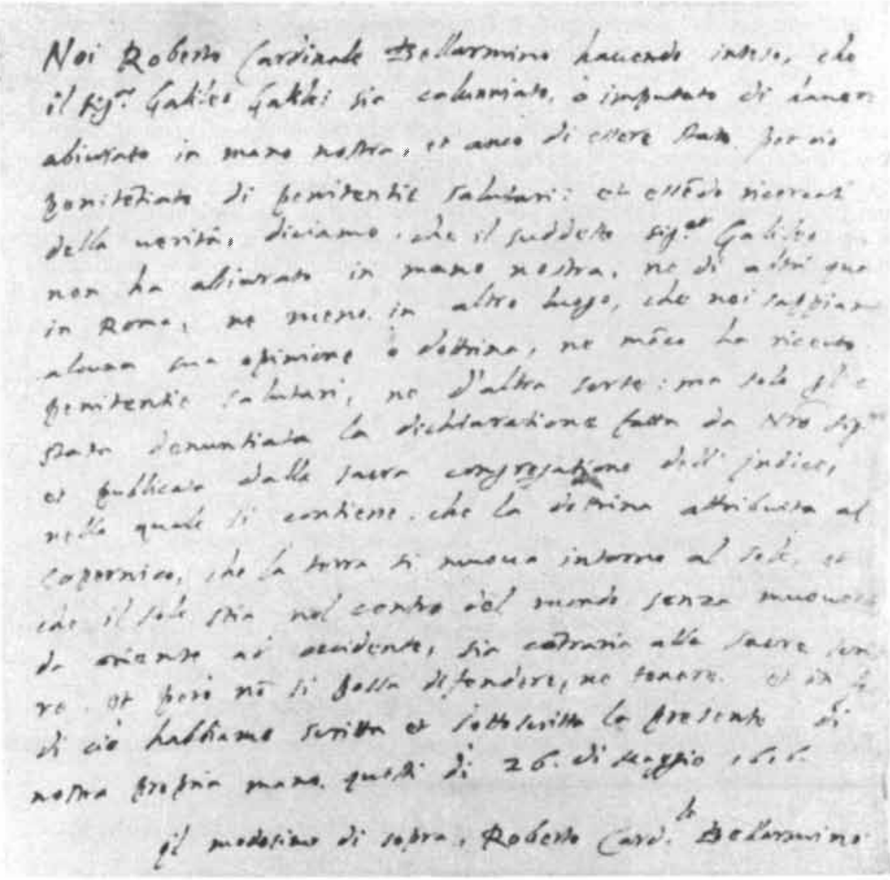
How *do* the heavens go, and how is their motion revealed by the Book of Nature? A flippant answer might be: by observing with the telescope. For Galileo the telescope had an enormous psychological impact. For years he had been at best a timid or even an indifferent Copernican, and he had taught his students in Pisa and later in Padua the standard arguments for a fixed, central earth. Then, in the fall of 1609, with an optical tube of his own making, a perspicillum as he called it, he turned his attention to the heavens and was staggered by what he saw. Within a few months his book reporting on his observations was off the presses: *Sidereus nuncius*, or *The Starry Messenger*. It told of mountains on the moon and of stars and satellites unknown to the ancients. The moon was earthlike and not the ethereal globe of pure crystal imagined by his predecessors. The Milky Way was revealed to be the confluence of innumerable stars. Most unexpected of all, Jupiter was circled by four companions. Galileo craftily called them the Medicean stars, with hopes of a government-supported position in Tuscany at the court of Grand Duke Cosimo II de' Medici.

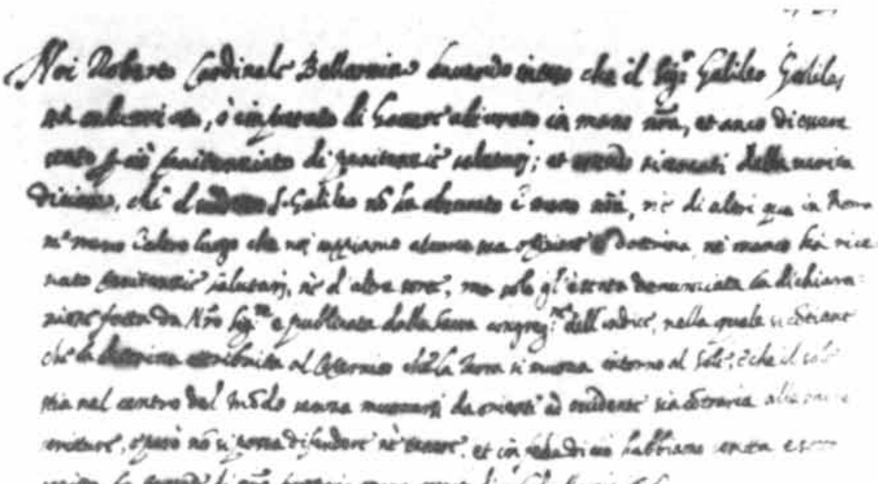
Galileo's observations with the telescope must have shaken his complacency, but his account in *The Starry Messenger* gives no unambiguous evidence that he espoused the Copernican system. The book had only just been printed, however, when he made another remarkable finding: the phases of Venus, which in a stroke falsified the Ptolemaic system.

Venus had been too close to the sun to observe when Galileo was making his



PROBABLE DISTRIBUTION of censored copies of *De revolutionibus* in 1620 suggests that the 1616 decree of the Holy Congregation of the Index was effective mainly in Italy; even in other Catholic countries, such as Spain, the decree was evidently not enforced. Censored copies are represented by colored circles and uncensored copies by black circles. Where there were multiple copies in one place the number is given. The map was compiled by examining the history of some 500 surviving copies of the first two editions. There were probably another 500 copies with a distribution similar to that of the approximately 380 books whose locations are given here.





We, Roberto Cardinal Bellarmino, having heard that it is calumniously reported that Signor Galileo Galilei has in our hand abjured and has also been punished with salutary penance, and being requested to state the truth as to this, declare that the said Signor Galileo has not abjured, either in our hand, or the hand of any other person here in Rome, or anywhere else, so far as we know, any opinion or doctrine held by him; neither has any salutary penance been imposed on him; but that only the declaration made by the Holy Father and published by the Sacred Congregation of the Index has been notified to him, wherein it is set forth that the doctrine attributed to Copernicus, that the Earth moves around the Sun and that the Sun is stationary in the center of the world and does not move from east to west, is contrary to the Holy Scriptures and therefore cannot be defended or held. In witness whereof we have written and subscribed these presents with our hand this twenty-sixth day of May, 1616.

astonishing discoveries at the end of 1609 and the beginning of 1610. Sometime late in the summer a former student, Benedetto Castelli, remarked to Galileo that in the Copernican system Venus should show the entire range of phases, from a dark disk through crescent and gibbous forms to a fully illuminated disk. In the Ptolemaic system, on the other hand, the epicycle of Venus is locked between the earth and the sun, and Venus therefore has only crescent phases; it never passes behind the sun for full illumination.

Not until October did Galileo train his perspicillum on Venus, which was then in its distant gibbous phase. By early December, when the planet had waned to a miniature half moon, he put forth his discovery in an anagram: "Haec immatura a me iam frustra leguntur o.y." ("These are at present too young to be read by me." The letters "o.y." are part of the original sentence but did not fit into the anagram.) He undoubtedly chose this veiled form of announcement to give himself time to be sure of his finding; after all, Venus might lie always beyond the sun, in which case it would go back into a gibbous phase. By this stratagem Galileo also guarded his priority; since Castelli had mentioned the possibility in the first place, others might have been on the verge of making the same discovery.

Galileo was a scrambling social climber. His discoveries had gained him a new post as mathematician to the Medici and had brought him the fame he greatly relished. Fame in turn brought power of a kind, perhaps the power to persuade the entire Catholic hierarchy to adopt the Copernican system. At least Galileo was egotistical enough to expect that it would.

In Galileo's rush to assert a claim of priority he was sometimes more aggressive than might seem prudent. He got into a squabble with the Jesuit Christoph Scheiner when each asserted he had been the first to observe sunspots. Moreover, Scheiner preferred to believe the sun was unblemished and the spots were intervening clouds. Galileo proved otherwise, with rather little charity to Scheiner. Giorgio de Santillana, in his book *The Crime of Galileo,* hints darkly that Scheiner never forgot and years later led the Jesuits in a vendetta. It is true Scheiner was in Rome at the time of

LETTER TO GALILEO from Roberto Cardinal Bellarmino sets forth the nature of the warning issued to Galileo in 1616. The original is at the top and Galileo's copy is in the middle. In 1633, when the Inquisition hinted that Galileo had been forbidden to write on the Copernican cosmology, he responded by producing first his copy and later the original letter. Both are now among the Galileo papers in the Secret Archives of the Vatican.

Galileo's trial, but there is no evidence he had anything to do with those machinations. Nevertheless, the story might make a play along the lines of *Amadeus,* the current Broadway production about the supposed poisoning of Mozart.

By New Year's Day of 1611, just after Venus had rounded its western elongation, the crescent phase began to emerge and Galileo unscrambled his anagram for Kepler. It read, "Cynthiae figuras aemulatur mater amorum." ("The mother of love imitates the shapes of Cynthia"), or in other words, Venus goes through the same series of phases as the moon. When Galileo realized that the observed phases of Venus are incompatible with the Ptolemaic arrangement, he could hardly fail to notice that the Book of Nature was indeed saying something about how the heavens go. With the Ptolemaic scheme eliminated Galileo threw his support behind the Copernican system, ignoring the Tychonic plan.

At a breakfast with Cosimo de' Medici and his mother, the Dowager Grand Duchess Cristina, the question of the reality of the Jovian satellites came under discussion. Galileo himself was not present but Castelli was there. Through Galileo's influence Castelli had just become professor of mathematics at Pisa, and he entered into a spirited discussion with Cristina on the issue of whether there is any conflict between the Bible and the heliocentric theory. As a direct result of that debate Galileo was challenged to defend his view that the Book of Scripture raises no insuperable objections to the Copernican system. Galileo wrote a cogent analysis, including the splendid epigram about the Bible's teaching how to go to heaven, not how the heavens go. (Actually Galileo had borrowed the saying from Caesar Cardinal Baronius, the librarian of the Vatican.)

It was one thing to argue that the heliocentric arrangement is compatible with the Book of Scripture and quite another to prove that the Book of Nature speaks unmistakably in favor of Copernicus. To understand this part of the controversy it is necessary to keep in mind the two forms of Aristotelian logic: induction and deduction.

Induction is the process of drawing general conclusions from particular instances; it is, I think, the basic process whereby learning takes place. Consider the reproduction of birds: chickens lay eggs, robins lay eggs, ostriches lay eggs and so on, and thus we generalize that all birds reproduce by laying eggs. We have not proved this conclusion, however, since there is always the possibility that a counterexample will be found. For this reason inductive reasoning, as all the scholastic philosophers of Galileo's time knew, cannot lead to indubitable truth.

Deduction is another matter. Given true premises, a conclusion reached by valid deduction must be rigorously true. Consider this syllogism:

A. If it is raining, the streets are wet.
B. It is raining.
C. Therefore the streets are wet.

Now consider the converse:

A. If it is raining, the streets are wet.
B. The streets are wet.
C. Therefore it is raining.

To students of logic this procedure of confirming the consequent was a well-known fallacy. After all, the streets could be wet for other reasons: the winter snow could be melting, the street-cleaning department might be out in force or the Lippizaner horses might have been on parade.

How does this logical analysis apply to Galileo's defense of Copernicanism? Consider this syllogism:

A. If the planetary system is heliocentric, Venus will show phases.
B. The system is heliocentric.
C. Therefore Venus will show phases.

True enough, but this was not the form of Galileo's argument. He had exchanged the second premise and the conclusion:

A. If the planetary system is heliocentric, Venus will show phases.
B. Venus shows phases.
C. Therefore the planetary system is heliocentric.

Clearly Galileo had committed an elementary blunder of logic, and even Kepler criticized him for it. There might well be other explanations for the observed phases of Venus; indeed, the Tychonic system also predicted them.

When Galileo's "Letter to Cristina" was circulated in Rome in 1616, it elicited the following response from Roberto Cardinal Bellarmino, the leading Catholic theologian of the day, who wrote to another Copernican, Father Paolo Antonio Foscarini:

"I have gladly read the letter in Italian and the essay in Latin that Your Reverence has sent me, and I thank you for both, confessing that they are filled with ingenuity and learning. But since you ask for my opinion, I shall give it to you briefly, as you have little time for reading and I for writing.

"First. I say that it appears to me that Your Reverence and Signor Galileo did prudently to content yourselves with speaking hypothetically and not positively, as I have always believed Copernicus did. For to say that assuming the earth moves and the sun stands still saves all the appearances better than eccentrics and epicycles is to speak well. This has no danger in it, and it suffices for mathematicians.

"But to wish to affirm that the sun is really fixed in the center of the heavens and that the earth is situated in the third sphere and revolves very swiftly around the sun is a very dangerous thing, not only by irritating all the theologians and scholastic philosophers, but also by injuring our holy faith and making the sacred Scripture false. For Your Reverence has indeed demonstrated many ways of expounding the Bible, but you have not applied them specifically, and doubtless you would have had a great deal of difficulty if you had tried to explain all the passages that you yourself have cited....

"Further, I say that if there were a true demonstration that the sun is in the center of the universe and that the sun does not go around the earth but the earth goes around the sun, then it would be necessary to be careful in explaining the Scriptures that seemed contrary, and we should rather have to say that we do not understand them than to say that something is false. But I do not think there is any such demonstration, since none has been shown to me. To demonstrate that the appearances are saved by assuming the sun at the center and the earth in the heavens is not the same thing as to demonstrate that in fact the sun is in the center and the earth in the heavens. I believe that the first demonstration may exist, but I have very grave doubts about the second." (The translation is abridged from one done by Stillman Drake.)

Galileo knew he could not logically establish the Copernican system by deduction, but the situation was not quite that simple. The Copernican system not only predicted the phases of Venus but also, as a model, explained many other things. If the earth was a planet, the other planets might well be earthlike, and so indeed the moon turned out to be when he examined it with his telescope. The Copernican system arranged the planets naturally by period; similarly, when the telescope revealed the satellites of Jupiter, they were found to be arranged sequentially by period, as in a miniature solar system.

Galileo's process of reasoning was similar to induction but more sophisticated. It was, in an embryonic state, what is now called the hypothetico-deductive method: the testing of a hypothetical model, which attains ever more convincing likelihood as it passes each test successfully. Today it is not the word "truth" but the word "model" that continually decorates the pages of scientific journals.

As far as the theologians were concerned, the Copernican system was not really the issue. I can hardly emphasize this point enough. The battleground was the method itself, the route to sure knowledge of the world, the question of whether the Book of Nature could in any way rival the inerrant Book of Scripture as an avenue to truth. In the opinion of Cardinal Bellarmino and the other Catholic theologians Galileo's procedures were essentially inductive and therefore potentially fallacious. Such contingent arguments were insufficient to force a reinterpretation of scripture that might erode the concept of the inerrancy of Holy Writ.

To be quite sure of avoiding confusion in the popular mind (particularly because issues of interpretation were central in the ongoing battle with the Protestants) the church officials found it prudent to condemn the Copernican teaching. The first step was to seek a theological opinion on two separate propositions: the immobility of the sun and the mobility of the earth. The report, which was essentially an internal memorandum, said the immobility of the sun was foolish and formally heretical because it violated the literal meaning of the Scriptures, but the mobility of the earth was merely erroneous. The question then was what to do about the report. Two actions were planned: to rein in Galileo and to put *De revolutionibus* on the Index of prohibited books.

The latter measure, however, entailed certain practical difficulties. Copernicus' book was considered an important contribution to the reform of astronomy, on which the calendar and the accurate determination of the date of Easter depended. Accordingly the Holy Congregation of the Index decided not to proscribe the book but instead to expurgate and emend it.

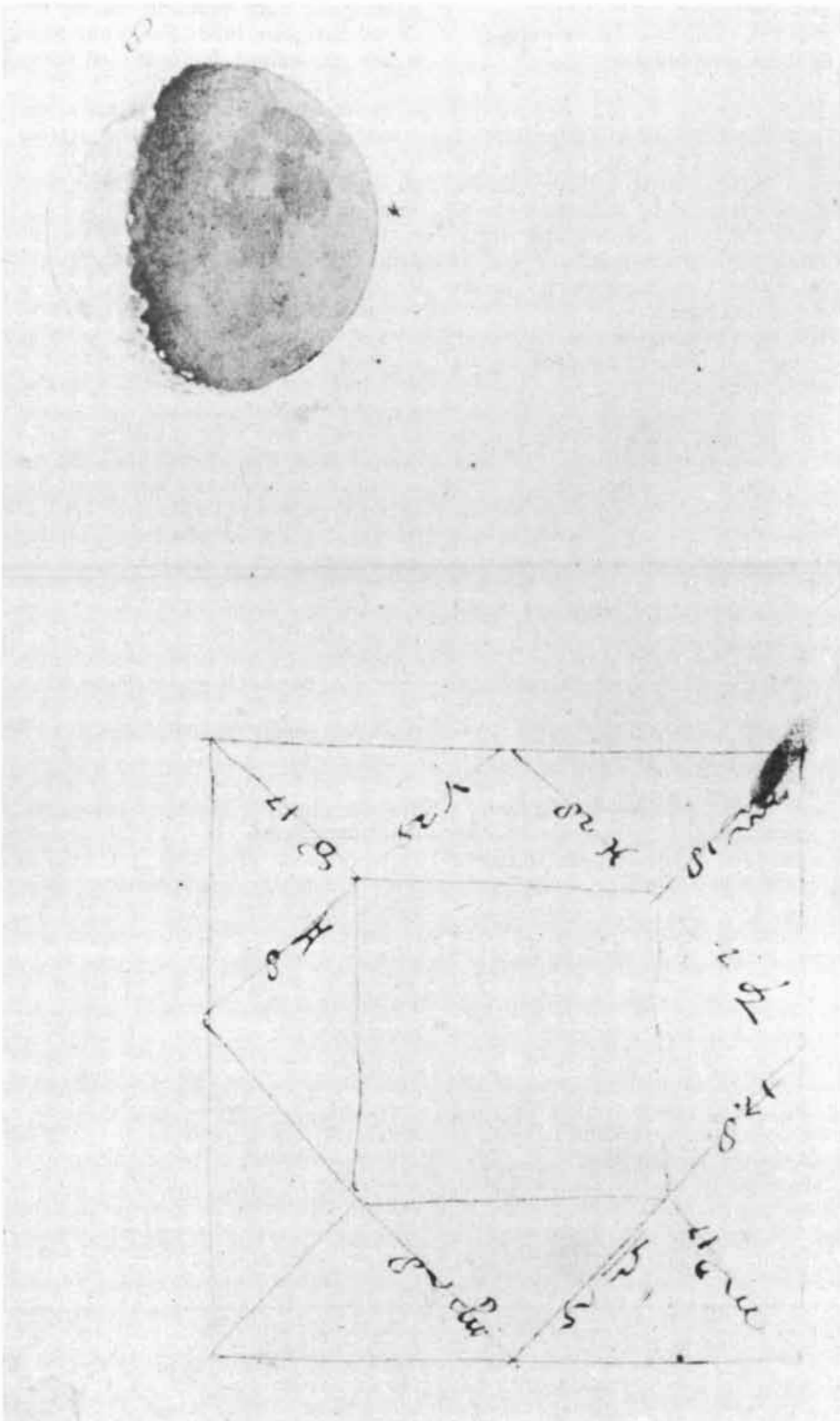
From these deliberations a bit of gossip has survived in the diary of Giovanfrancesco Buonamici, a diplomatic secretary from Galileo's province of Tuscany. Buonamici wrote that "Pope Paul V was of the opinion to declare Copernicus contrary to the faith; but Cardinals Bonifacio Caetani and Maffeo Barberini withstood the Pope openly and checked him with the good reasons they gave." The two cardinals were central figures in the cosmological controversy. Barberini was later to have an even larger role in the story of Galileo's life, and Caetani drafted the opinion recommending censorship of *De revolutionibus.*

Caetani's opinion declared that the Copernican teaching was false and opposed to Scripture, but not that it was heretical. This may seem to be a distinction without a difference, but it was certainly not so in the 17th century. The instructions for the censorship read:

"If certain of Copernicus' passages on the motion of the earth are not hypothetical, make them hypothetical; then they will not be against either the truth or the Holy Writ. On the contrary, in a certain sense they will be in agreement with them, on account of the false nature of suppositions, which the study of astronomy is accustomed to use as its special right."

Even as the Holy Congregation of the Index was moving against Copernicus' book, Galileo was in Rome aggressively lobbying on behalf of the heliocentric system. It seems he was convinced he could single-handedly sway the Catholic leaders to his view. Indeed, he had powerful friends in Rome who were sympathetic to his ideas, even among the churchmen, but the conservative forces were also strong, and they included Pope Paul V.

While Galileo was in Rome the other part of the pope's response to the theological opinion was put into action. Ga-



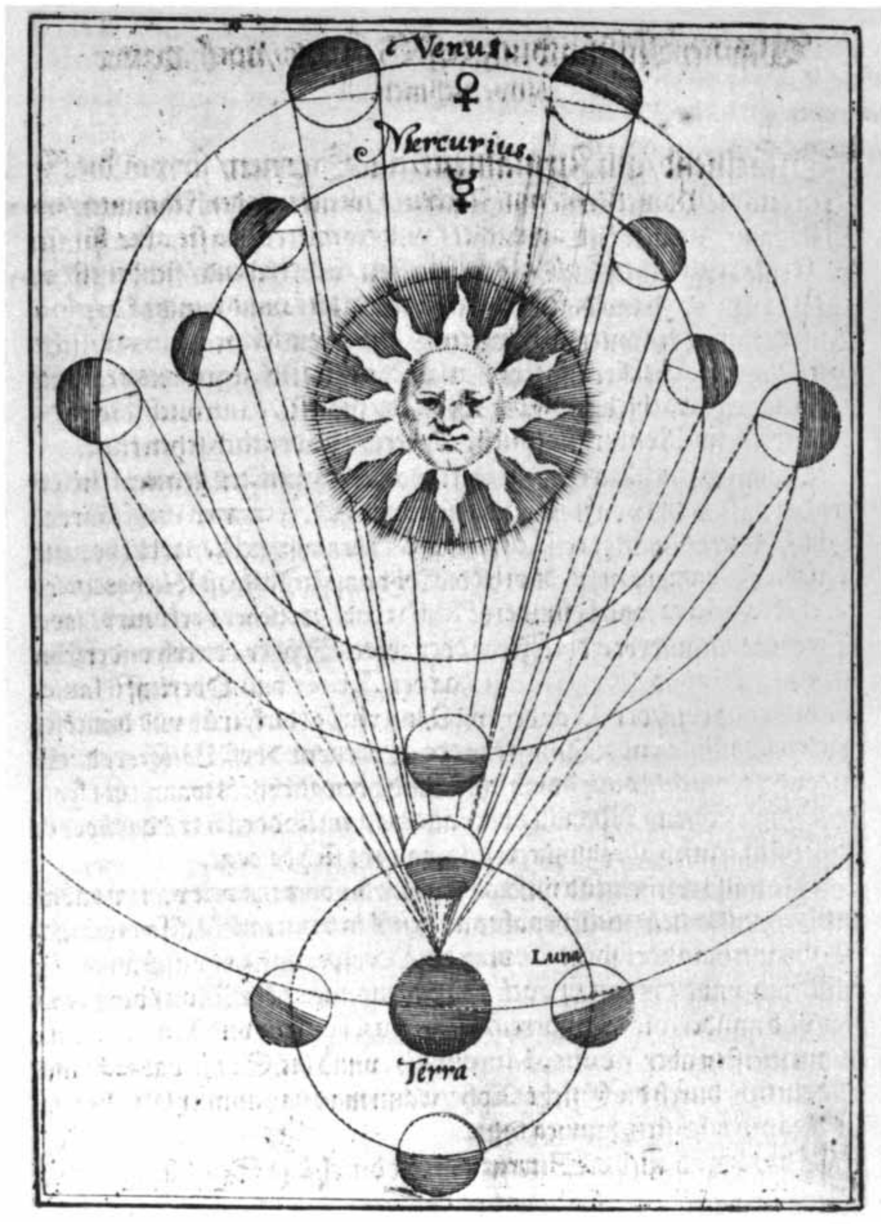
EARTHLIKE FEATURES on the surface of the moon were among the observations cited by Galileo in support of the heliocentric theory. The presence of mountains, craters and other "blemishes" indicated that the heavenly bodies are not fundamentally different from the earth. It therefore became reasonable to suppose the earth is a planet and not a fixed sphere with a quite different status. The drawing was made by Galileo after he constructed an astronomical telescope in 1609. On the same page is the start of a horoscope he cast for Cosimo II de' Medici.


WHY OUR VERSION OF A BMW OUTPERFORMS EVERYONE ELSE'S.

lileo was to be called before Cardinal Bellarmino and cautioned against speaking out too forcefully on behalf of the Copernican system. The pope told Bellarmino that if Galileo proved intractable, he was to be ordered to keep quiet. To be sure the pope's wishes were enforced the interview was conducted in the presence of two Dominican friars, members of the order charged with administering the Inquisition.

As it turned out Galileo was cooperative in accepting Bellarmino's warning. After the conference, however, rumors began to circulate in Rome that Galileo had been officially enjoined against teaching the Copernican doctrine. Galileo was naturally disturbed by the rumors, and he sought and received a letter from Bellarmino saying that no such thing had happened. It read in part:

"We, Roberto Cardinal Bellarmino, having heard that it is calumniously reported that Signor Galileo Galilei has in our hand abjured and has also been punished... declare that the said Signor Galileo has not abjured... any opinion or doctrine held by him; neither has any salutary penance been imposed on him; but that only the declaration made by the Holy Father and published by the Sacred Congregation of the Index has been notified to him, wherein it is set forth that the doctrine attributed to Copernicus... is contrary to the Holy Scriptures and therefore cannot be defended or held." (The translation is by de Santillana.)



PHASES OF VENUS had an important role in Galileo's own conversion to the Copernican view. A student, Benedetto Castelli, pointed out that in the Ptolemaic system Venus would show only crescent phases because it would always remain between the earth and the sun; in the Copernican system, on the other hand, Venus would show a full range of phases. In 1610 Galileo trained his telescope on the planet and was able to observe its progress from a gibbous to a crescent form. He considered the evidence for a heliocentric planetary system compelling, but others contended that such empirical findings could not supply a rigorous proof because other arrangements of the planetary system leading to the observations could be imagined. The illustration, which shows the phases of Mercury and the moon as well as those of Venus, is from a treatise by the Swiss mathematician Matthias Hirzgarter published in 1643.

Thus for the time being Galileo was silenced. For seven years he remained in Florence and complied with Cardinal Bellarmino's advice. He was as feisty as ever, but he reserved his scrappiness for other subjects, such as the comets of 1618. In his book on the comets (*Il Saggiatore,* or *The Assayer*) he avoided discussing the Copernican system, but he included so many interesting remarks on the nature of science that the book is sometimes called his scientific manifesto. He stated, in Italian:

"Philosophy is written in this grand book, the universe, which stands continually open to our gaze. But the book cannot be understood unless one first learns to comprehend the language and read the letters in which it is composed. It is written in the language of mathematics, and its characters are triangles, circles and other geometric figures.... Without these one wanders about in a dark labyrinth."

The printing of *Il Saggiatore* was not finished when news arrived that cheered all liberal Catholics. The newly elected pope, who had taken the name Urban VIII, was Maffeo Barberini, one of the cardinals who had intervened to prevent the proscription of *De revolutionibus.* Barberini was also a friend of the arts and a fellow member with Galileo of the small Academy of Lynxes, one of the earliest scientific societies. The delighted Lynxes had just enough time to change the title page on Galileo's book so that it could be dedicated to the new pontiff. Before a year had passed Galileo was in Rome for a series of papal audiences. Urban assured him that *Il Saggiatore* had been read to him, to his great pleasure. Galileo hinted he would like to write more, in particular a book on the relative merits of the Copernican and the Ptolemaic systems, but his enemies prevented him.

From what is known of the two men it is possible to speculate on how the conversation went. "Nonsense," the pope may have responded. "I helped to keep this from becoming heresy before, and I can protect you now. But remember, your account should be neutral, since you have no physical proof of the Copernican system."

"Ah," replied Galileo, "but I do. I believe the tides are the proof of a moving earth, and I propose to call my book *On the Flux and Reflux of the Sea.*"

"No," said Urban, "that won't do at all. That title would give too much

Finally you can afford to satisfy your lust for power.

It's been said that the future belongs to those who understand computers.

Well, this is your chance.

What you see here is the Sinclair ZX81 personal computer for only $99.95.

It's a personal computer that you can order right now. That you can learn to use. And that will open the door to understanding the technology that is rapidly changing the very way you live and work.

The computer for everyone.

The ZX81 is a true personal computer—not a business computer or an elaborate games computer. Everyone can afford it and learn to use it.

It hooks up to your black and white or color TV for video display. And it comes with a comprehensive instruction book that takes you step-by-step into the world of computers.

You and your children will learn how computers work. How to write your own programs. Even how to create computer games.

And Sinclair offers dozens of ready-to-use software packages. Try your skills at a game of chess. Learn famous composers and musicians. Computerize your personal budget. These programs and more are described in our software catalog which you'll receive with your ZX81. The programs come on cassette tapes, and work with a standard cassette recorder or player.

The $99.95 personal computer.

For less than $100, the Sinclair ZX81 will get you started in personal computing right now. Your children will gain an understanding of computers that will benefit them the rest of their lives. And you will be prepared to make informed decisions about using and buying computers, both in your career and in your home.

Free guide to programming

There's never been a better time to buy a personal computer.

And for only $99.95, there's never been a better price.

How to order today.

Call our toll-free number and use your MasterCard or VISA. Or send the coupon along with a check or money order.

Then try out the ZX81 for 10 days. If you're not satisfied, just return it to us and we'll refund your money.

Call toll free: 800-543-3000.

Ask for operator #509. In Ohio call: 800-582-1364; in Canada call: 513-729-4300. Ask for operator #509. Have your MasterCard or VISA ready.

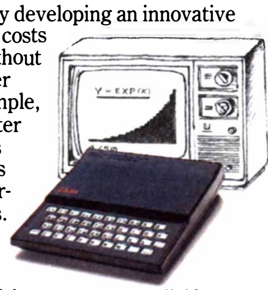
Phones open 24 hours a day, 7 days a week. These numbers are for orders only.

If you just want information, please write: Sinclair Research Ltd., 2 Sinclair Plaza, Nashua, NH 03061.

How did we do it?

Other personal computers cost hundreds, even thousands of dollars more. How did Sinclair create a personal computer for less than $100?

We did it by developing an innovative design that cuts costs dramatically without cutting computer power. For example, our unique Master Chip replaces as many as 18 chips used in other personal computers.



Sinclair programs are available

The success of the ZX81 speaks for itself. It is now the fastest-selling personal computer in the world. And we stand behind our product. If anything goes wrong in the first 90 days, we'll repair or replace it free of charge. Even after that, our national service-by-mail facilities are as close as your post office.

Why now?

Why should you take the time right now to order your Sinclair?

Because computers have become an important part of everyone's life. In business, you'll find them on the desks of both secretaries and presidents. In schools, you'll find them in a growing number of classrooms. Even in grade schools.

To order call toll free 800-543-3000 Operator #509

AD CODE: 08SA

Sinclair Research Ltd., One Sinclair Plaza, Nashua, NH 03061.

Yes! I'd like to try the Sinclair ZX81 personal computer. I understand that if I'm not completely satisfied, I can return it in 10 days for a full refund.

	PRICE†	QTY.	AMOUNT
ZX81 personal computer	$99.95		
Add shipping	$4.95	–	$4.95
†U.S. Dollars		TOTAL	

☐ Check or money order enclosed.

Name ____________________

Street ____________________

City ____________ State ________ Zip ________

sinclair

Sinclair technology is also available in Timex/Sinclair computers under a license from Sinclair Research Ltd.

prominence to what you take to be a physical proof, but God could have created the tides in any way he liked, and not necessarily by moving the earth." Note that Urban's argument was the same as Bellarmino's: even if a moving earth would produce tides, the observation of tides does not necessarily imply the movement of the earth. The case is particularly ironic, because Galileo's physical argument based on the tides was quite wrong. (He attributed them to the daily change in velocity that results from the earth's compound motion of rotation and revolution.)

Galileo was elated to have the gag order removed by the highest possible authority, and he returned to Florence to work on his book. He adopted the popular form of a dialogue, just as his father, a distinguished musician, had done in writing a *Dialogue on Ancient and Modern Music*. Galileo's three speakers are Simplicio, a traditionalist, named after a sixth-century commentator on Aristotle; Salviati, who most often speaks for Galileo himself, and Sagredo, an open-minded man of the world who asks intelligent questions and is generally persuaded by Salviati's reasoning.

The arguments marshaled on behalf of the Copernican system include the phases of Venus, the harmony of the arrangement of the planets and the existence of the tides. The work could hardly be considered neutral, but it ends with the pope's argument in the following words spoken by Simplicio:

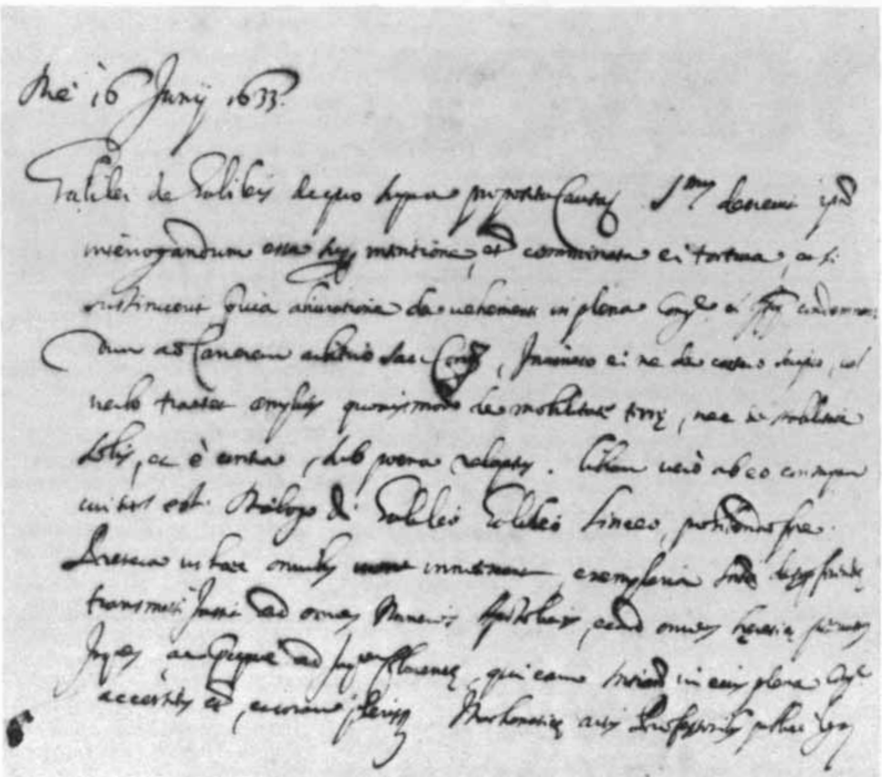
"I confess that your hypothesis on the flux and reflux of the sea is far more ingenious than any of those I have ever heard; still, I esteem it neither true nor conclusive, but, keeping always in mind a most solid doctrine I once received from a most eminent person, I know that if you were asked whether God in his infinite power and wisdom might confer upon the element of water the reciprocal motion in any other way, both of you would answer that he could, and in many ways, some beyond the reach of our intellect." (The translation is based on one by Drake.)

The passage seems quite innocuous, and yet it is singularly inappropriate as the closing argument of the preceding four days of dialogue. Throughout the work Galileo had attempted to show that reasoning from the Book of Nature can, at the very least, establish that one world view is far likelier than another. This has surely been the method of science ever since. Indeed, one might argue (as Alfred North Whitehead did) that since an omnipotent Creator could have made the world in any way he liked, it is all the more incumbent on scientists to discover which way God chose to make it.

When the *Dialogue* appeared, Galileo's enemies were outraged, and they quickly persuaded the pope that the book was heavily weighted in favor of the Copernican system. Furthermore, they convinced the pope that he had been made to look a fool by having his argument given to Simplicio, whose very name suggested "simpleton." The pope, agreeing that Galileo had gone too far, unleashed the Inquisition.

There were two stumbling blocks to prosecution: Copernicus' doctrine had never been publicly declared heretical, and the *Dialogue* had received a license from the censors. From the Vatican Archives, however, the Inquisitors produced a fascinating document: a report of the 1616 meeting between Galileo and Bellarmino. The report stated that an official injunction had indeed been served on Galileo, and that the astronomer had promised not to teach or defend the Copernican doctrine in any way. The pope was furious; it appeared Galileo not only had made him into a fool but also had deceived him about the outcome of the proceedings of 1616.

In February, 1633, Galileo was ordered to Rome, and he was told to come immediately in spite of the rigors of winter travel for a man of almost 70. Before a tribunal of 10 cardinals he was accused of disobedience. The archival evidence, however, was quite irregular: the document was neither signed nor notarized, as such an injunction should have been. Bellarmino had died, and so it was difficult to clarify the status of the document. Hence the Inquisitors, without revealing the source of their accusations, tried to get Galileo to admit that he had



16 June 1633

Galileo Galilei, for the above reasons, as decreed by his Holiness, is to be interrogated concerning the accusation, even threatened with torture, and if he sustains it, proceeding to an abjuration of the vehement [suspicion of heresy] before the full Congregation of the Holy Office, sentenced to imprisonment at the pleasure of the Holy Congregation, ordered, in either writing or speaking, not to treat further in any way either the mobility of the Earth or the stability of the Sun; or otherwise he will suffer the punishment of relapse. The book actually written by him, whose title is *Dialogo di Galileo Galilei Linceo*, is to be prohibited. Furthermore, that these things may be known by all, he ordered that copies of the foregoing sentence shall be sent to all Apostolic Nuncios, to all Inquisitors against heretical depravity, and especially the Inquisitor of Florence, who shall publicly read the sentence to his whole congregation and even in the presence of as many of those who teach mathematics as he can summon together.

BOOK OF DECREES of the Congregation of the Inquisition records the sentencing of Galileo in 1633. The proceedings against him had come to a halt after he had produced Cardinal Bellarmino's letter of 1616. Thereafter an agreement had been reached: Galileo would repent and would promise to write no more on cosmology. The agreement was overruled, however, and he was forced to submit to the humiliating ritual of abjuration, followed by house arrest, the banning of the *Dialogue* and a prohibition of further writing on the Copernican system.

been served an injunction, which would have established the legitimacy of the earlier document. Ultimately Galileo played his trump card. Having been alerted by his friends and their spies, he knew that the Inquisition was looking into his 1616 visit to Rome, and so he had brought a copy of Bellarmino's letter. Galileo's unexpected move threw the Inquisition into disarray, and the cardinals decided to adjourn.

It was a duel of wits, and Galileo had outwitted the pope. Nevertheless, all the secular power remained in the hands of the church, and the pope could not afford the embarrassment of bringing Galileo to Rome for naught. Even Galileo could appreciate this, and so some plea bargaining ensued. It could all be settled out of court: Galileo would confess that he had gone too far, would repent and then would be sent home and enjoined to avoid writing about cosmology.

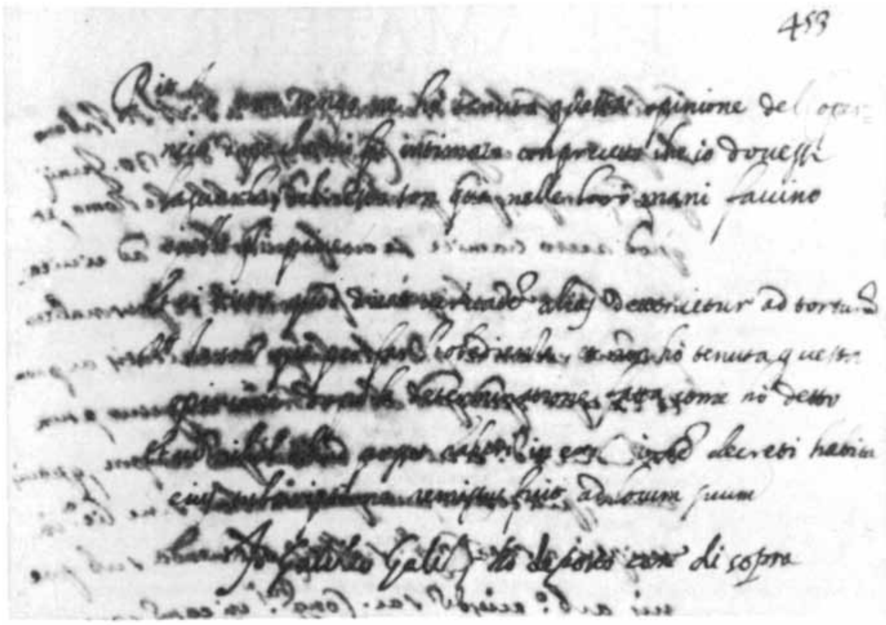
One can imagine Galileo's shock on June 16, 1633, when he found that the agreement had been overruled and the following sentence was entered in the Book of Decrees: "Galileo Galilei . . . is to be interrogated concerning the accusation, even threatened with torture, and if he sustains it, proceeding to an abjuration of the vehement [suspicion of heresy] before the full Congregation of the Holy Office, sentenced to imprisonment. . . ." He was also forbidden to write further on the mobility of the earth, and the *Dialogue* was banned.

On the next page the results of the interrogation are recorded. In Italian are Galileo's words: "I do not hold and have not held this opinion of Copernicus since the command was intimated to me that I must abandon it." Then he was again told to speak the truth under the threat of torture. He responded: "I am here to submit, and I have not held this opinion since the decision was pronounced, as I have stated." Finally, there is a notation that nothing further could be done, and this time the document is properly signed in Galileo's hand.

Galileo was sent back to his house at Arcetri, outside Florence, where he remained under house arrest until his death in 1642. Partly as a consequence of his persecution, the center of creative science moved northward to the Protestant countries, notably the Netherlands and England.

I am fascinated by the choices the Vatican confronts today in reopening Galileo's case. In the first place, it would do no good to announce that the Copernican doctrine should never have been declared heretical, since strictly speaking it never was. Second, Galileo was tried not so much for heresy as for disobeying orders, and it seems clear beyond question that he ignored the earlier decree of the Index when he published his *Dialogue.*

Where there is room for maneuver, it seems to me, is in accepting Galileo's arguments about the reconciliation of science and Scripture. The truth of the Bible, for those who wish to affirm it without rejecting the findings of science, must not be found in a literal six days of Creation, in the sun standing still for the battle of Gibeon or in a physically real star of Bethlehem. I quote Galileo, as he quoted Cardinal Baronius: "The Bible teaches how to go to heaven, not how the heavens go." Such a judgment, it seems to me, would confirm what has long since been accepted by both Catholic and Protestant theologians. It would also speak to the current controversy over Darwinian evolution and the so-called Creation science.



"I do not hold and have not held this opinion of Copernicus since the command was intimated to me that I must abandon it; for the rest, I am here in your hands—do with me what you please."

Being once more bidden to speak the truth, otherwise recourse would be had to torture:

"I am here to submit, and I have not held this opinion since the decision was pronounced, as I have stated."

And since nothing further could be done in execution of the decree, his signature was obtained, and he was sent back to his place.

"I, Galileo Galilei, testify as above."

GALILEO'S ABJURATION appears in the Book of Decrees following his sentencing. He retired to his house at Arcetri outside Florence, where he was confined until his death in 1642.

I had a sense of déjà vu when the creationists in California tried to have evolution presented in biology textbooks as a mere hypothesis. This was precisely the tactic the Inquisition adopted with Copernicus' book: they made it acceptable by making it appear hypothetical. I expect the creationists will have about as much success as the Holy Congregation of the Index did. Of course, Galileo believed the Copernican system could be defended as physically real, and not simply as a hypothetical geometric arrangement. It is an irony of history that Galileo's own methods of scientific argument were instrumental in showing that what passes for truth in science is only the likely or the probable; truth can never be final and never absolute. What makes science so fascinating is the task of pushing ever closer to the unattainable goal of complete knowledge.

It is this process that the poet Robinson Jeffers had in mind when he wrote: "The mathematicians and the physics men have their mythology; they work alongside the truth, never touching it; their equations are false but the things *work.*" The mathematicians and the physicists cannot really claim truth, but they have certainly sorted out a lot of things that do not work, and they are building a wondrously coherent picture of the universe. The Copernican system is surely a part of that coherency. A universe billions of years old and evolving is also part of that coherency. Galileo made a noble effort to convey such a picture of beauty and rational coherency to his public. Scientists today would honor him by helping their own public to understand better not only the majesty and the beauty of the modern scientific picture of the universe but also the process of hypothesizing and testing by which that view is achieved.

THE AMATEUR SCIENTIST

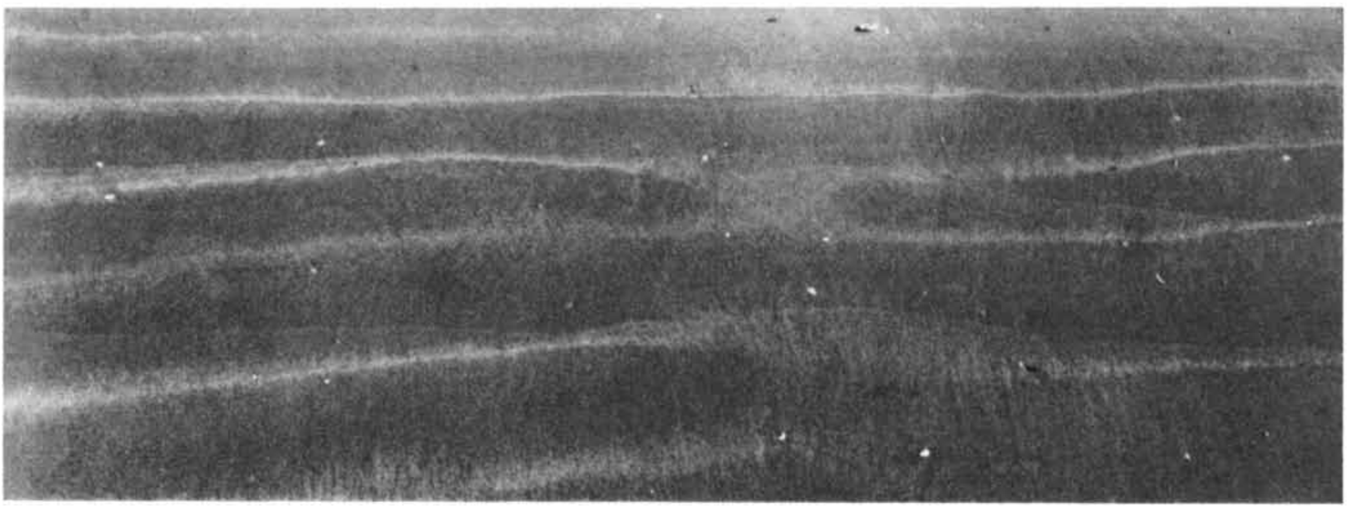
Walking on the shore, watching the waves and thinking on how they shape the beach

by Jearl Walker

A keen-eyed observer walking along a sea beach can find many interesting things having to do with the movement of water onto and away from the shore. They include tiny crisscrosses etched into the sand by the surf as it spends itself on the shore and large-scale cusps dug out by the incoming waves. In addition one sees the ceaseless and highly varied movements of the surf itself.

One thing to notice is that whereas the waves in deep water move in many directions, the waves coming onto a beach have a uniform direction of movement that is more or less perpendicular to the shoreline. The explanation starts with the fact that the waves are slowed as they move into the shallow water near the shore and begin to "feel" the bottom. A wave coming into shallow water at an angle that is not perpendicular to the shoreline is slowed first at the inshore end while the remainder of the wave front is still traveling relatively fast. The result is that the wave is slewed so that its direction of travel becomes perpendicular to the shoreline. In other words, the wave is refracted.

Another thing to notice is that the waves in deep water have a fairly uniform undulating shape, whereas waves in shallow water display a variety of shapes. Four groups of breaking waves can be identified. First, in what is called spilling, the wave retains its usual shape but a layer of foam spills down the forward slope. Second, a plunging breaker develops when the crest of the wave overruns the lower section and drops



Swash marks left on a sea beach by waves at high tide and receding tide

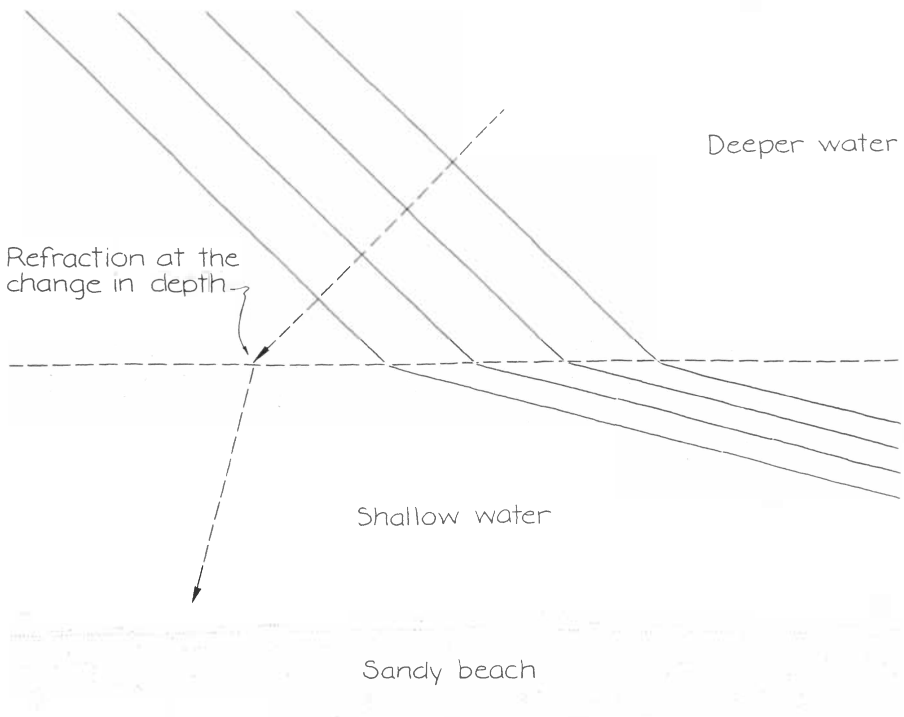
A complex pattern left in the sand by a receding tide in front of the base in a sheet. (Surfboarders value this kind of breaker. In "shooting the pipeline" the surfer maneuvers his board to run under the sheet of falling water.) Third, a collapsing wave breaks into foam and turbulence forward of the crest. Fourth, in surging, the wave quietly ascends the sloping shore with little turbulence.

I have examined beaches and breaking waves in two places: along the ocean near Charleston and Kiawah Island in South Carolina and along Lake Erie near Cleveland. In South Carolina I had the aid of Mary Golrick. Along the shore we found that many waves showed several of the breaking characteristics simultaneously. For example, part of the wave could be spilling while another stretch was plunging, or a spilling wave could begin to collapse or surge. Even at a particular spot the appearance changed from wave to wave, presumably because of variations in the height of the waves and the speed of the wind. At some places the shape of the waves was more consistent, probably because the configuration of the bottom determined what type of breaker would predominate.

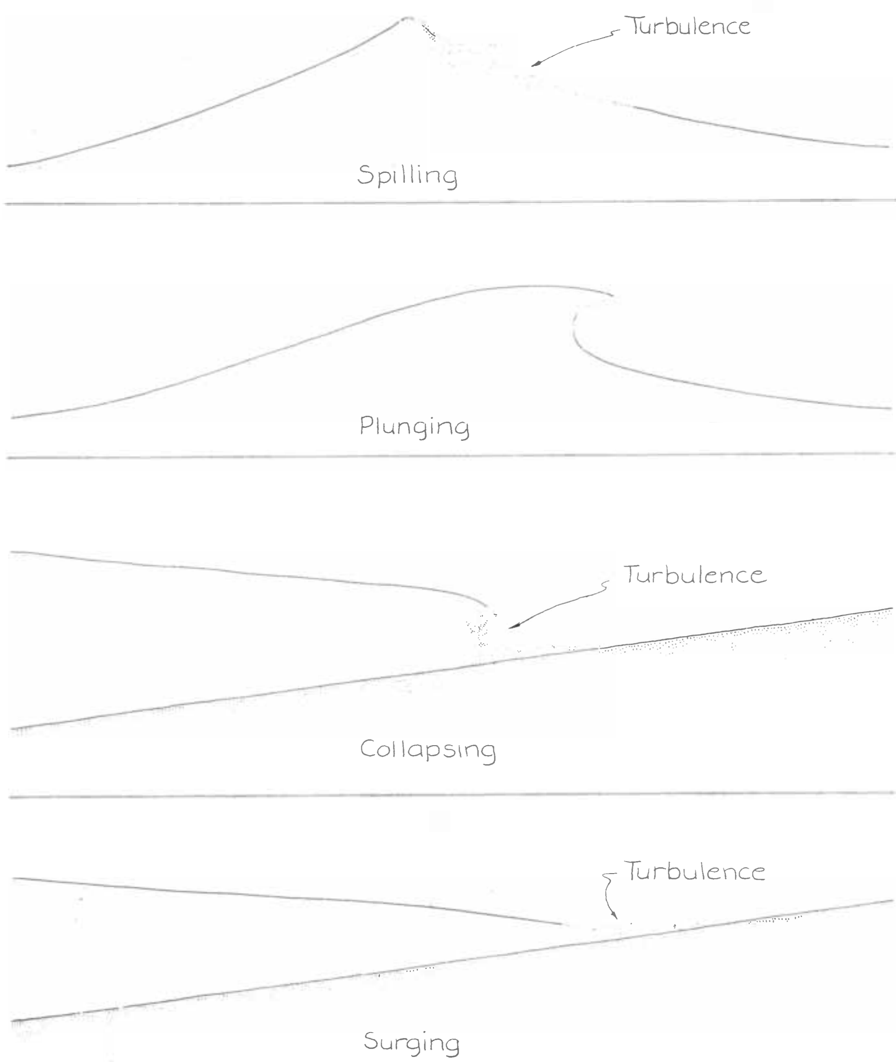
Often a wave rolling up the beach collapsed or plunged, creating a nearly vertical wall of highly turbulent water. For a time such a wave would move up the beach fast and noisily, but then it would abruptly lose its energy and become slower. These waves are examples of the solitary waves called bores or hydraulic jumps, which I described in this department for April, 1981.

What happens is that as a surge of water approaches the beach it begins to move faster than a wave is physically able to travel over the surface of water. This motion is called supercritical flow. It is relieved by a shock transition to subcritical flow brought about by an elevation of the water level, which enables the wave to move faster. The transition develops because the supercritical flow is unstable to the many perturbations introduced by irregularities in the sand bottom. It is the shock transition that comes loudly up the beach as a vertical wall. The energy of the wall is dissipated against the rise of the beach.

Even the relatively docile surging waves Golrick and I observed turned into bores as they advanced on the beach. The vertical wall, however, was often no more than a centimeter in height. Just ahead of it one sees a shallower layer of water the vertical wall seems to be pushing along. In the shallower layer we could see small waves separated by half a centimeter or less, that is, they had a wavelength of the same dimensions. The small waves appeared to be stationary with respect to the vertical wall. Although they could have been formed by the wind or irregularities in the beach, they are part of the



The refraction of waves near the shore



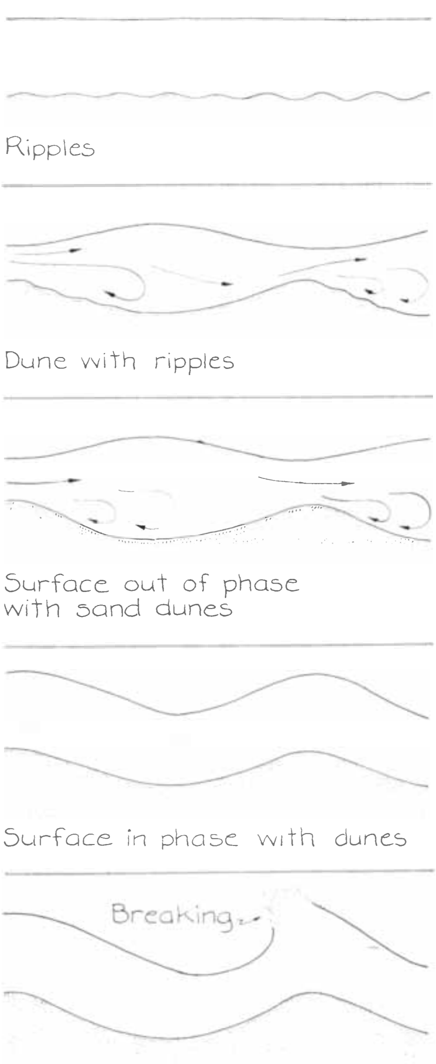
Four types of breaking wave

instability associated with supercritical flow and so are part of the bore.

A wave that had reached its upper limit on the beach seemed to recede faster than it had risen. We were not able to time the movement accurately enough to verify this impression, but the action of the running water around our feet suggested that the impression is correct. A receding wave removed more sand around my feet than an incoming one.

Sometimes the receding water displays the sudden change in depth that characterizes a bore. A layer of water running down a sloping beach is of course accelerated by gravity. The flow can become supercritical and thus unstable to perturbations introduced by the sand bottom. Waves generated by the perturbations create the sudden increase in water depth that marks the shock transition to subcritical flow.

It was interesting to see an outgoing flow meet an incoming wave. Even with a strong incoming bore the result was often a standoff as the backwash dissipated the bore's energy. The bore was reduced to a quiet surge unless it was overtaken from the rear by a new bore.



Various waves in sand

An intriguing feature of the water in a backwash is the appearance of capillary waves. Such waves are governed by surface tension rather than by gravity. Golrick and I had trouble tracking them because they move fast and also because they are hard to see unless the light is reflected from them at a certain angle. My impression is that capillary waves seldom have more than three or four crests. The waves move faster than the line of water retreating from the shore does, and they often move in a different direction. I do not know what causes capillary waves, but I would guess they result from a gust of wind or from an irregularity in the sand.

Except in shallow areas along a beach a given parcel of water associated with a wave has no net motion. In deep water such a parcel moves in a vertical circle as a wave passes over it. Closer to the beach the movement is elliptical with the short axis vertical. Along the beach, however, a parcel in an advancing or retreating sheet of water hops. The hopping movement picks up grains of sand that are then carried along with the moving sheet. Watching a mild backwash, I could follow the grains as they were picked up, carried and dropped.

A backwash running over a small object buried in the sand gouges a short trough seaward of the object. The trough behind a shell I buried in sand was about two centimeters long. Apparently the outgoing water grew turbulent enough seaward of the shell to lift grains of sand and carry them for that distance.

Waves can leave a variety of patterns in sand. At low tide Golrick and I found irregularly spaced marks that were concave toward the ocean. The marks delineated the upper reaches of waves that had come in at high tide. Each runup leaves a high-reach line on the beach because it has carried sand to that point before receding. While the tide is coming in most of these lines are erased by backwash, but while the tide is going out a series of unerased swash marks remains on the beach.

Seaward of a swash mark on some beaches one is likely to find smaller diamond-shaped markings left by the backwash. Willard Bascom discusses them in his book *Waves and Beaches: The Dynamics of the Ocean Surface,* a valuable source of information for any student of beach physics. He says the marks are made by the backwash on a beach with a fairly steep slope. On a beach with a gentle slope the backwash does not move fast enough to dig these tiny valleys, but even there you can find small, irregular lines that point more or less toward the ocean.

Many small domes can be found on a sand beach. They are made by the water sinking into the sand. Before the water arrives the spaces between grains of sand are mostly filled with air. The sinking water pushes air upward through a small opening. You can usually spot such an opening by the bubbles the escaping air makes in the adjacent water.

A dome forms when successively deeper layers of sand are affected by the sinking water. One incoming wave wets a layer of sand near the surface. If more water enters from later waves, the air in the deeper reaches of sand is trapped by the wet top layer of sand. Water moving into the spaces between grains of deeper sand increases the pressure of the trapped air, which pushes the surface of the sand up into a dome.

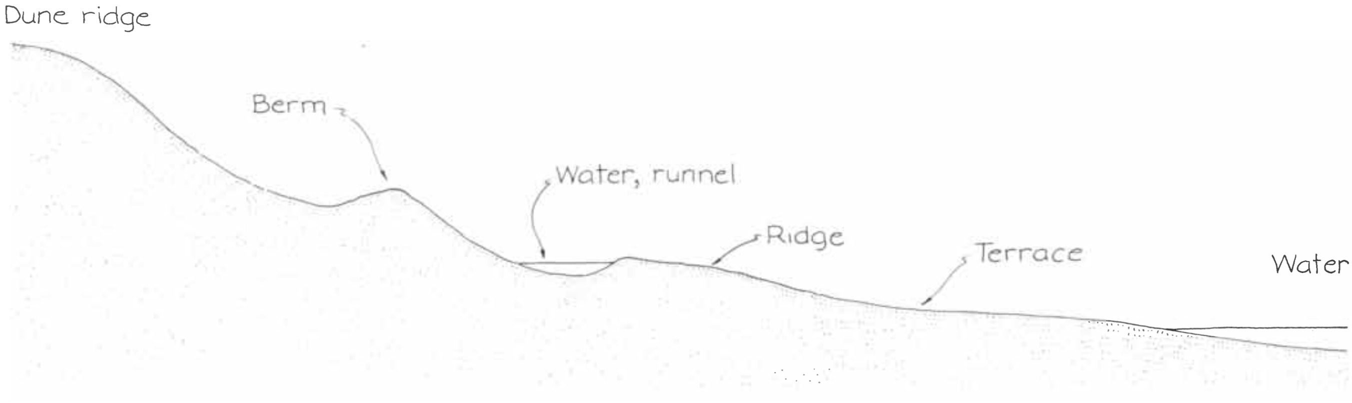
On both the damp area of beach left by a receding tide and the area still being washed by waves one finds systems of sand waves with wavelengths of several centimeters. They are complex structures because they are created by moving water and then they alter the movement of the water. Bascom describes two types. In one type the moving water causes the crests of the sand waves to oscillate, with individual grains of sand being transported back and forth in the trough of the wave. In the second type the crests are almost stationary. As water sweeps over them, swirls are set up in the troughs. You can see the deposition of sand grains in a swirl as the receding water becomes thinner.

Water left in a trough usually drains along it, since the sand waves are seldom exactly perpendicular to the slope of the shore. A breeze will send scores of capillary waves along a trough of water. The troughs make even shallow sand waves easy to see because their crests stand above the flowing water.

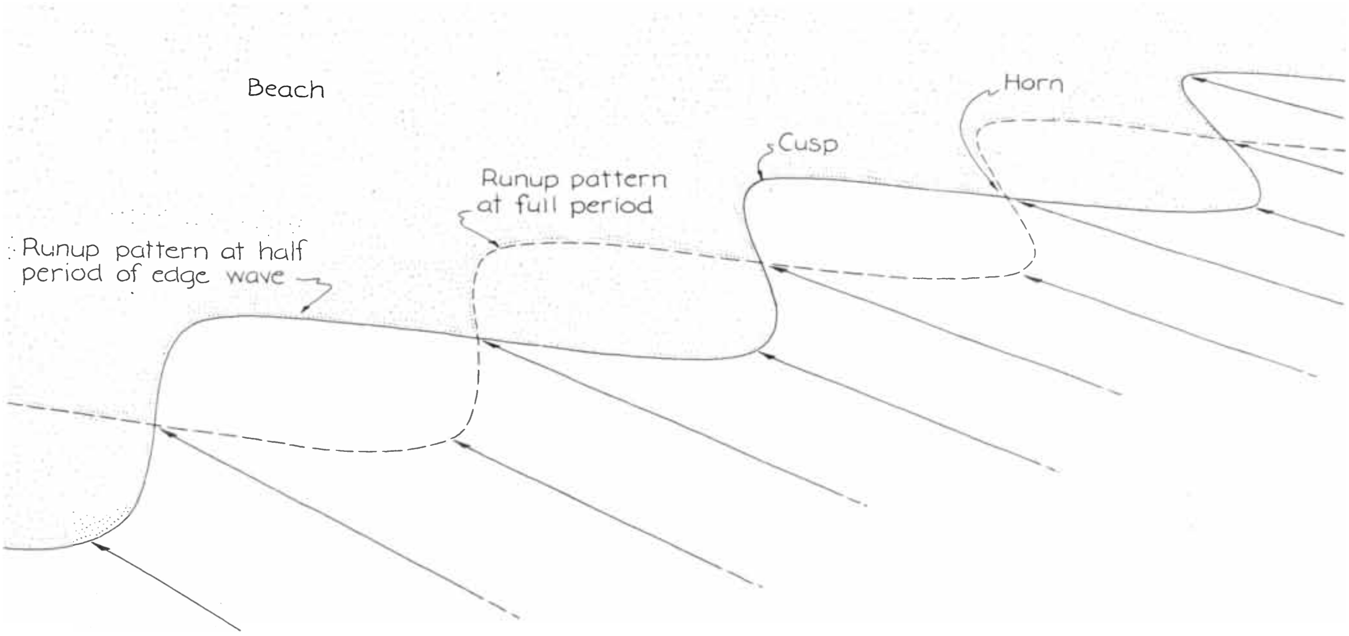
The shape of sand waves depends on how fast the surf is moving. Shapes range from fine ripples to large dunes. The dunes modify the flow of water. When the waves of water are out of phase with the dunes, the water flowing over a dune breaks into a vortex that digs into the valley between the dune and the next dune. Because of the flow the dunes are irregularly shaped. When waves of moving water are in phase with dunes, the shape of the dunes is smoother and rounder.

On a larger scale a typical beach usually has several variations in height superposed on the general slope toward the water line. The structure is best seen at low tide. At the top of the beach are dune ridges that define the upper boundary of the beach. Below them is a smaller elevation of sand called a berm. It is the line the water reaches at high tide. Another ridge may lie between the berm and the ocean at low tide. Between the ridge and the berm may be a small pool of water, which is called a runnel. A terrace of sand may run from the ridge to the low-tide line.

The creation and destruction of these features are major concerns for engineers studying the erosion of beaches. A severe storm is likely to change the features drastically. Even in the ab-



Features of a beach



How cusps are formed by the interaction of an incoming wave and an edge wave

sence of storms the features are slowly modified by the tides.

The beaches along Lake Erie do not show the effects of tide and appear to be affected more by storms than by a gradual movement of sand up onto the shore. On one shore near Cleveland the lake floor drops off sharply only about a meter from the line representing the upper reach of incoming waves. The ridge of sand left by those waves is narrower but more prominent than the wave-built ridge on Kiawah Island. I found no runnels along the lake, although the high-water debris inshore of the ridge shows that at times the waves reach higher than the ridge.

Many shorelines exhibit cusps, which constitute one of the strangest and prettiest sights to be seen on a beach. A cusp is a thin sheet of water forming a horizontal curve as it washes up the slope of the beach. Between each cusp and the next one is a section of beach forming a curve in the opposite direction. The diameter of a cusp can range from less than a meter along the shore of a lake to 100,000 meters or more for major shoreline features along an ocean. The large-scale features are often called giant cusps or shoreline rhythms. Sometimes they are so irregularly spaced that they go unnoticed. Elsewhere the periodicity of the cusps is so precise that the shoreline looks artificial.

According to Bascom, the cusp pattern is maintained by the interaction of two waves that impinge consecutively on the shore. The first one enters an existing cusp carrying a load of suspended sand. Then the wave divides so that the sand is carried to the inner border of the cusp. During the backwash the water flows to the center of the cusp, where there is a troughlike configuration that carries it back to the sea.

At the mouth of the trough the incoming second wave is arrested by the backwash, whereas at the horns of the cusp the new wave is essentially undeterred. It is therefore able to bring in more sand and to deposit it along the inner border of the cusp. The process continues with succeeding waves.

The source of the periodicity of cusp formations along a shore was unknown until recently. Now a number of workers, including Anthony J. Bowen of Dalhousie University and Robert T. Guza and Douglas L. Inman of the Scripps Institution of Oceanography, have ascribed the periodicity to edge waves, which are created along a shoreline by other waves coming in from deep water.

An edge wave varies sinusoidally in height on an axis parallel to the shore, as is shown in the upper illustration on the next page. Points of maximum and minimum height are antinodes; between them are nodes, where the water level does not vary. (The illustration shows only the variations in height due to the edge waves. On a sea beach the variations caused by the incident deep-water waves are superposed on the edge waves to yield the observed differences in water height.)

Seen in cross section along an axis perpendicular to the shore an edge wave has several modes representing variations in its amplitude. They are shown in the lower illustration on the next page. In the simplest mode the amplitude of the edge wave drops off continuously toward the deep water; this is the zero mode. I shall focus on it because it seems to have the most to do with the features that are due to edge waves.

Edge waves are sometimes difficult to see on a natural beach because the waves coming in from deep water vary so much in shape, strength and direction and also because of irregularities along the bottom. To simplify matters Guza and Inman experimented with edge waves formed in a wave basin at Scripps. Their beach was a sloping section of concrete. The deep-water waves were generated by a plunger.

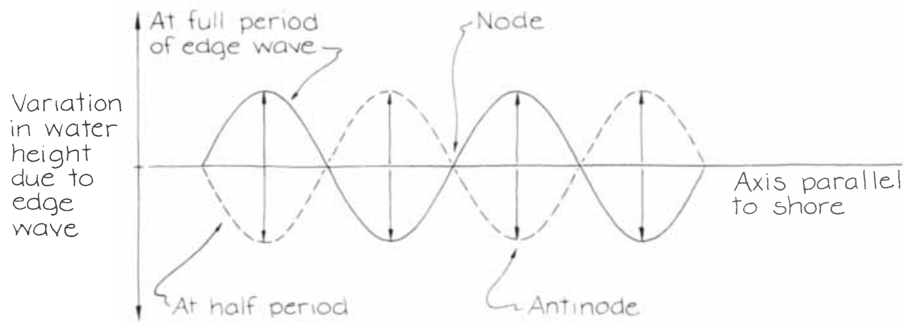
At the nodes of the edge waves the runup of water onto the shore was the same for each wave coming in from deeper water. At a given antinode the runup varied periodically, being at a maximum when the incoming wave was in phase with the high-water level of the edge wave and at a minimum when the two were out of phase.

Since the period of the edge wave's variations in height was twice the period of the waves from deep water, the observers saw an alternation between constructive and destructive interference. If one deep-water wave arriving at a particular antinode of the edge wave interfered constructively with the edge wave, yielding a large runup, the next interfered destructively and the runup was small. Since this periodic interference took place at every antinode along the shore, the borderline between wet and dry concrete on the shoreline was formed into a pattern of cusps centered on the antinodes. The horns of the cusps marked the nodes.

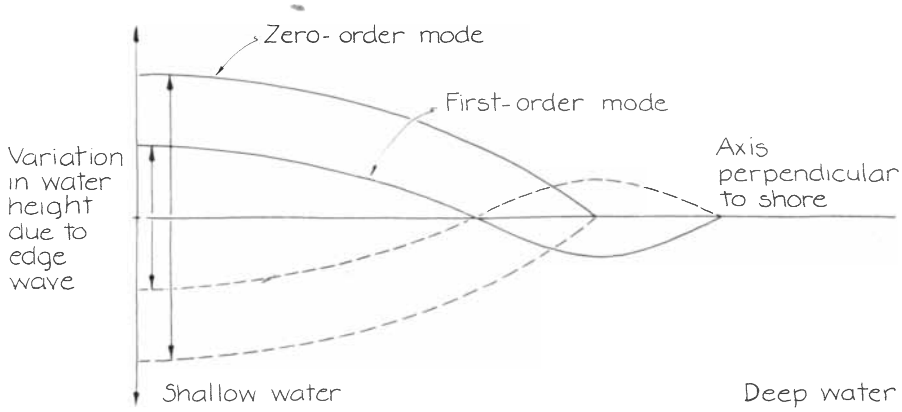
To simulate an erodable beach Guza and Inman covered the concrete with a thin layer of fine sand, adding more as the edge waves developed. At the antinodes of the edge waves the sand was eroded by the swash of water. Each maximum runup lifted and suspended the sand and then redeposited it in the area of the nodes. Soon the characteristic cusp formation appeared in the sand. With time the horns of the cusps extended toward the deeper water and grew higher.

As more sand was added a cusp was likely to show a buildup of sand down the center, where the flow of water was less vigorous. Each such cusp soon split into two cusps, thereby doubling the periodicity of cusps along the shore. This arrangement of a small periodicity laid down on a larger dominant pattern can be seen along some natural beaches.

As even more sand was added the pattern of cusps began to disintegrate. Apparently the sand formations themselves began to interfere with the waves coming in from deep water and with the ability of those waves to generate edge waves. Probably what happens on a real beach is that the incoming waves pump energy into edge waves until the edge waves have redistributed a certain amount of sand. Then the transfer of energy ceases and the patterns in the sand disappear.



Edge-wave variations parallel to a shore



Edge-wave variations perpendicular to a shore

SCIENCE/SCOPE

Intelsat VI communications satellites can be flown on either the Ariane 4 or the Space Shuttle. For an Ariane launch, the spacecraft is mated to the booster with a conical adaptor and clamp. The compact launch arrangement provides maximum payload capacity. A Shuttle launch is also economical because of the satellite's length-to-weight ratio. Its overall length in the Shuttle is only 44.6% of the orbiter bay; its weight uses 45.4% of available capacity. Hughes heads an international team building Intelsat VI for the International Telecommunications Satellite Organization.

Ambitious manned missions in space will become possible with development by NASA of advanced solar-power platforms capable of electric-power output as high as 100 kilowatts or more. Hughes, under contract to the NASA Lewis Research Center, is building a breadboard 25-kilowatt dc-to-dc power converter that uses a transistorized series-resonant inverter. This module provides the basic building block required to match solar-array characteristics to payload requirements. The converter will use technology demonstrated by a lightweight 10-kilowatt converter that operated at over 91% electrical efficiency.

Electronic circuits as complex as 100 Los Angeles street maps printed on a thumb tack are being developed for military systems of tomorrow. The so-called Very High Speed Integrated Circuits (VHSIC) will give military electronics systems a tenfold increase in data processing capability. These "super chips" will be more reliable and need less power than the integrated circuits now in use. Hughes is one of six firms involved in a tri-service program to develop VHSIC. In addition to developing chips for various high-speed signal processing uses, Hughes will develop the VHSIC high-speed electron-beam lithography system to process the chips, a requirement of all VHSIC contractors. Hughes also will build a demonstration processor for a two-way system for Army troops to communicate among themselves and find and report their positions automatically.

A new anti-jam, secure communications system is now under development by the U.S. Navy. A family of JTIDS (Joint Tactical Information Distribution System) terminals will be deployed aboard surface ships, command aircraft, and fighters. Using the Navy's advanced DTDMA (Distributed Time Division Multiple Access) system, the fleet will be fully interoperable with the U.S. Air Force and NATO. Full communication, navigation, and identification functions will be provided even in a hostile electronic environment. TADCOM, a joint venture of Hughes and ITT Avionics Division, is developing this newest generation of JTIDS equipment.

Hughes is seeking engineers to develop advanced systems and components for many different weather and communications satellites, plus the Galileo Jupiter Probe. Immediate openings exist in applications software development, data processing, digital subsystems test, microwave/RF circuit design, power supply design, digital communications, signal processing, spacecraft antenna design, system integration test and evaluation, and TELCO interconnection. Send your resume to Tom W. Royston, Hughes Space & Communications Group, Dept. SE, Bldg. S/41, M.S. A300, P.O. Box 92919, Los Angeles, CA 90009. Equal opportunity employer.

Creating a new world with electronics

HUGHES

HUGHES AIRCRAFT COMPANY
CULVER CITY, CALIFORNIA 90230

(213) 670-1515 EXTENSION 5964

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

THE COMPLEAT STRATEGYST. J. D. Williams. McGraw-Hill Book Company, 1954.

GAME THEORY: MATHEMATICAL MODELS OF CONFLICT. A. J. Jones. Ellis Horwood, Ltd., 1980.

A UNIFIED MODEL FOR BARGAINING AND CONFLICT. Douglas Heckathorn in *Behavioral Science,* Vol. 25, pages 261–284; 1980.

CARBON DIOXIDE AND WORLD CLIMATE

ENERGY AND CLIMATE. Geophysics Study Committee, Geophysics Research Board, Assembly of Mathematical and Physical Sciences. National Research Council, National Academy of Sciences, 1977.

ON THE DISTRIBUTION OF CLIMATE CHANGE RESULTING FROM AN INCREASE IN CO_2 CONTENT OF THE ATMOSPHERE. Syukuro Manabe and Richard T. Wetherald in *Journal of the Atmospheric Sciences,* Vol. 37, No. 1, pages 99–118; January, 1980.

MAJOR CLIMATIC EVENTS ASSOCIATED WITH A PROLONGED CO_2-INDUCED WARMING. Hermann Flohn. Institute for Energy Analysis, Oak Ridge Associated Universities, October, 1981.

MICROBIOLOGICAL MINING

MICROBIAL TRANSFORMATIONS OF METALS. A. O. Summers and S. Silver in *Annual Review of Microbiology,* Vol. 32, pages 637–672; 1978.

BACTERIAL LEACHING. Corale L. Brierley in *CRC Critical Reviews in Microbiology,* Vol. 6, pages 207–262; November, 1978.

MICROBIAL METHODS FOR THE EXTRACTION AND RECOVERY OF METALS. D. P. Kelly, P. R. Norris and C. L. Brierley in *Microbial Technology: Current State, Future Prospects (The Society for General Microbiology, Symposium 29),* edited by A. T. Bull, D. C. Ellwood and C. Ratledge. Cambridge University Press, 1979.

GEOMICROBIOLOGY. Henry Lutz Ehrlich. Marcell Dekker, Inc., 1981.

THE CORONAS OF GALAXIES

ON A POSSIBLE INTERSTELLAR GALACTIC CORONA. Lyman Spitzer, Jr., in *The Astrophysical Journal,* Vol. 124, No. 1, pages 20–34; July, 1956.

ULTRAVIOLET ABSORPTION BY INTERSTELLAR GAS AT LARGE DISTANCES FROM THE GALACTIC PLANE. Blair D. Savage and Klaas S. de Boer in *The Astrophysical Journal,* Vol. 243, No. 2, Part 1, pages 460–484; January 15, 1981.

SEARCHING BETWEEN THE STARS. Lyman Spitzer, Jr. Yale University Press, 1982.

GAS IN THE GALACTIC HALO. D. G. York in *Annual Review of Astronomy and Astrophysics,* in press.

ALLERGY

NASAL ALLERGY. Niels Mygind. Blackwell Scientific Publications, 1978.

THE MAST CELL: ITS ROLE IN HEALTH AND DISEASE. Edited by J. Pepys and A. M. Edwards. University Park Press, 1979.

THE MAST CELL—A FASCINATING RIDDLE. Michael A. Kaliner in *The New England Journal of Medicine,* Vol. 301, No. 9, pages 498–499; August 30, 1979.

IMMUNOLOGICAL AND CLINICAL ASPECTS OF ALLERGY. Edited by M. H. Lessof. J. B. Lippincott Co., 1981.

ANALYSIS OF TRIGGERING EVENTS IN MAST CELLS FOR IMMUNOGLOBULIN E–MEDIATED HISTAMINE RELEASE. Teruko Ishizaka in *The Journal of Allergy and Clinical Immunology,* Vol. 67, No. 1, Part 2, pages 90–96; January, 1981.

SURFACE DIFFUSION

SURFACE DIFFUSION. S. Chandrasekhar in *Reviews of Modern Physics,* Vol. 15, No. 1, pages 2–89; January, 1943.

FIELD EMISSION AND FIELD IONIZATION. Robert Gomer. Harvard University Press, 1961.

FUNDAMENTALS OF STATISTICAL AND THERMAL PHYSICS. F. Reif. McGraw-Hill Book Company, 1965.

FIELD ION MICROSCOPY: PRINCIPLES AND APPLICATIONS. Erwin W. Müller and Tien Tzou Tsong. American Elsevier Publishing Co., 1969.

CHEMISORPTION ON METALS. Robert Gomer in *Solid State Physics,* Vol. 30, pages 93–225; 1975.

DIFFUSION OF HYDROGEN AND DEUTERIUM ON THE 110 PLANE OF TUNGSTEN. Rocco DiFoggio and Robert Gomer in *Physical Review B,* Vol. 25, No. 6, pages 3490–3511; March 15, 1982.

THE COEVOLUTION OF A BUTTERFLY AND A VINE

SEARCH IMAGE FOR LEAF SHAPE IN A BUTTERFLY. Mark R. Rausher in *Science,* Vol. 200, No. 4345, pages 1071–1073; June 2, 1978.

PLANT CHEMISTRY AND THE EVOLUTION OF HOST SPECIFICITY: NEW EVIDENCE FROM *HELICONIUS* AND *PASSIFLORA*. John Smiley in *Science,* Vol. 201, No. 4357, pages 745–747; August 25, 1978.

FOOD WEB ORGANIZATION AND THE CONSERVATION OF NEOTROPICAL DIVERSITY. Lawrence E. Gilbert in *Conservation Biology: An Evolutionary-Ecological Perspective,* edited by Michael E. Soulé and Bruce E. Wilcox. Sinauer Associates, 1980.

INSECTS AS SELECTIVE AGENTS ON PLANT VEGETATIVE MORPHOLOGY: EGG MIMICRY REDUCES EGG LAYING BY BUTTERFLIES. Kathy S. Williams and Lawrence E. Gilbert in *Science,* Vol. 212, No. 4493, pages 467–469; April 24, 1981.

OPHIOLITES

OPHIOLITES: ANCIENT OCEANIC LITHOSPHERE? Robert G. Coleman. Springer-Verlag, 1977.

OPHIOLITES: PROCEEDINGS OF AN INTERNATIONAL OPHIOLITE SYMPOSIUM HELD IN NICOSIA, CYPRUS, IN 1979. Edited by A. Panayiotou. Geological Survey Department, Nicosia, Cyprus, 1980.

THE SEA, VOL. 7: THE OCEANIC LITHOSPHERE. Edited by C. Emiliani. John Wiley & Sons, Inc., 1981.

THE GALILEO AFFAIR

THE CRIME OF GALILEO. Giorgio de Santillana. University of Chicago Press, 1955.

DISCOVERIES AND OPINIONS OF GALILEO. Stillman Drake. Doubleday Anchor Books, 1957.

GALILEO GALILEI: A BIOGRAPHY AND INQUIRY INTO HIS PHILOSOPHY OF SCIENCE. Ludovico Geymonat. McGraw-Hill Book Company, 1965.

THE CENSORSHIP OF COPERNICUS' *DE REVOLUTIONIBUS*. Owen Gingerich in *Annali dell'Istituto Museo di Storia della Scienze di Firenze,* Vol. 4, No. 2, pages 45–61; 1981.

THE AMATEUR SCIENTIST

EDGE WAVES AND BEACH CUSPS. Robert T. Guza and Douglas L. Inman in *Journal of Geophysical Research,* Vol. 80, No. 21, pages 2997–3012; July 20, 1975.

BREAKING WAVES. E. D. Cokelet in *Nature,* Vol. 267, No. 5614, pages 769–774; June 30, 1977.

EDGE WAVES AND SURF BEAT. Anthony J. Bowen and Robert T. Guza in *Journal of Geophysical Research,* Vol. 83, No. C4, pages 1913–1920; April 20, 1978.

WAVES AND BEACHES: THE DYNAMICS OF THE OCEAN SURFACE. Willard Bascom. Doubleday Anchor Books, 1980.

Pack a lab in your bag.

A quality processing lab. Filled with the experience and expertise it takes to do justice to the creativity of your 35 mm photography.

You can get your quality Kodak color processing started without wait, worry, or carrying around your exposed rolls. Simply stock your bag with Kodak processing mailers. Then all you have to do is shoot, stamp, and drop in the nearest mailbox, here or abroad. Often, your prints or slides will be waiting for you when you return from your trip.

So choose your processing as carefully as you choose your camera and film. And choose the shoot-and-send convenience of Kodak mailers.

Kodak

Kodak color processing. Quality worth asking for.

WE'RE SHAPING OUR CARS FOR
AND LESS FRICTION

Overcoming the wind. It's absolutely critical to achieve good mileage.

At 55 mph, over half of a car's fuel is used just to overcome aerodynamic drag.

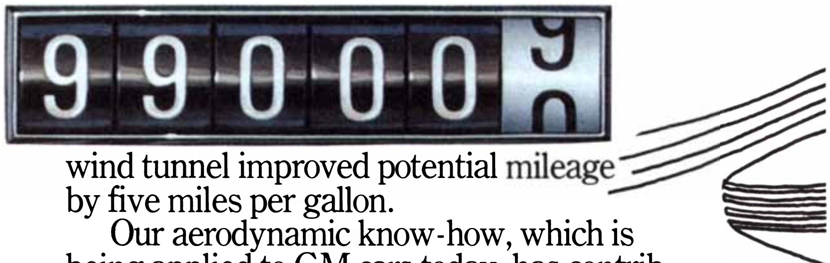
Aerodynamics.

At GM, we've put millions of dollars into aerodynamic research. That includes everything from modifying the overall shape of our cars to scrutinizing details like the angle of the outside mirrors.

GM is the only manufacturer in the U.S. with its own full-scale automotive wind tunnel—and it is the most advanced. It gives us an opportunity to build the most aerodynamic cars in the world.

Great gas mileage.

The aerodynamic efficiency of a five-passenger concept vehicle developed in our

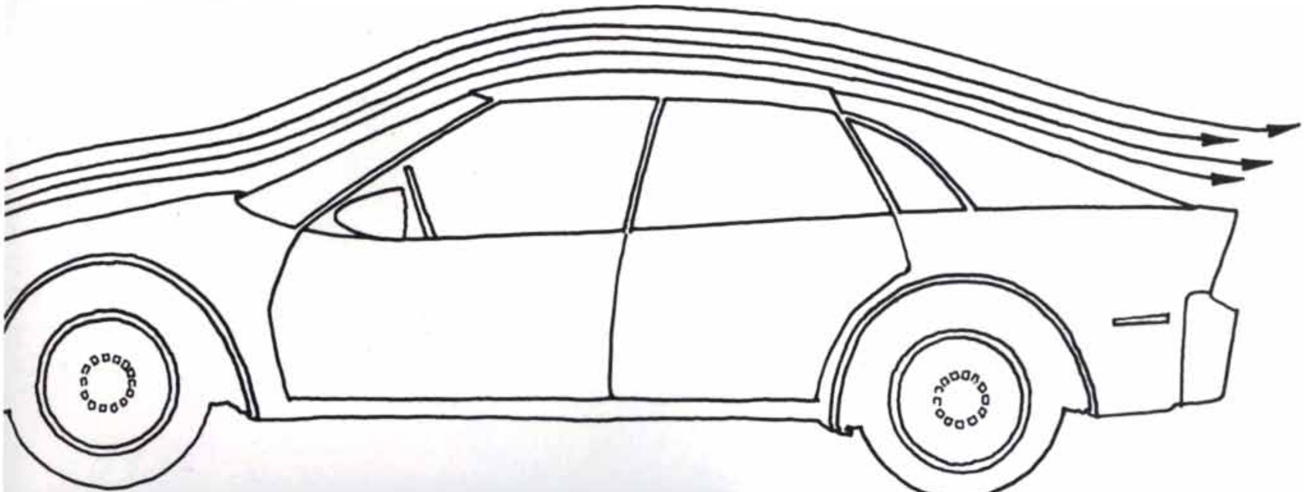


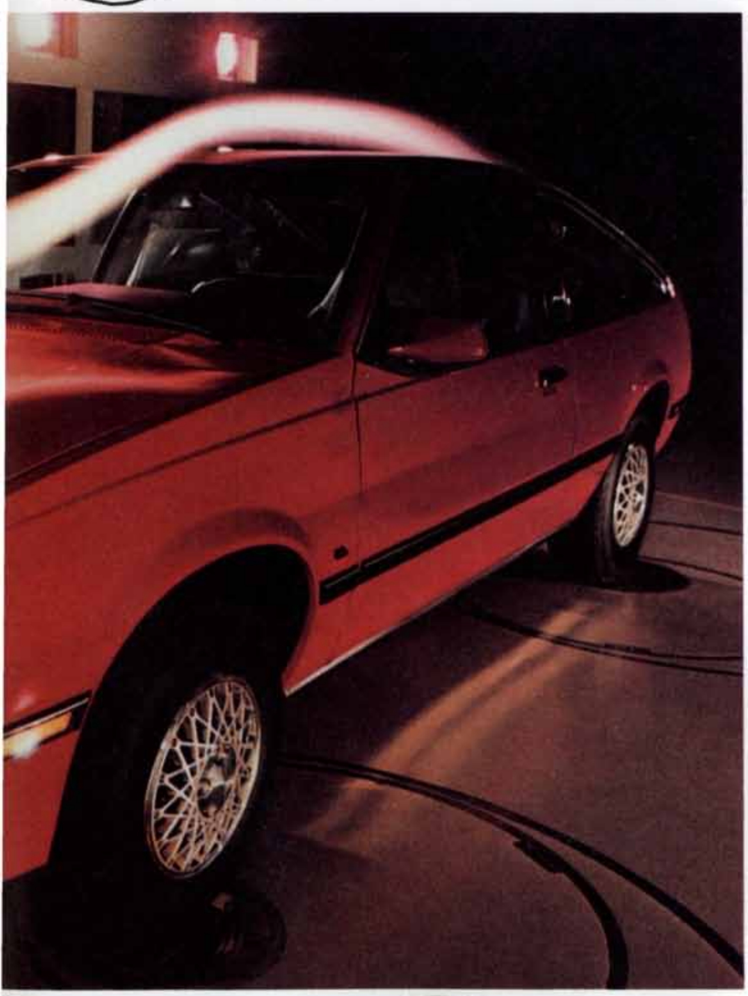
wind tunnel improved potential mileage by five miles per gallon.

Our aerodynamic know-how, which is being applied to GM cars today, has contributed to the 100% mileage improvement our cars have achieved in the last eight years.

WE'RE THE B

LESS DRAG ON YOUR WALLET, AT THE GAS PUMP.





Great looks.

Aerodynamically designed cars look terrific. Sleek, sophisticated and elegant.

Corners are rounded, surfaces are smooth, trim is minimized. Great mileage married to great appearance.

A new kind of excellence.

We believe that GM stands for something special in the eyes of the American public.

It stands for excellence.

And continuing excellence. So that now, next year, and the years after that, we need always to be:

the best GM ever.

GM
MARK OF EXCELLENCE

Chevrolet
Pontiac
Oldsmobile
Buick
Cadillac
GMC Truck

EST GM EVER