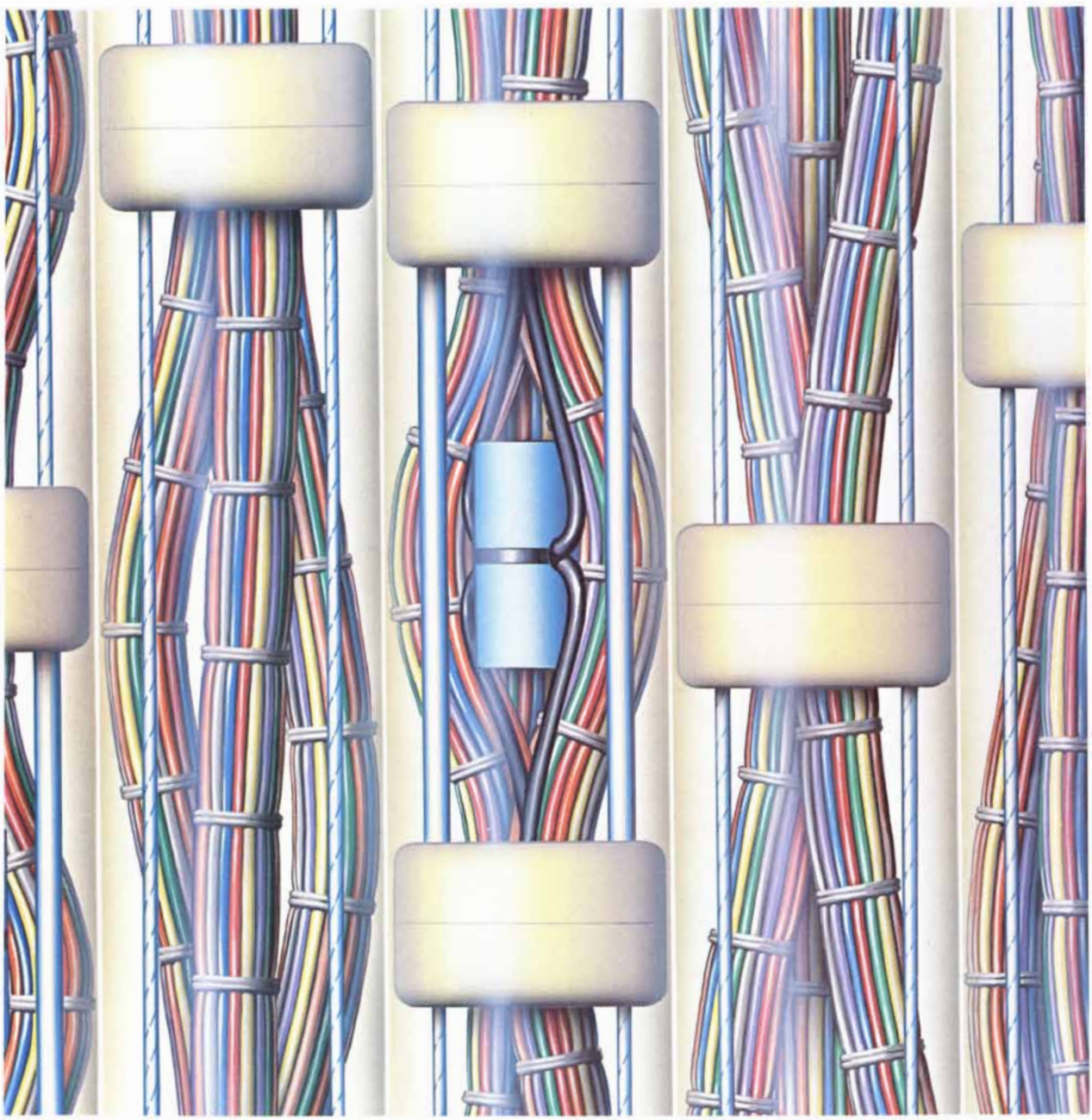
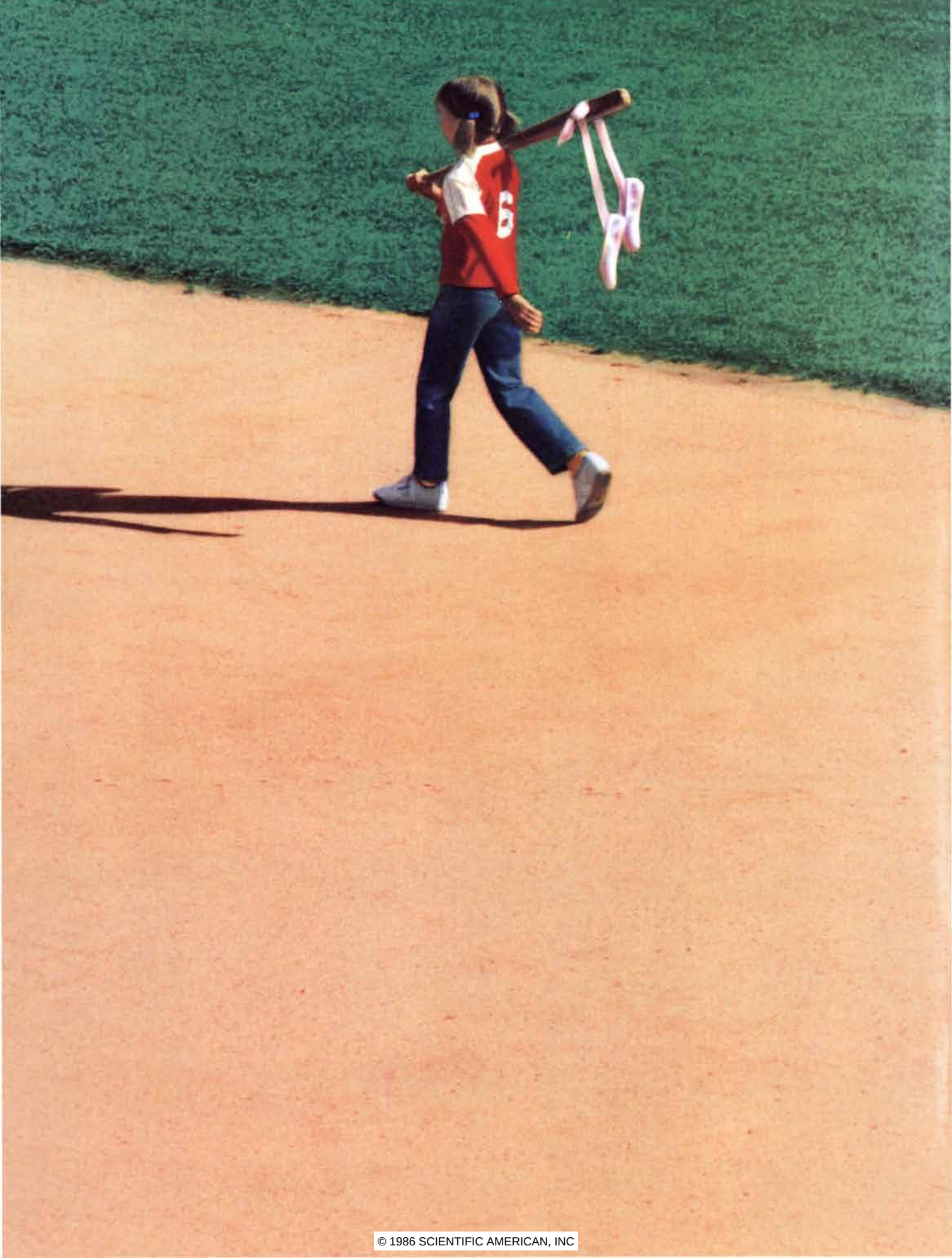
SCIENTIFIC AMERICAN



IMAGING THE EARTH'S CRUST

$2.50

February 1986



Diversification.

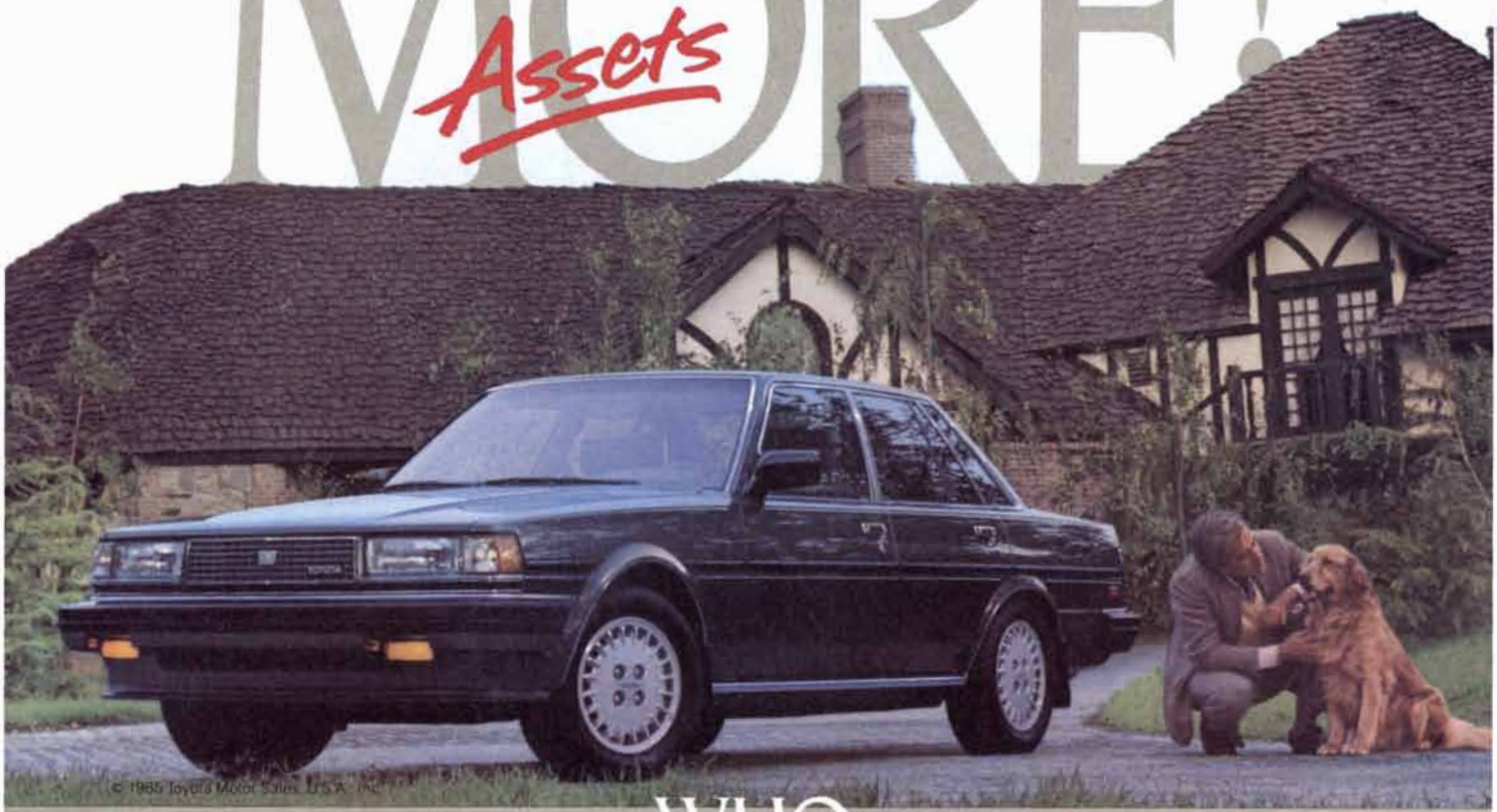
Getting into new things. Because the future belongs to those who anticipate change. To Eaton.

EATON

Growing into the future.

Automotive • Electronics • Defense • Capital Goods

MORE! Assets



WHO COULD ASK FOR ANYTHING MORE! TOYOTA

1986 TOYOTA CRESSIDA.

The measure of a true luxury car's value lies in its tangible assets. And nowhere are assets more tangible than in the 1986 Toyota Cressida.

And with appropriate reason.

Inside, you'll enjoy lavish comfort. Automatic climate control system. Speed-sensing power steering. Power windows and door locks. A state-of-the-art Technics* AM Stereo/FM/MPX with integral cassette and 9-band graphic equalizer. Comfort you'll wish you had at home: rich thick velour or

optional glove-soft leather seats, and a refined Theft Deterrent System. And of course, there's a tilt/slide sunroof available.

Cressida's 2.8 liter, Electronically Fuel-Injected, dual overhead cam, 6-cylinder engine unleashes 156 charging horses for more driving excitement. Stopping power is no less superb thanks to new power-assisted ventilated 4-wheel disc brakes.

Your Electronically Controlled 4-speed automatic overdrive transmission adjusts to your driving needs. Choose "Normal" for fuel savings or "Power" for acceleration. All at the touch of a button.

And its 4-wheel fully independent suspension treats road obstacles with the disdain they deserve.

The new 1986 Toyota Cressida. It puts you into the lap of luxury without requiring a lapse of fiscal acumen.

*Technics is a trademark of the Matsushita Electrical Industrial Co., LTD.

GET MORE FROM LIFE—BUCKLE UP!

Established 1845 **SCIENTIFIC AMERICAN** February 1986 Volume 254 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: $24 PER YEAR IN THE U.S. AND ITS POSSESSIONS; $33 PER YEAR IN ALL OTHER COUNTRIES. POSTMASTER: SEND ADDRESS CHANGES TO SCIENTIFIC AMERICAN, 415 MADISON AVENUE, NEW YORK, N.Y. 10017

SCIENCE/SCOPE®

A two-way television system enables Minnesota students to be taught by teachers miles away. A consortium of seven school districts in east-central Minnesota operates this system in conjunction with local cable television operators as part of an elementary and high-school program to promote telecommunications technology in education. It allows each district to produce and transmit live video and audio signals to other schools. Thus, specialized classes such as foreign languages and advanced mathematics and science are shared among the seven districts. The system also allows administrators to attend regional meetings without traveling. It consists of a microwave path from Cambridge to six outlying school districts 6 to 26 miles away. The microwave transmitters and receivers were manufactured and provided by Hughes Aircraft Company.

A new ion thruster that operates on xenon gas has proven to be vastly superior to previous designs that operated on mercury propellant, offering a number of benefits to spacecraft designers. The xenon thruster would be used by communications satellites to stay precisely on station 22,300 miles above Earth. It is much simpler mechanically and yet more powerful than the mercury ion thrusters developed earlier. In fact, it could reduce the weight of a large satellite by 15%—meaning that launch costs could be trimmed, or more communications equipment could be packed into the same satellite. The thruster would operate for about an hour and a half each day, accumulating 5,000 hours during a 10-year period. Hughes Research Laboratories conducted the studies under contract to INTELSAT.

A private company has assumed the operation of U.S. earth resources satellites, the purpose being to save taxpayers money. Earth Observation Satellite Company (EOSAT), a joint venture of Hughes and RCA, is now operating Landsats 4 and 5. EOSAT will be the primary source for marketing, ordering, and distributing data from these spacecraft. In addition, EOSAT will begin constructing the next-generation commercial Landsat spacecraft. Landsats previously had been acquired by NASA then turned over to the Commerce Department's National Oceanic and Atmospheric Administration. For the past 13 years, Landsat data has been used to map vegetation, mineral resources, and water pollution, and to update land use maps. The new spacecraft, Landsat 6 and 7, are being built by RCA and will carry imaging devices built by the Santa Barbara Research Center, a Hughes subsidiary.

Computers monitor the work flow at a Hughes facility for making printed circuit boards for advanced missiles. Once planning instructions are entered into the network, planning route sheets and tool sheets are printed and follow the work order through the shop. Route sheets are printed with bar-code labels so work can be logged in after each operation. The bar codes also are used to log in quality inspections. The computer network allows management to immediately determine the status of any program or of any specific piece of hardware. The facility is located in Tucson, Arizona.

Support Systems in southern California designs, develops, and manufactures some of the most sophisticated training simulators and a wide array of automatic and manual test systems. In addition, field engineering and technical support of a wide range of electronic systems keep Hughes' systems operating at top efficiency worldwide. Opportunities are available for a variety of engineers qualified by degree or extensive work experience. They include systems engineers, radar engineers, and software and hardware design engineers. Please send your resume to Lowell Anderson, Professional Employment, Dept. S2, Hughes Aircraft Company, P.O. Box 9399, Long Beach, CA 90801-0463. Equal opportunity employer. U.S. citizenship required.

For more information write to: P.O. Box 45068, Dept. 78-9, Los Angeles, CA 90045-0068

HUGHES
AIRCRAFT COMPANY

Subsidiary of GM Hughes Electronics

SCIENTIFIC AMERICAN

CORRESPONDENCE

Offprints of more than 1,000 selected articles from earlier issues of this magazine, listed in an annual catalogue, are available at $1.25 each. Correspondence, orders and requests for the catalogue should be addressed to W. H. Freeman and Company, 4419 West 1980 South, Salt Lake City, UT 84104. Offprints adopted for classroom use may be ordered direct or through a college bookstore. Sets of 10 or more Offprints are collated by the publisher and are delivered as sets to bookstores.

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

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

Subscription correspondence should be addressed to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 5969, New York, NY 10017. The date of the last issue on your subscription is shown in the upper right-hand corner of each month's mailing label. For change of address notify us at least four weeks in advance. Please send your old address (if convenient, on a mailing label of a recent issue) as well as the new one.

Name

New Address

Street

City

State and ZIP

Old Address

Street

City

State and ZIP

THE COVER

The painting on the cover shows several coiled sections of a "streamer," a device used by marine seismologists to make structural images of the crust under the sea floor (see "Seismic Images of Plate Boundaries," by John C. Mutter, page 66). A streamer is a plastic tube, often several kilometers long, containing a string of thousands of hydrophones, or sound detectors. (One of these small cylinders is visible at the center of the painting.) The tube is filled with a low-density oil that enables it to float in the water as it is towed behind a research vessel. Stainless-steel cables strengthen the streamer, and plastic cylindrical "spacers" prevent it from collapsing on the hydrophones. As the ship advances, air guns deployed in the water are fired at regular intervals. The low-frequency sound from the guns penetrates the sea floor and is reflected by underlying rock boundaries. The reflections are detected by the hydrophones, converted into electric signals and stored on magnetic tape. Reflections recorded by many hydrophones from the same point on the sea floor can be combined to create enhanced images of the crust.

THE ILLUSTRATIONS

Cover painting by Hank Iken

Page	Source
14–21	Ilil Arbel
30	Ray Pfortner, Peter Arnold, Inc.
31–32	Alan D. Iselin
33–34	Edward Bell
35	Alan D. Iselin
37	Andrew G. Farr, University of Washington School of Medicine
38–45	Bunji Tagawa
47	Paul Weiner, Rutgers University
48–51	Sergei Anosov and Ilil Arbel
52	Aleksandr Chumichov
66–67	John W. Ladd, Lamont-Doherty Geological Observatory (*top*); George Retseck (*bottom*)
68–69	George Retseck
70–71	Lucien Montadert, Institut Français du Pétrole (*top*); George Retseck (*bottom*)
72–73	John C. Mutter, Lamont-Doherty Geological Observatory
74	George Retseck
75	Lamont-Doherty Geological Observatory
76	Andrew Christie
77	Marc Cantin
78	Kevin A. Somerville (*top*), Andrew Christie (*bottom*)
79	Kevin A. Somerville
80	Kevin A. Somerville (*top*), Andrew Christie (*bottom*)
83–92	Patricia J. Wynne
95	Jean McMann (*top*); Jean-Pierre Protzen, University of California at Berkeley (*bottom*)
96–98	Michael Goodman
100	Jean-Pierre Protzen, University of California at Berkeley
101	Michael Goodman
107	Sky Publishing Corporation
108	The Granger Collection
109	Quesada/Burke, courtesy of the New York Public Library
110	Quesada/Burke, courtesy of the New York Public Library (*top*); Royal Astronomical Society (*bottom*)
111	Royal Astronomical Society
112	Quesada/Burke, courtesy of the New York Public Library
115–118	Michael Goodman

LETTERS

To the Editors:

In "The Federal Support of Mathematics" [SCIENTIFIC AMERICAN, May, 1985], Edward E. David, Jr., has argued convincingly that research and education in mathematics are threatened unless funds are increased substantially in the near future.

Being concerned with the relations between technology and the educational system, we find, however, that this problem (which exists also in other industrialized countries, including our own) has deeper roots. In particular we should like to point out that the general opinion of what mathematics is, and can do, is antiquated and incorrect. Everyone knows there are great opportunities for creative work in both the technical and the natural sciences, but to most people outside the scientific community mathematics appears to be static and sterile. In our opinion this is at least part of the reason it has been difficult to convince politicians and administrators of the importance of mathematics and to attract gifted young people.

The role of elementary education in mathematics in forming most people's concept of the subject is therefore of great importance. If it is true—and we are sure it is—that most pupils leave school with an impression of mathematics as meaningless jiggling around of formulas and proving of obvious statements, financial measures alone will not produce durable results.

It is necessary to develop new educational methods that will allow more people outside the scientific community to appreciate the value of mathematics and to take part in the public debate about its role in society.

JENS BJORNEBOE

GUNHILD NISSEN

Roskilde, Denmark

To the Editors:

The article "Cricket Auditory Communication," by Franz Huber and John Thorson [SCIENTIFIC AMERICAN, December, 1985], describes an auditory-neuron direction-finding system that is remarkably similar in concept to an electronic system developed to measure the position of particle beams in the accelerators at the Fermi National Accelerator Laboratory.

In both systems, the common problem is to measure the relative power of two signals to within a fraction of a decibel when the absolute power levels can vary by 40 or 50 decibels. In the accelerator, passage of the beam between two small electrodes induces signals whose relative amplitudes are directly related to the position of the beam. These two in-phase signals are symmetrically split and recombined in quadrature (that is, with a 90-degree relative phase shift), so that the phase difference of the two resultant output signals varies by about 6.6 degrees per decibel of amplitude difference between the two input signals.

In the cricket, the relative external sound pressure on the two tympana is determined by the orientation of the cricket in relation to its chirping mate. The coupling of the two tympana by the tracheal tube apparently makes the conversion from amplitude to phase. The result is that the relative amplitudes of the two input signals are completely encoded in the relative phase of the two output signals. In the cricket, since there is little useful information in the amplitude of the nerve impulses from the auditory receptors, the omega neurons in the prothoracic ganglia must measure the relative phase of the two nerve signals. This is apparently accomplished by the reciprocal inhibition function in the omega-1 cells by a mechanism similar to that of two cross-coupled monostable multivibrators in an electronic system. The beam-position system differs from this only in that the phase-detection circuit generates an analogue signal proportional to the phase difference with an accuracy equivalent to about a tenth of a decibel in the ratio of the input signals.

The amplitude-to-phase conversion scheme was chosen to measure the beam position as it accomplished the necessary function with a minimal amount of signal processing. It is interesting that genetic selection in the cricket led to a similar result.

ROBERT E. SHAFER

Batavia, Ill.

To the Editors:

I should like to correct a possible wrong impression conveyed in the article entitled "The Search for Proton Decay," by J. M. LoSecco, Daniel Sinclair and me, which appeared in the June 1985 issue of *Scientific American*.

The first publication applying unified gauge-theoretical ideas to the problem of proton decay was by Jogesh C. Pati and Abdus Salam and appeared in *Physical Review Letters* in 1973. This paper marked the beginning of a development of a class of theories that led to the current widespread belief that protons do indeed decay. Following this early work, in 1974 Howard Georgi and Sheldon Lee Glashow discussed the so-called minimal SU(5) theory, putting forward a quantitative prediction of the proton lifetime (which has since been ruled out by the Irvine-Michigan-Brookhaven experiment described in our *Scientific American* article). As we mentioned in our article, long before the advent of a theory, evidence for proton decay was sought by experimentalists. This increasingly stringent series of experiments led William Kropp and me to propose (in mid-1975) a dedicated experiment designed to check proton stability more sensitively.

After we had several long discussions with Salam at the University of California at Irvine, during a visit he made for the express purpose of discussing the experimental status of proton decay, Pati and Salam spoke with officials of the U.S. Department of Energy, urging them to support our proposal. Not persuaded that the problem was worthy of further study (because the existing theory of Pati and Salam gave no firm prediction and the subsequent Georgi-Glashow theory was still numerically vague), they took no action for some time.

Meanwhile my colleagues at Irvine and Case Western Reserve University and I continued to ponder the results we had obtained during the previous decade in the deep South African mine and succeeded in pushing the limits somewhat further. Having held for some 25 years the general belief that conservation laws should be tested experimentally, it was gratifying to me that Pati and Salam had provided a base of a more specific kind and I was anxious to pursue the matter.

The increasingly attractive possibility that a unified gauge theory was compatible with—and even called for—an unstable proton under a wide range of assumptions was considered further by Georgi, Helen R. Quinn and Steven Weinberg, who spearheaded the continuing advance toward a testable quantitative prediction. This development, bolstered by the success of the gauge theory of electroweak interactions, led to a changed climate regarding expanded experimental tests, and the past seven years have seen a burst of activity worldwide.

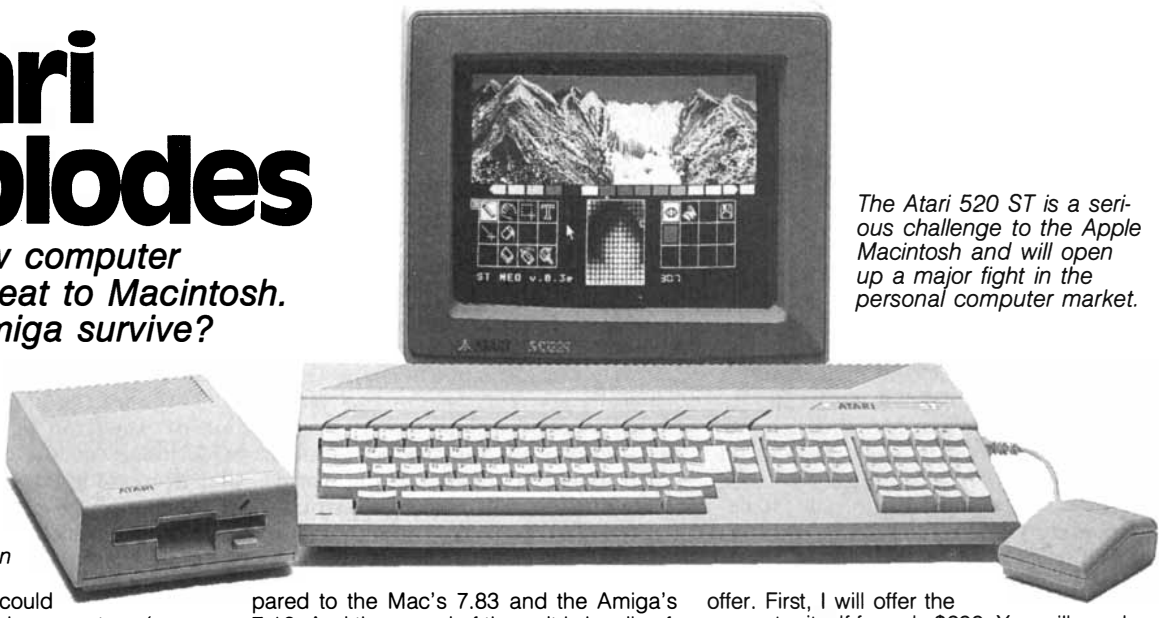
I hope these remarks will help to clarify the record regarding the seminal contributions of the theorists Pati and Salam to the increasingly central problem of proton decay.

FREDERICK REINES

Irvine, Calif.

Atari Explodes

Atari's new computer serious threat to Macintosh. Will the Amiga survive?



The Atari 520 ST is a serious challenge to the Apple Macintosh and will open up a major fight in the personal computer market.

By Joseph Sugarman

Imagine this. If I could offer you a Macintosh computer—(a computer that sells for over $2000)—for one third the price, you might wonder.

But what if I offered you a better computer with none of the disadvantages of the Mac and what if I added new features which improved its speed and performance? That's exactly what Atari has done in an effort to grab the ball from Apple and really explode into the personal computer market.

HEADING EFFORT

Heading the effort at Atari is Jack Tramiel—the same man who built Commodore into a billion dollar corporation, sold more computers than any other man in the world and believes in giving the consumer incredible value without sacrificing quality. The new Atari is a perfect example.

First, let's compare the new Atari ST to the Macintosh and the Commodore Amiga. Sorry IBM, we can't compare the ST to your PC because yours is almost five years old, much slower, and, in my judgement, over priced.

Price The cheapest you can get the Macintosh with 512K of memory is $1800 with a one-button mouse, a disk drive and a monochrome monitor. The Amiga sells for $1995 with a two-button mouse, a disk drive and a color monitor. The Atari ST sells for $699 with a two-button mouse, a disk drive and a monochrome monitor and for $200 more, a color monitor. Read on.

Monitor With the Mac you can only use its 9'' monochrome monitor and with the Amiga you can only use its 12'' color monitor. With the ST you have a choice of either a 12'' monochrome or high-resolution color monitor or your own TV set.

Resolution The number of pixels or tiny dots on a screen determine the sharpness of a computer monitor. The Mac has 175,104 pixels and has one of the sharpest screens in the industry. The Atari ST has 256,000 pixels or almost a third more than the Mac. And the Atari color monitor compared to the Amiga in its non interlace mode is 128,000 pixels or exactly the same.

Power All the computers have a 512K memory with a 68000 CPU operating with a 32-bit internal architecture. But Atari uses four advanced custom chips which cause the CPU to run faster and more efficiently giving it some tremendous advantages. For example, it has a faster clock speed of 8Mhz compared to the Mac's 7.83 and the Amiga's 7.16. And the speed of the unit is hardly affected by the memory requirements of the monitor which in the Amiga can eat up much as 70% of the unit's cycle time or speed.

Keyboard This is the part I love. The Mac has a small 59-key keyboard and a mouse. That's all. The 95-key Atari has both a mouse, cursor keys, a numeric keypad and ten function keys. The keyboard looks fantastic and is easy to type on. Although the 89-key Amiga has almost all the features of the Atari keyboard, it looks like a toy in comparison. (Sorry Commodore, but that's my opinion.)

Disk Drive The Mac's 3½'' disk drives run at variable speeds—slowing down as they run. The Atari 3½'' drives run faster at a constant speed—and quieter than any other unit.

Features The Atari ST comes equipped with the same printer and modem ports as the IBM PC—a parallel and RS232C serial port. The Mac comes only with a tiny non-standard serial and modem port. The ST has a hard disk interface capable of receiving 10 million bits per second. There are two joy stick ports and a 128K cartridge port for smaller programs or games. It has 512 colors (for the color monitor), it has a unique MIDI interface into which you can plug your music synthesizer and record or play back your music.

Software Right now, the Mac has more than the Atari ST and the Amiga combined. The Atari is a new system but the track record of Atari's Jack Tramiel and the potential of the new unit is causing a flood of new software titles. In fact, I'll predict that eventually the Atari will have more software than the Mac. There are now hundreds of titles, from word processing to spread sheet programs, from graphics and games to data base management—all with those easy drop-down menus and windows. There's plenty from which to select now and plenty more to come.

If you think I'm enthusiastic over the ST, listen to what the press is saying. *Byte Magazine* just called it the ''Computer of the year for 1986.'' *Creative Computing* exclaimed, ''Without question, the most advanced, most powerful micro computer your money can buy.'' and finally, the Atari ST is the best selling computer in Europe and acclaimed, ''The computer of the year,'' by the European personal computer press.

I am going to make the ST so easy to test in your home or office that it would be a shame if you did not take advantage of my offer. First, I will offer the computer itself for only $299. You will need, in addition, either one or two disk drives and either an Atari monochrome or color monitor or your own TV. If you order with your credit card during our introduction I will ship your order and only bill you for the postage and 1/3 the purchase price. I will also add a few software packages free including ''Logo''—a beginners programming language, a disk for programming in BASIC and Neochrome—a graphics paint program.

COMPARE THE TWO

After you receive the Atari ST, put it next to your Mac or Amiga or even IBM. See how extremely sharp the graphics appear, discover what a perfect word processor it is, how great the keyboard feels and finally how much faster and quieter it runs.

If you're not convinced that the Atari is far superior to your present computer and a fantastic value, simply return it and I'll refund your modest down payment plus our postage and handling charges. If you decide to keep it, I'll bill your credit card account for the remaining balance and enroll you in our discount software club (a $50 value) that lets you buy software for up to 50% off the retail price.

But act fast. We have only 2,000 units and 1,000 free memberships that we will offer as part of this introductory program and we are certain they will go fast. Order today.

To order, credit card holders call toll free and ask for product by number (shown in parentheses). Please add $20 per order for postage and handling. (If you pay by check, you must pay the full amount but we will provide you with a bonus software package.)

ST Keyboard, CPU & Mouse(4060Q) **$299**
Disk Drive (4056Q) **199**
Monochrome Monitor (4057Q) **199**
RGB Color Monitor (4058Q) **399**

Note: A list of software will come with the unit.

JS&A® PRODUCTS THAT THINK

One JS&A Plaza
Northbrook, Illinois 60062

CALL TOLL FREE 800 228-5000

IL residents add 7% sales tax. ©JS&A Group, Inc.,1985

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

FEBRUARY, 1936: "It is nearly a year since the great new star in Hercules appeared. When Mr. Prentice, the English amateur, first saw the star on the morning of December 13th, 1934, it was of the third magnitude. It decreased by half a magnitude and then brightened until, on the 23rd, it was of magnitude 1.5—equal to Alpha Cygni, and one of the most conspicuous stars in the sky. For three months after this it fluctuated irregularly in brightness, with a slow general downward trend, then on April 1 began a precipitate drop which took it from magnitude 4.5 to 10 in nine days. Early in May it began a steady rise in brightness, which brought it back to the 8th magnitude by the end of June. Since then it has varied but little."

"In the early 1880's aluminum was an expensive metal with few uses and a dubious outlook. Charles Martin Hall, 50 years ago this February, revolutionized the entire aluminum industry through the development of a process for electrolytically producing the metal from its oxide. Almost overnight the price of aluminum dropped, and 50 years have seen the creation of many diversified markets. Today aluminum ranks fifth in point of tonnage among the world's metals, being surpassed only by iron (and steel), copper, lead and zinc. Dreams of the '80's have come true: aluminum trains, bridges and buildings are actualities! What the next 50 years will hold for the metal is largely a matter of conjecture."

"It is commonly granted that the motion picture is important not only for its pervasive social effect but also because it is the only new art-form of modern times. Unfortunately films themselves are singularly evanescent. Generally speaking, a film two years old is a film that will not be seen again; the situation is comparable to that which would be created in the world of literature if only those books published within the past 12 months were available. In order to remedy this situation the Museum of Modern Art Film Library has been established for the purpose of collecting and preserving outstanding motion pictures of all types and of making them available to colleges and museums, thus to render possible for the first time a considered study of the film as art."

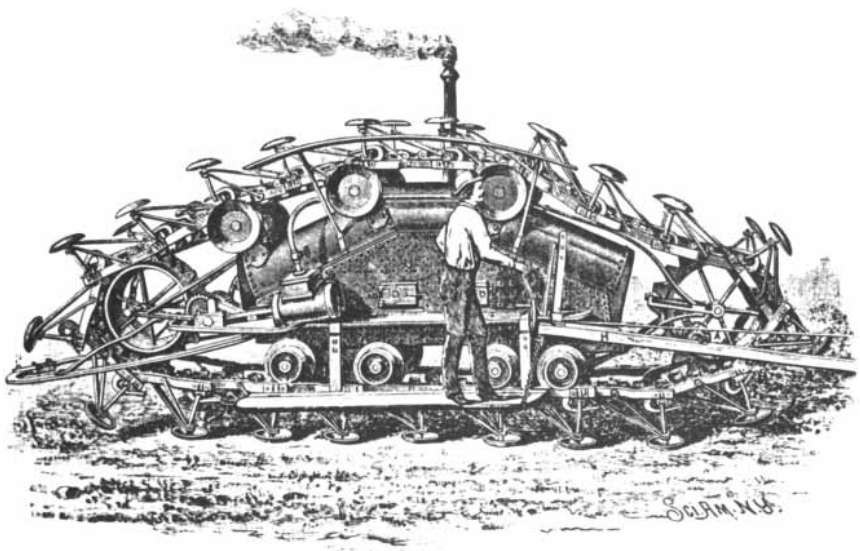
SCIENTIFIC AMERICAN

FEBRUARY, 1886: "The announcement made last summer that Mr. Thomas A. Edison was working out the details of a system of inductive telegraphy for sending and receiving messages on moving trains prepared the public for taking a lively interest in the practical trial of the system recently made on the Staten Island Railway. The main feature of the system is that of using the ordinary telegraph wires strung on the poles along the track, in place of a specially laid wire. The apparatus consists of an ordinary Morse key, phonetic receivers, an electromagnet, a vibrating reed, an induction coil and a battery. Similar apparatus is in use at the fixed stations. Of the system's usefulness in averting accidents, by keeping each train informed of the whereabouts of the one immediately ahead or following it, in intercepting criminals and in promoting general social and commercial intercourse, it is unnecessary to speak."

"The annually increasing destruction of American birds for purposes of fashion, and the consequent startling decrease in the numbers of many of the choice varieties, have aroused the American Ornithologists' Union to form a Committee on the Protection of North American Birds. The objects of the committee are the gathering of all possible information on the subject of the destruction of birds and the steps necessary for their preservation in the future; the diffusion of this information among the people in order to create a sentiment in favor of the birds, and in time the framing of a suitable statute, for enactment in the several States and Territories, which shall give the same protection to the smaller birds that the game laws afford the larger ones."

"A small apparatus that is placed in a central telephone office gives subscribers the exact time. This apparatus sends interrupted currents over the line every minute. The currents are so weak that they do not interfere with conversation, but they are sufficiently strong to produce quick, short, very distinct sounds that are separated by regular intervals. The subscriber, in order to know what time it is, has only to unhook his telephone and put it to his ear. Every minute he hears a feeble murmur that warns him to pay attention, and immediately afterward he hears successive interruptions that give the hour and minutes. The National Time Regulating Co., which is working this apparatus, charges subscribers one dollar a year."

"The novel farm locomotive herewith illustrated carries and lays its own track, which consists of an endless chain passing around two sprocket wheels, feet for resting upon the ground, and jointed side rail sections forming the track. One man can manage the locomotive and a gang of ten plows. A twenty horse power engine of this type plowed under test three acres six inches deep in one hour. Although the ground was wet and soft, the engine did not sink or slip and did the work well."



The track-laying farm locomotive

THE AUTHORS

FRED H. TSCHIRLEY ("Dioxin") is the executive director of the Michigan Agriculture/Business Council. He got his B.A. and M.A. at the University of Colorado in 1952 and 1954 respectively. He earned his doctorate in plant science from the University of Arizona in 1963. In 1954 he went to work for the Agricultural Research Service of the U.S. Department of Agriculture (U.S.D.A.). Between 1968 and 1974 he held a series of administrative posts in the U.S.D.A., the last one as coordinator of environmental-quality activities in the office of the Secretary of Agriculture. In 1974 Tschirley moved to Michigan State University, where he was chairman of the department of botany and plant pathology until 1979 and professor of botany and plant pathology until 1984; he took his current position last year.

PHILIPPA MARRACK and JOHN KAPPLER ("The *T* Cell and Its Receptor") are a wife-and-husband team at the National Jewish Center for Immunology and Respiratory Medicine in Denver. Both are professors of medicine at the University of Colorado Health Sciences Center, where Marrack is also professor of biochemistry, biophysics and genetics and Kappler is also professor of microbiology and immunology. Marrack had her undergraduate training at the University of Cambridge and got her Ph.D. there in 1970 for work done at the Medical Research Council's Laboratory of Molecular Biology. Kappler received his undergraduate degree at Lehigh University in 1965 and his Ph.D. from Brandeis University in 1970. The two investigators were postdoctoral fellows together at the University of California at San Diego before going to the University of Rochester Medical School in 1973 and to Colorado in 1979. Marrack traces her interest in immunology to an interest in the direct and indirect effects of genetics; Kappler came to immunology from studies of cellular differentiation.

VITALII I. GOLDANSKII ("Quantum Chemical Reactions in the Deep Cold") is a department chief at the Institute of Chemical Physics in Moscow. He was graduated in 1944 from Moscow State University, where he later received his Cand.Sc. in chemistry (in 1947) and his D.Sc. in physics and mathematics (in 1954). Goldanskii has written a number of books on such topics as applications of the Mössbauer effect, the general principles of gamma lasers, chemical applications of positron annihilation and the properties of nuclei far beyond the limits of beta-stability.

JOHN C. MUTTER ("Seismic Images of Plate Boundaries"), a marine seismologist, is adjunct professor of geology at Columbia University and associate research scientist at Columbia's Lamont-Doherty Geological Observatory. Born in Australia, he got a B.Sc. in physics and pure mathematics at the University of Melbourne and an M.Sc. at the University of Sydney. His association with Lamont-Doherty began during a collaborative experiment in which he took part as an employee of the Australian Bureau of Mineral Resources. Further joint studies led to and incorporated the work Mutter did to earn a Ph.D. from Columbia.

MARC CANTIN and JACQUES GENEST ("The Heart as an Endocrine Gland") work together at the Clinical Research Institute of Montreal, where Cantin is director of the laboratory of pathobiology and of a multidisciplinary research group on hypertension and where Genest is a consultant. They are both professors in the department of experimental medicine at McGill University, and Cantin is professor of pathology and Genest is professor of medicine at the University of Montreal. Cantin has a B.A. and an M.D. from Laval University, earned in 1953 and 1958, and a Ph.D. from the University of Montreal, earned in 1962. From 1962 to 1965 he was at the University of Chicago as a postdoctoral student and as instructor in the department of pathology. He joined the staff of the University of Montreal in 1965. Cantin's association with the Clinical Research Institute dates from 1980; he went to McGill in 1983. Genest's B.A. (1937) and his M.D. (1942) are from the University of Montreal. He did postgraduate work at the Hôtel-Dieu Hospital in Montreal, at Johns Hopkins, at Harvard University and at the Rockefeller Institute. He has practiced at Hôtel-Dieu since 1952. Genest's association with the University of Montreal goes back to 1965 and with McGill to 1970.

CHARLES G. SIBLEY and JON E. AHLQUIST ("Reconstructing Bird Phylogeny by Comparing DNA's") are respectively professor of biology and an associate research scientist and lecturer in biology at Yale University; they are also both affiliated with Yale's Peabody Museum of Natural History, where Sibley is curator of birds and Ahlquist is a research associate in ornithology. Sibley earned a Ph.D. in zoology from the University of California at Berkeley in 1948, and he held positions at San Jose State College and at Cornell University before going to Yale in 1965. He first began applying biochemical techniques for manipulating proteins and DNA to the study of the systematics and phylogeny of birds in 1956. From 1970 to 1976 Sibley was director of the Peabody Museum. Ahlquist has a B.S. in biology from Cornell and M.S. and Ph.D. degrees from Yale. From 1972 to 1977 he was a curatorial associate at the Peabody Museum. Ahlquist took his present position with the museum in 1977 and joined the faculty of Yale's biology department in 1981.

JEAN-PIERRE PROTZEN ("Inca Stonemasonry") is chairman of the department of architecture at the University of California at Berkeley. In 1954 he was graduated from the Collège St.-Michel in Fribourg, Switzerland. He went on to study architecture at the Swiss Federal Institute of Technology and then at the University of Lausanne, where he received his diploma in 1962. He was a practicing architect, specializing in the industrialized production of housing units, from 1956 to 1967. Then a grant from the Swiss National Foundation for Scientific Research allowed him to move to Berkeley to investigate the problem-solving behavior and decision-making procedures of designers. Since 1968, when he took a position at Berkeley, Protzen has concentrated on the methods and theories of design.

MICHAEL HOSKIN ("William Herschel and the Making of Modern Astronomy") is head of the department of history and philosophy of science at the University of Cambridge and president of Churchill College at Cambridge. He studied mathematics at the University of London, getting a B.A. in 1951 and an M.A. in 1952. His Ph.D. in algebraic geometry from Cambridge was awarded in 1956. He then changed his field of interest to the history of science, in which he was a lecturer at the University of Leicester from 1957 to 1959, when he joined the faculty at Cambridge. Hoskin's special interest is the history of stellar (as opposed to planetary) astronomy.

LEON J. KAMIN, who reviews *Crime and Human Nature,* by James Q. Wilson and Richard J. Herrnstein, is professor of psychology at Princeton University.

BUICK

The performance car that's tough as a mule.

The Buick Electra T Type is a performance car born of an extraordinary development program that took engineering prototypes — test "mules" — to the far corners of the world. We pushed it to the limit in the Australian Outback, the Bavarian Alps, in Kapuskasing, Canada. It was the only way to ensure that under the civilized skin there was bone and sinew and muscle enough to set a world standard.

But now it is time to see if it can meet the most demanding standard of all: yours. Fire up the 3.8-litre V-6 with sequential-port fuel injection and electronic ignition. Test the cornering of the special, firm suspension, tuned for serious driving and fitted with low-profile, steel-belted Eagle GT radials on wide 15" x 6" alloy wheels.

Grasp the leather-covered sport steering wheel, feel its quick-ratio steering and know that this is a performance car with the soul of a Buick and the heart of a mule.

Buckle up and visit your Buick dealer to put it to your own test. He can also tell you about a GMAC leasing program to make glorious driving more affordable than you thought.

Wouldn't you really rather have a Buick?

GM Buick Electra and the mule. For more information or a test drive, call 1-800-86-BUICK (1-800-862-8425).

YOU'VE ALWAYS HAD A LOT OF COMPETITION. NOW YOU CAN HAVE AN UNFAIR ADVANTAGE.

Nobody ever said it was going to be easy. But it just got easier. Now there's Amiga.™ The first and only computer to give you a creative edge. Amiga makes you look better, sound better, work faster and more productively.

You can't buy a computer at any price that has all of Amiga's features. Nor can you find one that's easier to use. Amiga lets you point at symbols instead of learning complicated commands.

Amiga is friendly, but it's a powerhouse, too. It has twice the memory of Macintosh™ or IBM® PC. It costs less than either of them and can do everything they can do, better.

No other personal computer gives you over 4,000 colors, stereo sound and incredible dimension. Imagine the advantage of preparing business presentations with color graphics and sophisticated animation right on your computer.

Need to make creative use of your time? Amiga can do as many as four or five things at once in separate windows on the screen. Not just display them. Work on them. No other personal computer can.

Amiga is IBM-compatible, too. A simple piece of software teaches Amiga to emulate the IBM operating system, so you can run most IBM programs. You'll have instant access to the largest library of business software in the world, including favorites like Lotus® 1,2,3, and dBase®.

And Amiga is endlessly expandable and adaptable. You can plug in printers (almost any kind), modems, musical keyboards, extra disk drives. You can even expand the memory to a whopping 8 megabytes with an optional expansion module.

See an Authorized Amiga Dealer near you. And don't wait. Your competition is gaining on you. Is that fair?

Amiga by Commodore.

Amiga's 4,096 colors give your business graphics a visible advantage.

Amiga makes telecommunications fast, easy and colorful.

Amiga's 4 channels of stereo give you a sound advantage.

AMIGA GIVES YOU A CREATIVE EDGE.

COMPUTER RECREATIONS

The king (a chess program) is dead, long live the king (a chess machine)

by A. K. Dewdney

If CRAY BLITZ had a memory for anything except the moves of chess, it would never forget the evening of October 15, 1985. It is the last round of the North American Computer Chess Championship, held at the annual meeting of the Association for Computing Machinery. A space at the front of a meeting room in the Radisson Hotel Denver is taken up by five tables, which are separated from the audience by a barrier. At each table two teams of programmers and advisers face each other. Sometimes they joke and sometimes their faces fall into the blankness of wondering and waiting. Behind each table is a display screen on which an overhead projector casts the image of a current board.

The tournament features 10 contenders for the North American title. Their names are odd and angular, betraying differing origins and aspirations: AWIT, BEBE, CHAOS, CRAY BLITZ, HITECH, INTELLIGENT SOFTWARE, LACHEX, OSTRICH, PHOENIX and SPOC [*see illustration on next page*]. Missing are three of the big names that have dominated computer chess in recent years: BELLE, CHESS 4.7 and NUCHESS.

Most of the interest is focused on the championship game between CRAY BLITZ and HITECH. On the CRAY BLITZ side of the table are Robert Hyatt of the University of Southern Mississippi, Albert Gower, a chess expert from the same institution, and Harry Nelson of the Lawrence Livermore National Laboratory. Facing them are Hans Berliner of Carnegie-Mellon University and Murray Campbell, one of his graduate students, who is an expert player. Berliner fills the dual role of chess adviser and programmer for the HITECH team. As the game wears on and tension mounts, Berliner rises often from the table, a weary smile on his face. Once he strolls past my chair and mumbles, "This is too much like my U.S. championship days!" (For several years in the late 1950's and early 1960's Berliner was rated among the top dozen players in the U.S.)

Unlike the U.S. championship, in which a deathly silence reigns, this tournament is filled with conversation, occasional laughter, the rattle of keyboards and a continuing microphone commentary by adjudicator Michael Valvo, a flamboyant computer consultant and international chess master from Sedona, Ariz. "A weak move by Black. The king is still too exposed and the doubled pawns on c5 and c6 continue to hamper the defense." Nearby a member of the CRAY BLITZ team exclaims to no one in particular, "That's funny, I thought it would play king to f3." An international master can still spot flaws in computer chess and programs still surprise their creators.

Throughout this final round of play it has been obvious that HITECH has the advantage over its rival: early in the game CRAY BLITZ has fallen into a zugzwang, a critical position from which any conceivable escape involves either a bad move or a loss of material. In this case CRAY BLITZ has been forced to structure its pawns badly. HITECH continues to exploit the advantage.

By midnight it is all but over. Most of the games are finished and the experts claim a win for HITECH. The CRAY BLITZ team asks adjudicator Valvo for permission to resign. He suggests two more moves: if the CRAY BLITZ position is no better then, the team may resign. It is not and they do. HITECH is North American champion and de facto king of computer chess. Although CRAY BLITZ is the official world champion (it won the title in 1983 and does not have to defend until June), HITECH's win, along with its three other tournament victories, is impressive. HITECH is almost certainly the world's strongest chess-playing computer.

There are smiles and more conversation. Did the absence of BELLE, CHESS 4.7 and NUCHESS make a difference? "It would have been nice if BELLE and some of the other programs could have made it," said one organizer of the tournament, "but I don't think the outcome would have been much different." He went on to point out that in terms of the programs and machines entered, there was no effective difference between the North American and the world championships. The talk turns to Kasparov and Karpov and then to theory. "I'm not kidding," says an apparently knowledgeable participant. "A 20-ply program that looks only at material can beat any grand master." There is some argument, but in a few more minutes the room is empty. The North American championship is over.

The claim about the 20-ply program is an interesting one. The game of chess can be represented by a vast tree consisting of nodes and lines. I visualize the tree upside down, so that the root node is at the top. Each node represents a possible position, namely a chessboard on which the pieces and pawns have arrived at their squares through legal play. A node is joined to a descendant node by a line if the move of a single piece or pawn converts the board represented by the former node into the board represented by the latter node. A game of chess can always be identified with a particular path through the chess tree from the root node (in which no moves have been made) down through the tree to some node where, as a general rule, few pieces are left and one player has been checkmated or forced to resign.

A chess-playing program attempts to explore only as much of the game tree as is necessary. From the node representing a current position it examines all the descendant boards (ply 1), examines the descendants of the descendants (ply 2) and so on. The average depth of its exploration is called the lookahead. This measure comprises the greater part of what might be called a chess program's intelligence. The lesser part arises in the program's evaluation of the boards constituting the horizon of its lookahead. It analyzes these boards and attaches a numerical value to each one. The value reflects the desirability of reaching that position. Using a procedure called minimax, the program causes some of the values thus assigned to percolate up the tree to the nodes at ply 1. The node receiving the highest value is the play to make.

There is an interesting tradeoff be-

tween the two parts of the program's intellection: the better its evaluation scheme is, the less deeply it needs to search the game tree. Indeed, if it had a perfect evaluation scheme, it would never have to search deeper than one ply. Conversely, a program with a very simple evaluation scheme must search much deeper if it is to play effectively. How deep must a search of the kind that looks only at material be in order to be effective against a grand master? Would a 20-ply search suffice?

The title of grand master is awarded by the Fédération Internationale des Échecs to players who distinguish themselves in international play. (The federation bars computers from consideration.) Grand masters generally have ratings higher than 2,400, the level of a senior master. Up to the time of the North American Computer Chess Championship, HITECH had played 21 games in human tournaments, earning a rating of 2,233. This made it the highest-rated chess-playing computer in the world. According to Berliner, who was rated at 2,443 in his competition days, HITECH's rating has increased by an average of eight points a game in national tournaments. Dare one suppose that in just 14 more games the machine will surpass its designer?

All of this raises the question of just how good chess-playing computers will eventually become. Will a computer ever be the best chess player in the world? David Levy, former player and present author-entrepreneur, has committed the question to a series of wagers. In 1968 Levy bet John McCarthy of Stanford University £500 that no computer would succeed in beating him in a chess match for the next 10 years. Levy collected in August, 1978, at the Canadian National Exhibition in Toronto. There he toyed with CHESS 4.5, a program created at Northwestern University. The basic bet was thereafter renewed in the amount of $6,000 for a period of six more years. In April, 1984, Levy in London played a telephone match with CRAY BLITZ. He won again.

Levy's short streak emboldened him to offer the following £100,000 wager in Denver: within 10 years of the offer each computer challenger will have been defeated by a human player selected by Levy. If Levy finds a taker, it will probably be not a mere program but a specialized computer. So far there have been no takers.

The two top finishers in the North American tournament, HITECH and BEBE, were essentially such chess machines. Interestingly, Levy's own entry, a program named INTELLIGENT SOFTWARE, came in third. It runs on an Apple IIe computer that features nothing more sophisticated than an accelerator board, which is a special circuit card that doubles the speed of the machine. Perhaps Levy has developed a superior evaluation scheme.

The chess cognoscenti at the championship agree that the best game of the tournament was played during round two between CRAY BLITZ and BEBE, a product of private enterprise. Tony Scherzer, whose company SYS-10, Inc., developed BEBE, has transported his charge to a number of tournaments in recent years. BEBE is no mere program but a chess-playing machine in its own right. The game was significant not only because it was the most interesting of the tournament but also because it was the first time CRAY BLITZ had lost in three years.

Readers with a chessboard can follow the CRAY BLITZ v. BEBE game by playing the 50 moves listed below. Pieces are denoted by capital letters: K, king; Q, queen; B, bishop; N, knight, and R, rook. Chessboard squares are referred to by letter-number coordinates: when the board is in the standard position, so that the lower left-hand square is black, the files, or columns, are labeled from left to right with *a* through *h;* the ranks, or rows, are numbered 1 through 8, beginning at the bottom of the board. Notation employed in listed games such as the one below varies from the straightforward Kb1 (king to square b1) to the puzzling Nf3 (knight to f3). Which knight? On that particular move only one knight can jump to f3. A move by a pawn is indicated by the designation of a square, for example e4. The game is annotated by Valvo.

	CRAY BLITZ (White)	BEBE (Black)
1.	e4	c5
2.	Nf3	d6
3.	d4	cxd4

(The x means a piece or pawn is taken.)

4.	Nxd4	Nf6
5.	Nc3	g6
6.	Bg5	Bg7
7.	Qd2	Nc6
8.	0-0-0	0-0

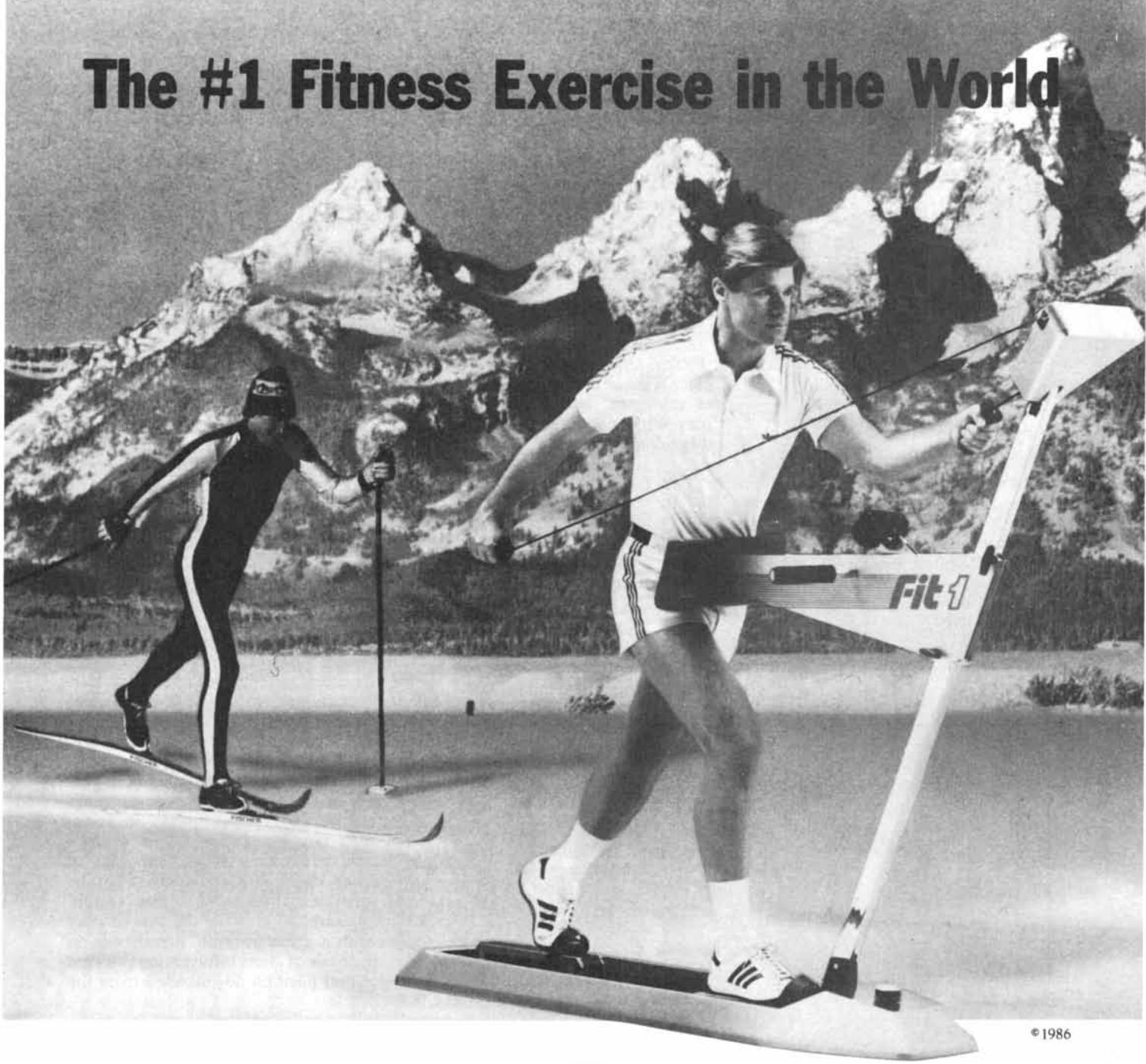
(White castles on the queen side and Black castles on the king side.)

9.	Nb3	Re8
10.	Bc4	Ng4

Black played Ng4 intending 11... Bc3xN in the next move. Black may have thought that 12 c3xB is forced on White, but Black changes its mind on White's next play. If Bxc3, then Qxc3!; Nxf2 fails if either rook is

PROGRAM	ORIGIN	COMPUTER	LANGUAGE	BOARDS/ SECOND	LOOK-AHEAD
AWIT	University of Alberta	Amdahl 5860	Algol W	10	3-ply
BEBE (second)	SYS-10, Inc., Hoffman Estates, Ill.	Custom machine	Assembler	20,000	7-ply
CHAOS	University of Michigan	Amdahl 5860	FORTRAN	70	4-ply
CRAY BLITZ	University of Southern Mississippi	Cray X-MP 48	FORTRAN/ Assembler	100,000	8-ply
HITECH (first)	Carnegie-Mellon University	Sun with custom VLSI	C	175,000	8-ply
INTELLIGENT SOFTWARE (third)	Intelligent Software, Inc., London	Apple IIe with accelerator	Assembler	500	7-ply
LACHEX	Los Alamos National Laboratory	Cray X-MP 48	FORTRAN/ Assembler	50,000	7-ply
OSTRICH	McGill University	Network of seven Novas and an Eclipse	Assembler	1,200	6-ply
PHOENIX	University of Alberta	Network of VAX 780's and 10 Suns	C	540	6-ply
SPOC	SDI/Cypress Software, San Jose, Calif.	IBM PC	Assembler	300	5-ply

Entrants in the 1985 North American Computer Chess Championship



...can now be duplicated at home!

You deserve the best. Now you can have it.

Cross country skiing is rated the number one cardiovascular and muscle toning exercise in the world. It's better than running. Better than exercise cycles, or rowing machines. Even better than swimming!

The unique design of ***Fit 1*** duplicates cross country skiing. It firms and tones every major muscle group in your body. And it gives you the best cardiovascular workout available.

Fit 1 features variable resistance settings for arm and leg movements. It can challenge you no matter what your level of fitness. Duplicate the resistance of a casual glide across a flat field . . . or a tough trek through the Rockies.

The electronic console provides immediate feedback on your heart rate, speed, distance traveled and elasped time. Data that guarantees you're getting the maximum cardiovascular workout in the minimum amount of time.

Fit 1 can get you in the best shape of your life. All on one simple machine.

To learn more on how you can enjoy the number one fitness exercise in your home, call anytime for our free brochure.

1-800-453-1212

Ask for operator 41

Fit 1™

for those who demand the best

1400 Raff Rd. SW • Canton, Ohio 44750

played to f1, whereas Bxf7, a check, would be fatal.

11.	h3	Nge5
12.	Bb5	a6
13.	Be2	a5
14.	Bb5	Be6
15.	Nd5	a4
16.	Nd4	Bd7

White's situation is desperate. The Black pawn (a4) threatens to create weaknesses around White's king.

17.	Nxc6	bxc6
18.	Nxe7 check	Rxe7
19.	Bxe7	Qxe7
20.	Be2	Qe6
21.	Kb1	Rb8

(The chess board at this point in the game is shown below.) Black threatens 22...Rb2, a check, which is followed by 23 Kb2 Nc4, a fork that wins the White queen.

22.	b3	axb3
23.	cxb3	Be8
24.	Kc2	Nd7
25.	f3	Ra8
26.	Kc1	Nc5
27.	Qc2	Qf6
28.	Bc4	Qa1 check
29.	Kd2	Qxa2

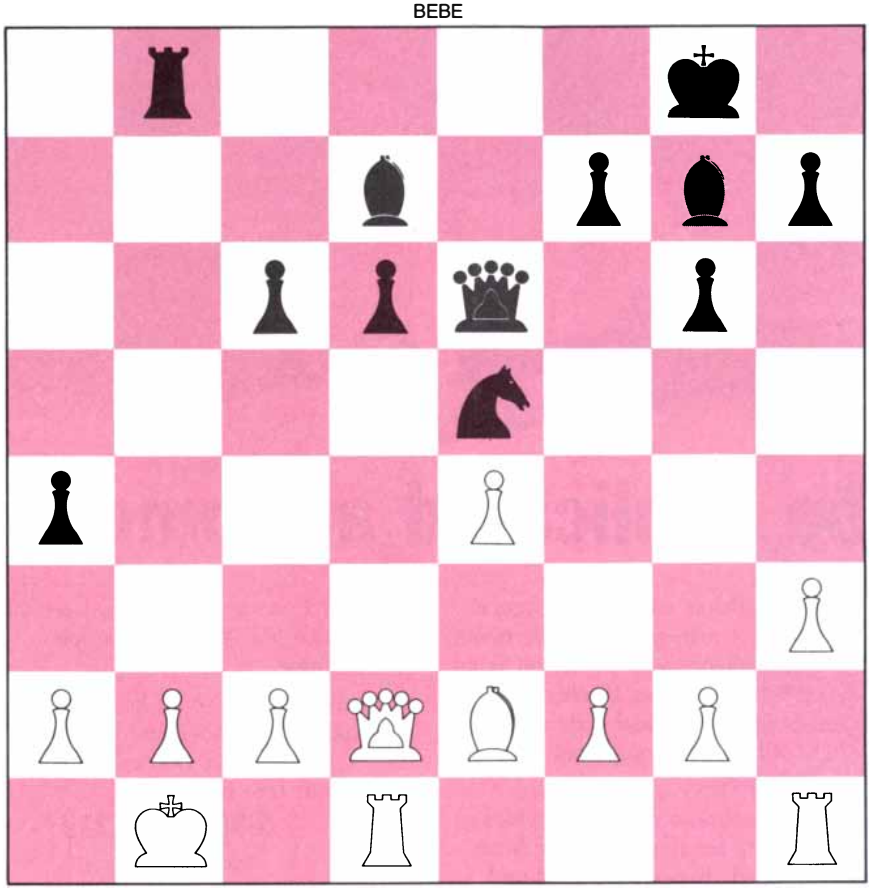
An even stronger move is 29...Bc3, a check, followed by 30 Ke2 Ra2!

30.	Qxa2	Rxa2 check
31.	Kc1	d5
32.	exd5	cxd5
33.	Bxd5	Bb5
34.	Rhe1	Nd3 check

Black's material advantage of one piece is about to be increased by another exchange. In human tournament play White could reasonably resign at this point.

35.	Rxd3	Bxd3
36.	Re8 check	Bf8
37.	g4	Kg7
38.	Re3	Ba3 check
39.	Kd1	Ra1 check
40.	Kd2	Bf1
41.	Kc3	Rc1 check
42.	Kd2	Rc5
43.	Ke1	Bxh3
44.	Bc4	h5
45.	gxh5	gxh5
46.	Kf2	h4
47.	Rd3	Bf5
48.	Rd4	h3
49.	Rh4	Rc7
50.	Rh5 (asks to resign)	

BEBE



CRAY BLITZ

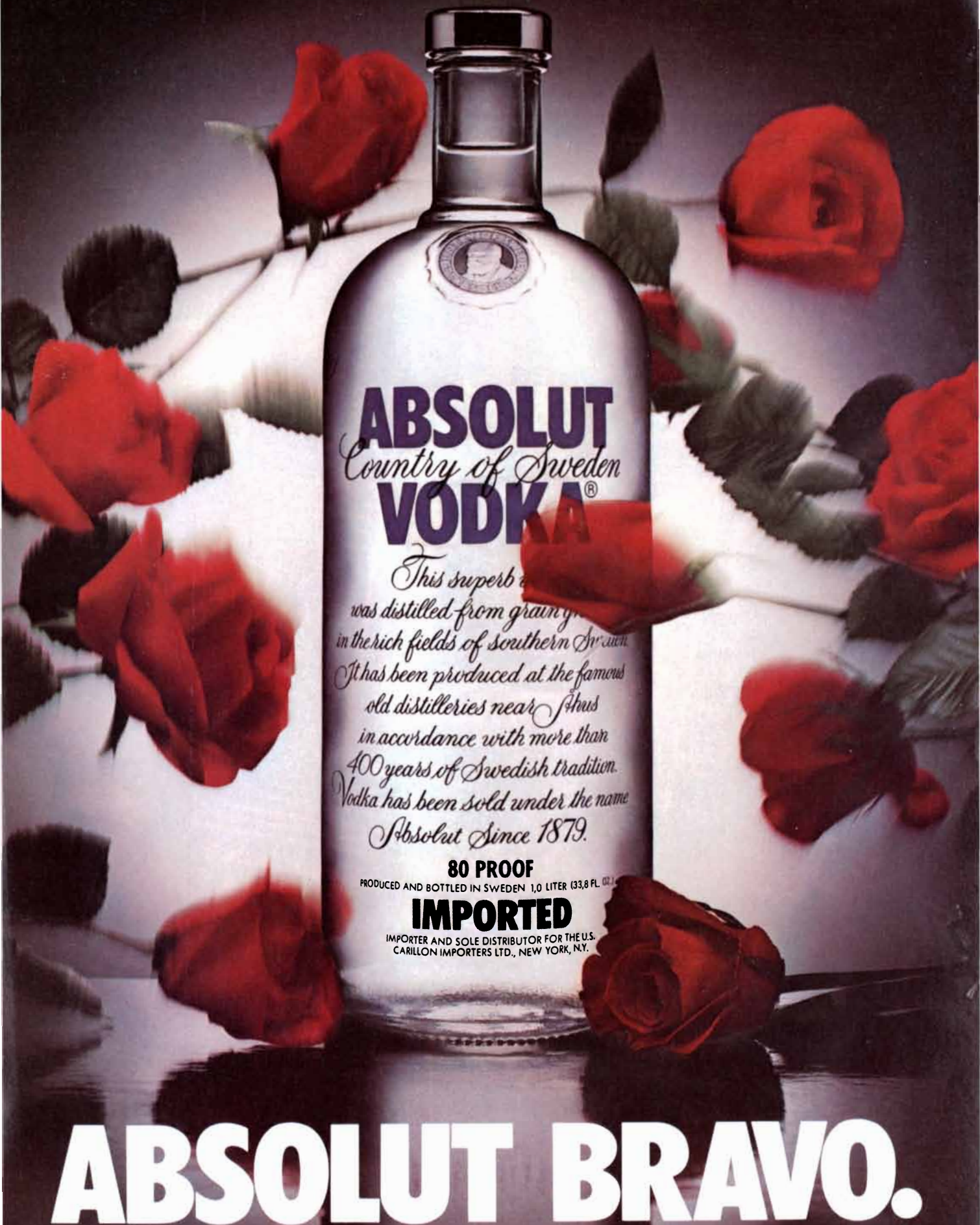
The board after move 21

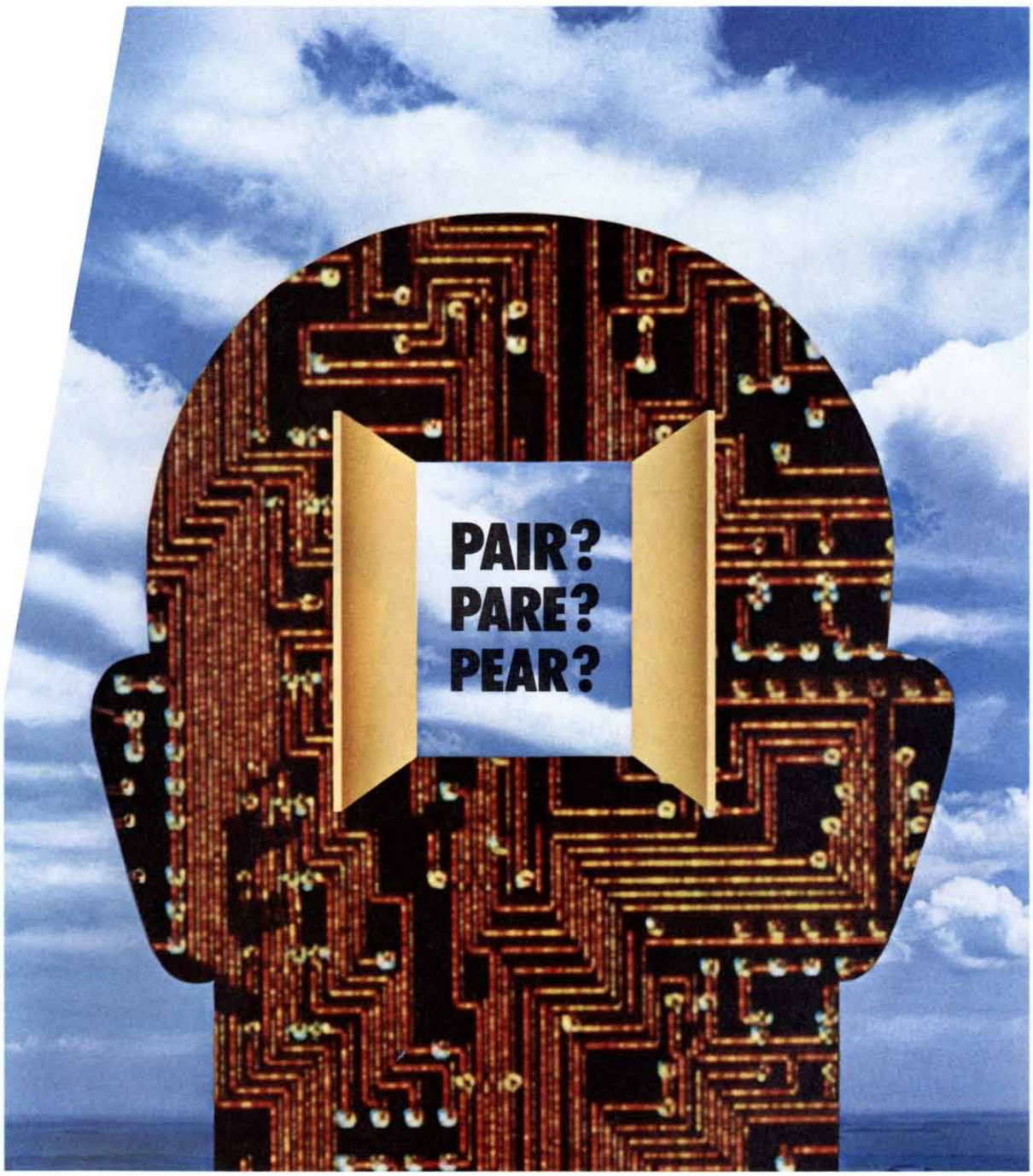
The CRAY BLITZ program runs on a Cray XM-P 48 computer. Famed for its speed as a multiprocessor, the Cray is nonetheless a general-purpose computer and not a chess machine. BEBE, whose circuits are devoted to chess playing, obviously outperformed the Cray–CRAY BLITZ combination in the game above.

HITECH is in a sense even more specialized. When Carnegie-Mellon University was the Carnegie Institute of Technology, a chess-playing program called TECH was developed there. The name HITECH reflects the fact that Berliner, Campbell and the other members of the HITECH team, Carl Ebeling, Gordon Goetsch, Andy Palay and Larry Slomer, have revived the TECH tradition in a world of very-large-scale integration (VLSI) and burgeoning parallelism. The HITECH machine combines a Sun computer equipped with a specially designed processor that Berliner calls the searcher. The Sun computer runs three programs: a user interface, a task controller and an oracle. The oracle embodies what computer chess experts call the book. This is a large catalogue of chess openings and variations that human chess experts commonly know. The oracle's data base contains a great deal of other chess knowledge that can be easily expanded. When the searcher examines the possibilities of play from a given position, it proceeds on the basis of chess information relevant to that position downloaded from the oracle.

The searcher itself contains a microprocessor and several hardware modules that generate moves, evaluate moves, check for repeated moves and so on. The microprocessor coordinates their activities. The move generator consists of 64 VLSI chips, one for each square of the chessboard. Each chip examines the entire board in order to determine whether any piece or pawn can be moved to the square under its purview. It determines the best move in terms of standard criteria such as opportunities for capture or control of the center. At the same moment the other 63 chips are doing the same thing. If there are 10 pieces on the board, this architecture means that possible moves are generated 10 times faster, other factors being equal.

The evaluation of moves must keep





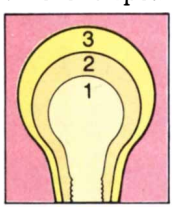
Would you like to know how computers can tell the difference between a pair in a poker game and a pear on a plate?

How they can understand a variety of speakers with a diverse variety of accents—and reply in pear-shaped tones, using normally connected speech?

Then read on to learn more about computers that recognize words, comprehend meaning from context, even synthesize human speech from a mere shadow of itself.

It's All In The Algorithms

Utilizing three levels of speech-processing algorithms, AT&T is giving the computer a more 'robust' understanding—the capacity to comprehend connected speech from different speakers.



Three levels to understanding

Acoustic pattern matching (1) identifies the spoken words.

Grammatical processing (2) figures out how the words are put together.

And semantic processing (3) extracts meaning from the context. With each successive step, the computer moves closer to accurate understanding.

Acoustic pattern matching determines how much latitude the waveform (pronunciation) of a word can have before it becomes unintelligible to the computer.

By isolating the specific characteristics the waveform of a word contains—independent of the accent of a speaker—we increase the probability that it will be correctly matched to a pattern stored in a computer's memory. But, correct recognition of words is only the beginning of computer understanding.

Computer Grammar 101

Grammatical processing further increases the probability of recognizing words. It analyzes them within the constraints imposed by language—the allowable sequences of syllables in a word or words in a sentence.

For a specific vocabulary and situation, it is possible to define every sequence the computer can recognize. Based on probabilities assigned to each word it recognizes—and where that word falls—the computer determines which of its possible sequences is the most likely. This process gains two advantages: It allows words that might not otherwise be recognized to be correctly accepted; and it speeds up processing time by using sequence position to limit the number of words it looks at for a pattern match.

A Meaningful Relationship

Semantic processing is the point where the computer crosses the line between recognition and understanding—the point where words are given meaning within a specific context. This endows a system with one of its most human qualities: knowing when a request isn't understood, and asking for appropriate clarification.

Talk Isn't Cheap

Making a computer listen intelligently is one thing; making it respond intelligibly, however, is another.

Enabling a computer to talk, reproducing the subtleties of human speech, has required large amounts of memory—a high cost item. Therefore, an 85 percent reduction in the amount of information needed to store and generate high-quality speech can mean significant cost reductions.

That's just what a new AT&T speech synthesis technique, called multi-pulse linear predictive coding (MP-LPC), provides. It reduces the 64 thousand bits per second previously needed to 96 hundred.

Speech signals mimic the human vocal tract—they have redundancies built in. MP-LPC codes speech to remove these redundancies, then tells the computer how to reconstitute the original speech from the mini-version in its memory. This coding eliminates unnecessary bits from being stored and transmitted.

Getting Down To Business

At AT&T, our goal is to make computers listen and understand as fast as people speak—and speak to and understand as many people as possible. Speech-processing algorithms, developed by AT&T Bell Laboratories, have moved us several steps closer to that ideal.

For example, most speech recognition systems make the speaker pause between words. But AT&T, using advanced recognition algorithms, has developed a Stock Quotation System that allows callers to enter and retrieve current market information in natural, normally-connected speech. Users simply speak the number codes for any of over 6,000 stocks, and the service provides current quotes—delivered in computer-generated speech. The system is now in initial service for customers of a major brokerage company.

Numbers are nice, but make for limited conversation. Closer to our goal of a conversational computer is the Flight Information System. It uses the Official Airline Guide as its data base. In its limited environment, this laboratory system converses with the user in natural speech in response to normal flight information queries.

One Of Our First Callings

AT&T has been deeply involved in speech technology since the genesis of the telephone. From the beginning, our goal was to make mechanical communications fast, foolproof and economical.

Today, with the advent of the computer, we're moving toward the ultimate ideal: creating machines that serve our needs and save our energy in the most natural manner—by voice command.

AT&T

The right choice.

The Audi 5000CS Turbo Quattro.
5,4,3,2,1…Prepare yourself for an unprecedented driving experience. For the Turbo's intercooled 2.2 liter engine achieves 0 to 50 in a mere 5.7 seconds.

Impressed? Now consider that the Turbo Quattro combines the unique Anti-Lock Braking System with Audi's acclaimed permanent all-wheel drive.

All-wheel drive maximizes traction by conveying engine power equally to all four wheels. All the time. While the ABS system virtually eliminates wheel lock-up by utilizing advanced computer technology. Under almost all conditions, superb stopping power is assured.

Aside from being the only production luxury sedan in the world with a fully galvanized steel body, the Quattro has aesthetic appeal as well. In fact, its sleek profile bears more resemblance to classic sculpture than an automobile.

The Audi 5000CS Turbo Quattro. Take one for a test flight today.

The art of engineering.

The only way to fly.

The Audi 5000CS Turbo Quattro is priced at $27,975* and is protected by a 3-year, 50,000 mile new vehicle limited warranty and a 6-year limited warranty against corrosion perforation. (*M.S.R.P. Title, taxes, transp., reg. and dealer delivery add'l.) Call 1-800-FOR-AUDI for details.

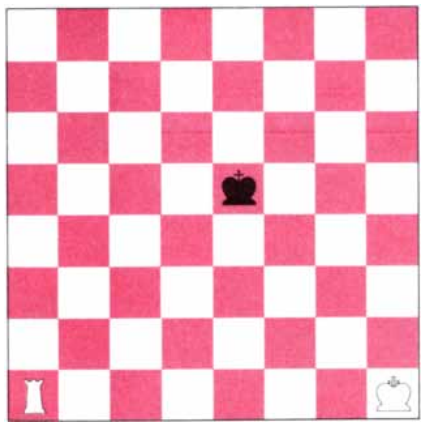
up with the generation process. A first stage of evaluation is carried out by the move generator itself. It houses a kind of supervisor that judges among the moves generated by the 64 chips. Each chip computes a number that estimates the strength of its best move and transmits the number to the supervisor. The chip-generated numbers are like cries for attention. The supervisor ranks them in order of loudness (read effectiveness).

HITECH then proceeds to search the game tree, following the ranking produced by the supervisor of the moves that are possible from the current position. A second stage of evaluation is performed by the evaluation module at each new position generated within the game tree. Using chess knowledge (downloaded from the oracle) relevant to the current position, the module evaluates each board, whether or not it is at the lookahead horizon. Such is the nature of parallelism. The additional effort costs no additional time. The Sun's task controller tells the searcher how deeply to search the tree and, when the search is completed, whether to probe still further. In this way HITECH manages an average lookahead of eight ply but will, on occasion, search as deep as 14 ply. This may seem a long way from the 20 ply that may be needed to beat a grand master. On the other hand, HITECH's use of parallelism and sophisticated deployment of chess knowledge into the tree search may compensate for the relative shallowness of the search. In any event, by the time of the next World Computer Chess Championship in June in Cologne, West Germany, HITECH may be unstoppable.

HITECH is the world's newest chess machine. The first such machine was invented in 1890 by Leonardo Torres y Quevedo, a Spanish engineer. Using mechanical levers, pulleys and electromechanical switches, it played a mean game of rook-and-king v. king. Human players were given the privilege of managing the affairs of a lone king seeking to evade checkmate by the machine's powerful combination of a rook and a king. Torres y Quevedo's machine always won.

Readers are invited to create a strategy that produces this result. It should be assumed that the human player's king does not begin life in a stalemate (unable to move without placing itself in check). Then the task is to specify in as few rules as possible how the machine achieves checkmate from an arbitrary position. The position shown in the illustration on this page is a reasonable starting point.

The machine plays White and is to move first. How do White's king and rook combine to drive Black's king into checkmate? White could begin by moving its rook to the *d* file. This would prevent Black from moving its king any farther to the left. The maneuver can be repeated if the Black king obligingly moves to the right, but what if it continues to occupy the *e* file, merely cruising up and down? I shall publish the most succinct solution, whether it is in algorithmic or in natural language.



How to checkmate with rook and king?

Last November's "Computer Recreations" featured flibs: finite living blobs that attempt to predict changes in their environment. In the primordial computer soup, during each generation the best predictor crosses chromosomes with a randomly selected flib. Increasingly accurate predictors evolve until a perfect one emerges.

A flib is essentially a finite automaton. That is, it has a finite number of states, and for each signal it receives (a 0 or a 1) it sends a signal and enters a new state. The signal sent by a flib during each cycle of operation is its prediction of the next signal to be received from the environment.

Some readers gave their flibs impossible prediction tasks. No flib will ever evolve that can predict a sequence of random bits. Nor will flibs ever develop to predict primes. It is perfectly reasonable to ask a flib to predict a repeating binary sequence. For example, there is a 4-state flib that will predict the repeating eight-symbol sequence 01100010. Even a repeating sequence, however, can tax the predictive abilities of a flib if its basic string is too long in relation to the number of states in the flib. As it happens, no 4-state flib will ever predict the repeating sequence 010010111. Why not?

The simplest answer to the question involves a process I call creeping induction. Imagine a 1-state flib. It might predict the endless repetition of the basic string 01. For each of the two possible signals the flib receives there is one response: if a 0 is received, the flib sends a 1 and then reenters the same state. If it receives a 1, it sends a 0. A basic string of three symbols, say 011, is beyond the ability of a 1-state flib to predict because the automaton simply does not have an adequate stock of responses. A 2-state flib, on the other hand, has four possible responses, two for each state. Thus it can predict a repeating string of four symbols but not one of five symbols; when the fifth symbol is reached, the flib must repeat an earlier response. The argument is clear. An n-state flib can predict a basic string that is $2n$ symbols long but not a string $2n + 1$ symbols long. There is some pleasant distraction to be had in devising a basic string eight symbols long and then constructing by hand the 4-state flib that will predict it. The perfect predictor thereby obtained is essentially unique. It is possible to measure the success of one's AUTOSOUP program by comparing the perfect predictor that evolves from it with the flib already constructed.

Several readers found ways to make AUTOSOUP run faster. For example, there is not much point in testing the current batch of flibs on a sequence of 100 environmental symbols if the basic string is only six symbols long. One repetition of the string will produce 12 environmental symbols, which should be enough for most purposes.

Philip Kaaret of Princeton University has pointed out that the program can also be shortened if two flibs rather than the entire population are scored on each execution of the main loop. After all, only two flibs (at most) have changed: the lowest-scoring flib has been replaced by a new hybrid, and one other flib has perhaps been struck by a cosmic ray.

The speedups obtained by shortening the environmental test sequence and by eliminating the test altogether for old flibs are roughly equivalent. Now there will be time to evolve n-state flibs that can predict repeated basic strings as many as $2n$ symbols long.

From his letter it appears that Ed Coudal of Park Ridge, Ill., was loath to send his lowest-scoring flib directly to the choir celestial. Instead he bred it with the highest-scoring flib at each cycle. By following this scheme Coudal could in fewer than 40 generations derive flibs capable of predicting a six-symbol basic string.

BOOKS

Is crime in the genes? The answer may depend on who chooses what evidence

by Leon J. Kamin

CRIME AND HUMAN NATURE, by James Q. Wilson and Richard J. Herrnstein. Simon & Schuster, Inc. ($22.95).

In this book James Q. Wilson and Richard J. Herrnstein are concerned with the explanation of "predatory street crime." They write that they would like also to discuss research on such white-collar "law-violating behavior" as fraud, bribery and embezzlement but that little such research exists. To understand street crime, they argue, we must redirect attention away from an excessive concern with social and economic factors and focus instead on differences among individual people. These often reflect biological and genetic differences, but different types of family upbringing also play a role; one way or another "bad families produce bad children" (page 215). The interplay of genes and environment creates, in some people but not in others, the kind of personality likely to commit crime. That is the broad outline of what Wilson and Herrnstein have to say.

This book, according to National Public Radio, is a "blockbuster." The dust jacket assures us that the study by Wilson and Herrnstein "will forever change the way we think about both crime and human nature." The book, in spite of "its dry-as-dust prose and its relentless blizzard of experimental reports," was immediately seen to be important by *Time; Newsweek, U.S. News and World Report, Vogue* and the *New York Times* all concurred. The conclusions reached by Wilson and Herrnstein are not new, however. They have been expressed repeatedly by Wilson, by Edward Banfield, and by other social critics of a similar ideological bent—notably and eloquently by Ronald Reagan. Why then all the fuss?

The fuss arises because Wilson, a professor of government, has now collaborated with Herrnstein, a professor of psychology, in what the two describe as a "truly interdisciplinary effort" combining the insights of political science, of economics and of a biologically informed psychology. We are not dealing here with the hortatory arguments of the evangelical right. This book, citing well over 1,000 research reports, is a product of *science,* and of Harvard science at that. This "magisterial survey of the now very extensive literature"—so it has been described by Banfield—is not mere political posturing. The research efforts of innumerable psychologists, sociologists and criminologists are said to support the authors' conclusions.

The appeal to the authority of science had been absent from Wilson's earlier, more openly speculative writings. In those prescientific works he had simply suggested that rising crime rates might be the result of such factors as "discordant homes, secularized churches, intimidated schools, and an ethos of self-expression." This tattering of the social fabric might in part be attributed to the behavior of the "educated classes," who no longer carried out their earlier function of providing "moral uplift." Rather, Freudian psychologists now taught that "repressing instincts was bad, not good," while anthropologists maintained that "this [our] culture was wrong, or at the very least no better than several competing cultural forms." Permissive child rearing, encouraged in such a climate, could have made the young "more impatient of restraints," thus predisposing them toward criminal behavior. This broad sketch might induce many or most of us to stand tall against the rising wave of secular humanism. But all of us can recognize such a sketch for what it is: a legitimate form of political argument, not science.

The new book, in what the authors clearly regard as a step forward, tries to buttress its political arguments with the support of social-science research. The trivial and often nonsensical nature of much such "research" is not, I believe, understood by many readers. The sketchy descriptions provided in general-interest books by writers such as Wilson and Herrnstein do not adequately convey what the research is really like. Therefore I want to begin by looking in considerable detail at a few of the research results cited by Wilson and Herrnstein, and at the overblown way they use those results.

The authors no longer merely suggest that today's young are more impatient. They now write (page 418) that there is "some evidence" that young people growing up in the 1960's were more "present-oriented and thus much less willing to delay gratification" than those growing up 15 years earlier. They take that evidence, from a single experimental study, at face value. They go on to suggest as one possible explanation for this tidal change in society the increasing survival rate of low-birth-weight babies. Those infants, Wilson and Herrnstein indicate, may have been damaged by the smoking or drinking of their pregnant mothers. The hapless infants survive nowadays, thanks to recent advances in medical technology, but they survive with mental defects that make them unable to delay gratification. So goes the magisterial sweep of social science, now presumably informed by biological thinking.

But what is the "evidence" that today's young people are less willing to delay gratification than those of earlier times? The single study cited by Wilson and Herrnstein compared 50 juvenile delinquents institutionalized in Rhode Island in 1974 with 57 delinquents studied in the same state in 1959. When they were asked what they would do if they were given a dollar, 16 percent of the 1974 subjects said they would save rather than spend it; in 1959, 30 percent would have saved the dollar. When the make-believe ante was raised to $100, 48 percent of the 1974 subjects, and 58 percent of the 1959 subjects, asserted that they would save the money.

Neither of these two outcomes deviates significantly from what might be expected as a result of chance and of sampling error. The Wilson and Herrnstein survey of the experimental literature seems a mite too magisterial. Their readers might profit from a more detailed description of the original studies than Wilson and Herrnstein provide. Few readers would be swept along to suppose that the decline of the West had been brought about by the smoking and drinking of pregnant women. More might appreciate that a pair of interdisciplinary social scientists, combining the insights of psychology and economics, based their analysis of societal trends on the amazing assumption that $1 in 1974 had the

same worth as $1 in 1959. The dollar had in fact lost 41 percent of its value between those dates, and any serious theory would *expect* fewer youngsters to save a dollar in 1974.

The idea of delaying gratification is central to many of Wilson and Herrnstein's ruminations about the personal defects of criminals. The authors themselves, they remind us, had to "forgo many days at the beach or playing tennis in order to write something that may or may not be purchased and read several years in the future." Criminals, however, are "more likely than authors and students to assign a very low value to distant rewards" (page 380). This claim in itself, it might be noted, fails to explain why criminals murder, pillage and rape rather than merely occupying the vacant places left at the beaches and tennis courts by scholars committed to the Protestant ethic. But make no mistake: the criminal is very different from the good citizen. He "cannot resist the rewards of an immediately available opportunity," and so he "snatches a purse if it is ready at hand." The type of citizen preferred by Wilson and Herrnstein "returns purses to their owners, waits long hours in line at the employment office, and saves his money for a rainy day" (page 314).

What is the evidence that criminals cannot delay gratifying their every impulse? There are in fact a few experimental studies, cited by Wilson and Herrnstein, purporting to show that delinquents or criminals are less likely than other people to "delay gratification" in testing situations devised and administered by psychologists. Those studies, merely listed by Wilson and Herrnstein as confirming their point, deserve more detailed examination. In one study, cited but not adequately described by the authors, delinquents and high school students in Anglo, Chicano and Mexican cultures were asked what they would do if they were to receive certain sums of money. The authors concluded that in each culture more nondelinquents than delinquents would save the money rather than spending it.

That was indeed true (if the subject's verbal declarations are thought to be credible) when the sum involved was either 25 cents or $2. But it was not true for sums of $20 or $200. With the larger sums, the proportions of delinquent and control subjects asserting that they would save the money did not differ. The researchers explained this anomaly by pointing out that the smaller sums were "realistic," whereas the larger sums "approached the fantasy level, so that all the subjects tended to start responding in terms of stereotypes that might have been quite different from what they would actually do." The fantasy responses of delinquents given large sums of make-believe money did show that "they were sufficiently cognizant of social mores to be willing to guide at least their verbal behavior by them."

That is to say, as long as the "verbal behavior" of delinquents conforms to the stereotypes of social scientists, it is taken to indicate "what they would actually do"; otherwise we are merely dealing with fantasy, and unwanted experimental results can be dismissed as meaningless. The actual experimental observation was that in response to a rather silly game forced on them by psychologists, nondelinquents asserted that they would save any amount of money given to them—25 cents no less than $200. The verbal behavior of the delinquents, in contrast, seems incomparably more flexible and realistic; they would spend the pittance but save real money.

This result seems eminently predictable from an observation in a New Zealand study cited by Wilson and Herrnstein in a very different context. That study administered a paper-and-pencil "personality test" to samples of normal people and of recidivist criminals. Tests of this kind ask the subject to describe his own personal traits by checking off items that apply to him. To detect subjects who paint too flattering a portrait of themselves—who "fake good"—such tests often contain a "Lie Scale." The authors of the New Zealand study reported as "an unexpected finding" that normal people were more likely to fake "good" answers (had higher Lie Scale scores) than criminals. The criminal evidently feels no compulsion to declare that he would rush to deposit a 25-cent windfall into his savings account. Whether this disarming candor of criminals about their less admirable traits would carry through to parole-board hearings has yet to be examined by social scientists.

The short list of studies cited in support of the claim that criminals cannot delay gratification includes a 1961 experiment carried out among black Trinidadian children. Two groups —one of institutionalized delinquents, the other of schoolchildren—were told to choose between receiving a small candy bar immediately or receiving a larger candy bar one week later. The proportion of delinquents choosing the immediate (and smaller) reward was significantly greater.

Whatever apparent meaning this bit of information might have was largely vitiated by two observations reported by the same experimenter in later papers not cited by Wilson and Herrnstein. First, the tendency of a child to select either an immediate or a delayed reward depends on who the experimenter is. Presumably that is because, unconsciously, different experimenters are likely to communicate subtly different messages to their subjects. Those messages are surely influenced by the experimenter's theoretical expectations and biases.

Second, whether a child chooses a small immediate reward or a large delayed one of the same type turns out to depend on the particular type of reward being tested. Thus, for example, the same child who prefers to receive a small candy today rather than wait a week for a candy five times as large may cheerfully forgo watching one television program today if in return he can watch five next week. That is, there is no evidence for (and considerable evidence against) a general personality trait such that some individuals more than others consistently delay gratification in order to obtain larger rewards of all sorts. Yet just such a general trait is precisely what is demanded by Wilson and Herrnstein's theory of the personal deficiencies of criminals. Troublesome details of this kind are not communicated to their readers by Wilson and Herrnstein in the course of their review of the research literature.

The same style—panache, one might call it—is exhibited by Wilson and Herrnstein in their survey of research relevant to crime rates in Japan. The level of crime in Japan is, by Western standards, astonishingly low. Why is this so? "Cultural differences," Wilson and Herrnstein declare, "may grow out of biological differences." Personality, they suggest, has a biological basis, and there is "evidence that the average Japanese personality is atypical in just such a way as to reduce the risk of crime" (page 457). To support their claim that atypical Japanese personality may account for the safeness of Tokyo's streets, they cite a single paper. That paper reviews the results obtained when translated versions of a standard paper-and-pencil test of personality were administered to people in a large number of different countries.

To be sure, in keeping with the notion that extroverts are impulsive "actors out" likely to engage in criminal behavior, the Japanese scored well toward the introverted end of a scale designed to distribute individuals along a dimension running from introversion to extroversion. In fact, this sort

of personality testing has demonstrated that only the citizens of Uganda and Ghana are more introverted than the Japanese. Although Wilson and Herrnstein do not compare crime rates in Uganda or Ghana with those in other countries, they suggest that marked introversion in Japan and high extroversion in the U.S., both measured by the personality test, might account for crime differences between those two countries.

Readers of the book have no way of knowing it, but the same personality test also provides a score placing individuals along another personality dimension, one running from normalcy to "psychoticism." The cited paper reports that with respect to this unfortunate personality trait the Japanese stand out, among all tested nationalities, as Number One. (Perhaps not surprisingly, the least psychotic countries in the world—so the science of personality testing informs us—are Canada, the U.K. and the U.S.) Psychoticism, according to that paper, "is strongly associated with a factor of brutality and insensitivity to the feelings of others." The Japanese personality, according to the study, is atypical with respect to its brutal insensitivity. The same paper goes on to point out that in Japan, Britain and Hungary prisoners have been found to score relatively high on psychoticism as well as on extroversion. That is, the Japanese have rather low scores on one trait found to be associated with criminality but stunningly high scores on another "criminal personality trait."

The manner in which Wilson and Herrnstein use these findings is not unusual in social science. Tiny snippets of data are plucked from a stew of conflicting and often nonsensical experimental results. Those snippets are then strung together in an effort to tell a convincing story, rather in the manner of a clever lawyer building a case. The data do not determine the conclusions reached by the lawyer. Instead the conclusions toward which the lawyer wants to steer the jury determine which bits of data he presents.

The examples so far discussed have come from what is often called soft social science. Perhaps more reliable data, and greater rigor, can be expected from the hard, more biologically oriented sector of the social-science spectrum. That is not, however, the case; consider how Wilson and Herrnstein make use in their book of the somatotype data accumulated by W. H. Sheldon and his followers.

The somatotyping procedure was developed by Sheldon at Harvard as a technique for measuring variation among individuals in physique, or body build. The procedure, based on the inspection and measurement of photographs of nude individuals, described any physique in terms of the relative preponderance of three basic components. That is, numerical ratings were assigned to describe the roundness, the muscular squareness and the vertical linearity of any given physique. When Sheldon reported in 1949 on a study of 200 Boston delinquents, he indicated that their physiques were preponderantly "mesomorphic," or of the squarish, muscular type. Their excessive mesomorphy was made evident by comparing them with a sample of 4,000 college students, also somatotyped by Sheldon. Wilson and Herrnstein, who wish to demonstrate that there are constitutional (and possible genetic) correlates of crime, follow Sheldon's 1949 presentation of his data accurately enough. They reproduce in their book Sheldon's old graphs, in which he compared the delinquents with the 4,000 students, and they assert that the delinquents are "sharply skewed toward the mesomorphic" (page 87).

They make no reference, however, to a number of informative somatotyping studies that have appeared since 1949. For example, one sample of Princeton students turned out to be more mesomorphic than the Boston delinquents. That was also true of samples of bus drivers and truckmen, of children growing up in California and of 400 Army recruits. The last two samples are of special interest, since they were somatotyped by Sheldon himself. To describe the delinquents as "skewed toward the mesomorphic" is clearly misleading. With respect to whom? To what control group should the delinquents be compared?

There is clear evidence that somatotypes differ across various ethnic groups, and they are doubtless affected as well by such factors as nutrition and exercise. There is therefore no point in comparing deprived delinquents from one set of ethnic groups with well-off college students from another. To pin down a particular body build as a "constitutional correlate of crime" it is necessary to compare the physiques of delinquents with those of a "matched control group" of nondelinquents. The two groups should be matched for ethnicity, for age and—as far as possible—for socioeconomic background.

That is what was supposedly done in a study reported in 1950 by Sheldon Glueck and Eleanor Glueck, in which 500 Boston delinquents and 500 matched controls were compared. The Gluecks reported that the delinquents—now compared with an apparently appropriate control group—were significantly more mesomorphic in body build.

There are problems, however, even with this seemingly clear-cut demonstration. Although all the boys were said to come from homes in poor neighborhoods, the parents of the delinquents had a lower income and less education than the parents of the controls. The delinquents' homes were more crowded and had fewer sanitary facilities. And so any differences in physique between the two groups were once again confounded with socioeconomic differences.

Moreover, the effort to match the two groups for age failed. To a significant degree the delinquents were older than their so-called matched controls. The boys in the Gluecks' study ranged from 10 to 17; most of them were 14 or 15. There were differences in age of six months or more (ranging up to 14 months) in more than half of the matched pairs. From data the Gluecks presented in an appendix it is possible to calculate that when such relatively large age differences occurred, the delinquent was more than twice as likely as the control to be the older one of the pair. The age differences bracket the time of male puberty and the very rapid spurt in growth associated with puberty. If postpubescent boys have more masculine-looking bodies than prepubescent boys, there is little reason for surprise.

Finally I should mention (although I shall not discuss the point in detail) that as part of their effort to show a genetic predisposition toward criminal behavior Wilson and Herrnstein review a number of studies of twins and of adopted children. The problems with studies of this kind are by now, of course, numbingly familiar to all scholars in the field.

The studies do not and cannot effectively separate genetic and environmental variables, and the data therefore cannot be unambiguously interpreted. Were that not the case, the debate over the relative importance of genetic and environmental factors in determining I.Q. scores would long since have ended. The "genetic" studies in the I.Q. area are greater in both number and sophistication than those concerned with criminality, but even they are wholly inconclusive.

I have written in such detail about particular experimental results because I hoped to make clear the tenuousness of the "facts" on which much social-science theory is based. To suppose that theory resting on so wobbly a base can provide useful guidance

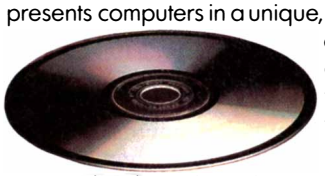
SHELF-BASED EXPERT SYSTEM.

An expert system is like borrowing someone else's experience and expertise on a given subject. UNDERSTANDING COMPUTERS by TIME-LIFE BOOKS *is* exactly that, an expert system on the subject of computers.

It's not software based, so it can't crash. And since it's not machine-specific, it won't become outdated easily, which means you'll find it an invaluable, long-lived reference whether you have an IBM®PC, an Apple IIe®, a multimillion dollar Cray®, or no computer at all right now.

A BRAND NEW SERIES THAT PRESENTS COMPUTERS IN A WHOLE NEW WAY.

UNDERSTANDING COMPUTERS is a new series of books that presents computers in a unique, broad-based way, unlike any other computer book you've ever seen. It gets "under the hood," right down to the nuts and bolts of computers to explain what you need to know about them in plain English.



On a digital record, sound is carried by pits and spaces, which are read by a laser beam.

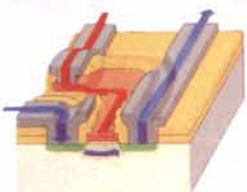
The first volume, *Computer Basics*, unwraps the mystery of writing binary code…the concept of logic gates…how chips are designed…how a light pen works for graphic effects…the sequence of events in a computer from the first clock pulse to the last. And much, much more.

Succeeding volumes cover *Software. Input/Output. Graphics. Networking.* The entire gamut of computer topics.

EVERY IMPORTANT CONCEPT IS ILLUSTRATED TO HELP ASSURE YOU UNDERSTAND.

Not only does UNDERSTANDING COMPUTERS give you all the computer background you want that you won't find elsewhere, it also *shows* you what it's all about. Full-color illustrations help make every concept crystal clear. And glossaries of terms help ensure your understanding. Each volume is a big 9¼" x 11⅛" hardcover book with approximately 128 pages.

FREE 10 DAY EXAMINATION. NO OBLIGATION.



How a transistor operates as a switch.

Begin to expand your computer know-how now by examining *Computer Basics* for 10 days free. Keep it and pay only $12.95 ($15.95 in Canada), plus shipping and handling. Then you can continue to receive other volumes every other month, always with a 10-day free trial. Keep only the books you want. And you can cancel any time. Or return *Computer Basics* and owe nothing.

Return the coupon today, and start to bridge the gap from computer buff to computer expert.

UNDERSTANDING COMPUTERS STARTS HERE.

Mail to TIME-LIFE BOOKS, Inc.
Time & Life Building,
Chicago, IL 60672-2058

YES! Send me *Computer Basics*, as my introduction to the shelf-based expert system, UNDERSTANDING COMPUTERS. I agree to the terms outlined in this ad.

D1AMR6

Name_____

Address_____

City_____

State or Province_____ Zip or Postal Code_____

All orders subject to approval. Price subject to change.

TIME LIFE BOOKS

UNDERSTANDING COMPUTERS

about the real world is to make wine out of water.

The lack of a sound empirical basis for their arguments is not the main criticism to be made of the Wilson and Herrnstein work, however. Far more serious is the fact that the book—in common with much of contemporary social science—suffers from a fundamental conceptual flaw. The authors often seem unwilling to distinguish clearly between correlation and causation. What if it really were the case that delinquents tend toward a particular type of physique? What would follow from the discovery of such a statistical correlation?

Wilson and Herrnstein assert (page 71) that if it can be shown that criminals do not differ from others in bodily constitution, "then it would seem plausible to suppose that crime results from a criminal's economic, cultural, social, and political circumstances, rather than his constitution." With such a simplistic and false dichotomy once set up, their next step is to argue: "Showing a constitutional correlate amounts to counterevidence against the purely environmental explanation." The task of research therefore boils down to a hunt for the external constitutional correlates of crime. The light at the end of the research tunnel is dazzling: "If it turned out that most criminals were, say, redheaded and freckle-faced, we would be on the trail of genetic correlates of crime, just because redheadedness and freckles have a genetic basis."

This kind of logic asserts that should it turn out that people with black skin make up a disproportion of the unemployed, we are on the trail of a genetic correlate of unemployment—just because black skin has a genetic basis. We are on such a trail, of course, but a "purely environmental explanation" attributing black unemployment to racism need not deny that the skin color of blacks is attributable to their genes. Although their genes are indeed correlated with the state of being unemployed, there is no meaningful sense in which the genes of black people cause their unemployment. We can easily imagine a society in which skin color would no longer be a genetic correlate of unemployment.

The confusion between genetic correlates and genetic causes is plainly illustrated by Herrnstein's observation (in an earlier magazine article) that "as technology advances, the tendency to be unemployed may run in the genes of a family as certainly as bad teeth do now." The simple fact is that unemployment, like crime, is a social phenomenon; neither of them can "run in the genes." (I shall leave undiscussed the effects of a fluoridated water supply on the relations between genes and teeth.)

The fallacy of confusing correlation and cause is compounded by Wilson and Herrnstein's repeated suggestion, equally fallacious, that behaviors with a genetic correlate cannot be changed by conscious social action. For example, they review the research literature on innate sex differences in behavior—a literature that makes the research on delay of gratification look like a Gibraltar of empirical solidity. They write (page 125): "The underpinnings of the sexual divisions of labor in human society, from the family to commerce and industry to government, may not be rigidly fixed in the genes, but their roots go so deep into the biological substratum that beyond certain limits they are hard to change." Whatever aid and comfort Messrs. Wilson and Herrnstein may draw from it, there is nothing in genetic science, or in any science, to justify their claim that what now is must always be. Tell that one to the marines—the women marines, that is.

Wilson and Herrnstein tread delicately when writing about race and crime, but not delicately enough. They state (page 466): "If blacks are more likely to have an impulsive temperament or a somewhat lower measured I.Q., these traits may be the result of patterns of prenatal care as well as of inheritance." They then cite research to show that blacks are less "normal" than whites in their personality test scores, have lower I.Q.'s and are more likely to be of low birth weight. Criminals, remember, are said to be impulsive and unintelligent.

The reference to poor prenatal care was not intended to rule out possible genetic explanations for black crime. Wilson and Herrnstein grant that a relative increase in the black homicide rate in Philadelphia over the past century compared with the rate for whites cannot plausibly be attributed to genetic changes. Nevertheless, "purely genetic factors . . . may have made the average black male more vulnerable to changing circumstances, such as the greater availability of handguns, alterations in economic opportunities, or the pressures of racial animosity" (page 472).

This display of evenhandedness—maybe too many guns have been made available to blacks, or maybe too few jobs—implies in either case that certain genes, frequent in blacks, interact with environmental conditions to cause excessive crime. The fatal (and insensitive) confusion between correlate and

cause is again obvious. If blonds were subject to pervasive discrimination and committed crimes, would it make sense to talk of their genes as causing both blondness and crime?

The book does, of course, contain some appropriately worded comments to the effect that the relations between genes and the social environment are complex, so that the two should not be thought of as "either-or." Lip service aside, however, Wilson and Herrnstein repeatedly imply that genes and environment are radically separate sources of causation, and that when variations in the two are correlated, as is usually the case, causation resides in the genes. *Time* did not vulgarize their argument or thought when it titled its review of their book "Are Criminals Born, Not Made?" That title clearly did not offend the authors; one month later, when they spoke to a public meeting in Cambridge, the title of their talk was "Are Criminals Made or Born?"

The Wilson and Herrnstein work ought not to be judged in isolation. Their selective use of poor data to support a muddled ideology of biological determinism is not unrepresentative of American social science in the sixth year of the Reagan presidency. The political climate of the times makes it easy to understand why social scientists now rush to locate the causes of social tensions in genes and in deep-rooted biological substrata. Not many years ago their forebears cheerfully enlisted as environmentalist shock troops in the subsidized research battalions of Lyndon Johnson's Great Society. What remains of social "science" if developments within it are so slavishly dependent on transient political developments?

Wilson has, to be sure, insisted to *Time,* "This [book] has nothing to do with the conservative times. Do not put the book in that framework." Yet expert authority can be quoted in support of the view that the flight from social toward biological "explanations" is politically motivated. The stress on biology arises now because biological factors "seem beyond the reach of legislation or politics. When expectations for public policy are high, such as in the recent past, sociological hypotheses are in tune with the times." And alas, "sometimes people do not choose theories at random; very often, they choose them in part because the central factors in the theories... are ones, which for political or ideological reasons, the defenders of the theories *want* to believe are central." What is the source of these apposite quotations? Wilson and Herrnstein (pages 80 and 42).

What will be.

It will be new. All that has been will be the beginning. It will be a special shape. One that will cut cleanly through the air.

It will be lower to the ground. It will be larger inside. It will be a joy to drive. It will be the luxurious Honda Accord LXi 4-Door Sedan.

The hood angles down for greater aerodynamic efficiency. Under the hood a larger 2-liter engine designed with Honda Programmed Fuel Injection delivers smooth power.

A unique suspension system provides outstanding handling. It's called double wishbone. Race cars have used it for years. Nothing uses it like the Accord does.

Driving one will be an experience you will want to repeat. For once, the future is yours to see. HONDA

The New Accord

Established 1845 SCIENTIFIC AMERICAN February 1986 Volume 254 Number 2

Dioxin

Concern that this material is harmful to health or the environment may be misplaced. Although it is toxic to certain animals, evidence is lacking that it has any serious long-term effect on human beings

by Fred H. Tschirley

Dioxin. The word evokes a variety of reactions. Much of the public worries that the compound will cause poisoning even at minimal exposure. Toxicologists, knowing the severe toxic effects of dioxin in experimental animals but being uncertain about comparable serious effects on people, call for more research. Regulators, who must make decisions based on this conflicting evidence, are left wondering what to do. As one who has spent many years studying the dioxin issue, I hope in this article to provide a useful perspective for making judgments about the potential hazard of the material.

Dioxin is actually a name for a family of chemical compounds. The name refers to their basic structure: two oxygen atoms joining a pair of benzene rings. Substitution of chlorine atoms for hydrogen atoms on the rings produces a chlorinated dioxin, of which there are many. The chlorinated dioxin of interest here is 2,3,7,8-tetrachlorodibenzo-p-dioxin, usually abbreviated to TCDD. It is a by-product of the manufacture of trichlorophenol, which serves in the manufacture of two herbicides (the best-known being 2,4,5-trichlorophenol, or 2,4,5-T, one of the ingredients of Agent Orange) and the antibacterial agent hexachlorophene. TCDD is also produced in a variety of combustion processes.

The legacy of Agent Orange, Times Beach, Seveso, several industrial accidents and other instances of human exposure to significant amounts of TCDD is widespread concern about adverse health effects resulting from minimal exposure to the material. Yet none of the many studies directed at this question have demonstrated that TCDD causes severe chronic human effects. Moreover, not one human death has been attributed to TCDD, even though exposure has been high in a number of cases. The issue points up the broader problem of the difficulty faced by regulators who must make judgments on the basis of incomplete scientific knowledge on the one hand and public fear on the other.

TCDD was first recognized in 1957 as a contaminant of 2,4,5-T, when 31 workers involved in the manufacture of the herbicide in West Germany developed the dermatologic affliction now called chloracne. It is a skin eruption resembling acne and takes its name from the fact that it is caused by exposure to various chlorinated organic chemicals.

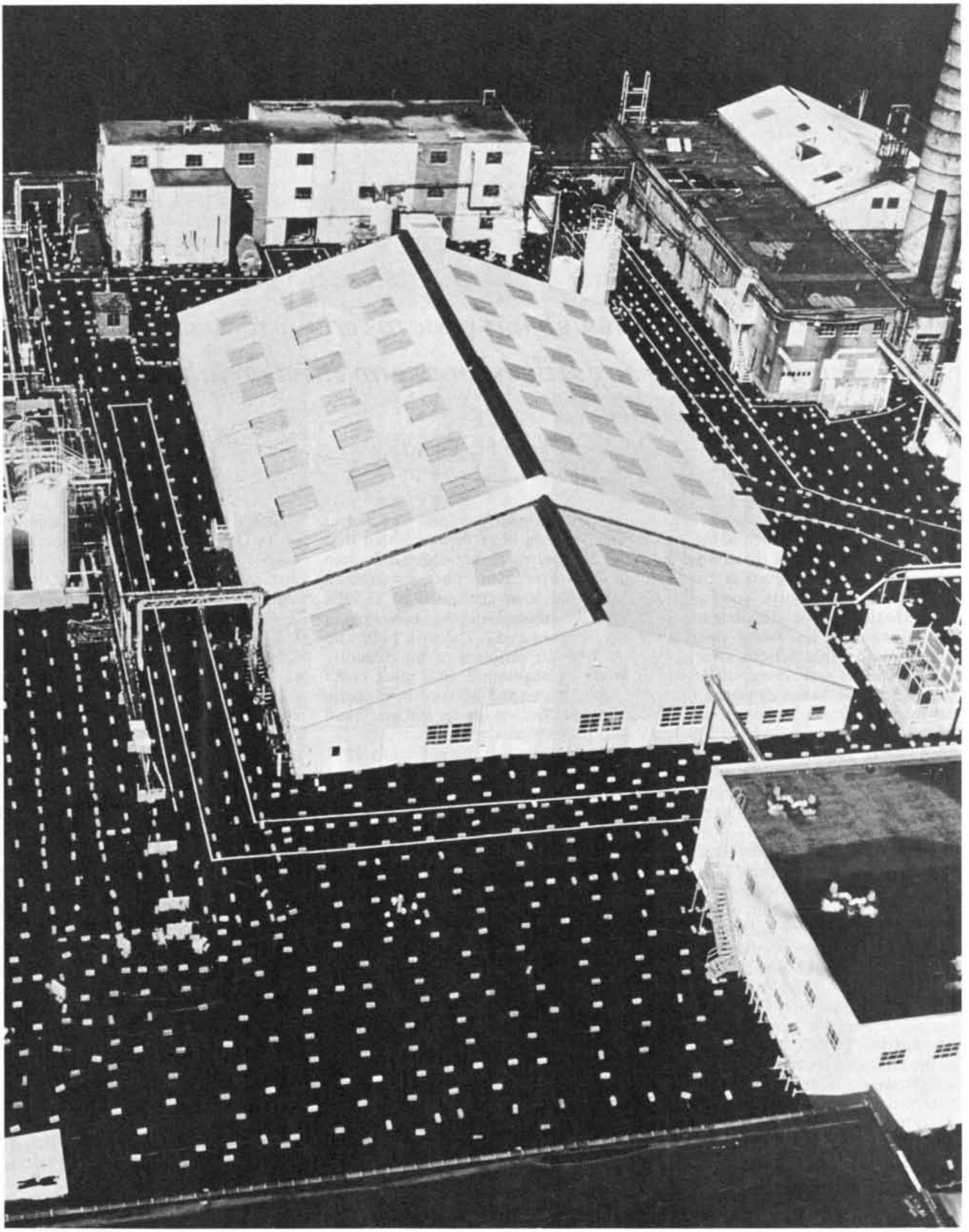
General awareness that TCDD is a potential hazard to health and the environment arose in 1970, when a House subcommittee held a hearing on "Effects of 2,4,5-T on Man and the Environment." The hearing dealt among other things with a study by the Bionetics Research Institute showing that 2,4,5-T caused birth defects in animals. Testimony suggested that the teratogenic component of 2,4,5-T may have been TCDD. In the sample of 2,4,5-T tested by Bionetics the dioxin occurred as a contaminant at an extremely high concentration: 27 ± 8 parts per million.

Since then there has been a steady accumulation of information about the sources of TCDD, its environmental fate and its toxic effects. One well-defined source is the formation of TCDD during the manufacture of 2,4,5-T. The amount of TCDD formed increases as the temperature of the reaction and the *p*H (degree of alkalinity) increase.

In 1977 investigators in the Netherlands reported that polychlorinated dibenzo-p-dioxins (PCDD's) were present in the fly ash from a municipal incinerator. Typically such an incinerator burns, among other things, organic wastes containing chlorine. Similar reports soon came from Switzerland, Canada and Japan. It was believed the compounds resulted from the condensation of chlorophenols. Later quantitative data demonstrated from a wide variety of combustion sources the presence of TCDD's that could not be explained on the basis of preexisting polychlorinated phenols.

These findings led R. R. Bumb of the Dow Chemical Company and 12 of his co-workers to put forward in 1980 the hypothesis that PCDD's can result from trace chemical reactions in fire. The hypothesis has been challenged because the reactions have not been defined. Nevertheless, PCDD's have now been found in the effluent and ash of so many combustion processes that there is no longer serious argument about their formation during combustion, even though the precise nature of the process remains obscure.

Moreover, TCDD has been specifically identified in soil and dust from numerous places, in soot from the chimneys of wood furnaces, as residues in river fishes (some from rivers whose watersheds do not have industrial operations known to form



PLASTIC TARPAULINS cover much of the open area of the Diamond Alkali Co. plant in Newark, N.J., because the land may be contaminated by a chlorinated dioxin: 2,3,7,8-tetrachlorodibenzo-p-dioxin, usually abbreviated as TCDD. From 1943 through 1968 the company manufactured chemicals at the plant, including the herbicide 2,4,5-T. TCDD is a by-product of the manufacture of 2,4,5-T. Federal and state officials had the tarpaulins put down as a precaution because TCDD is toxic to a number of animals tested in experiments. The material has not been shown to have severe chronic effects on human beings. It does produce some short-term effects.

TCDD), as residues in the eggs of herring gulls and recently in adipose tissue from more than 100 people in Canada, the U.S. and Vietnam. It appears that TCDD is a ubiquitous chemical, particularly in industrialized nations.

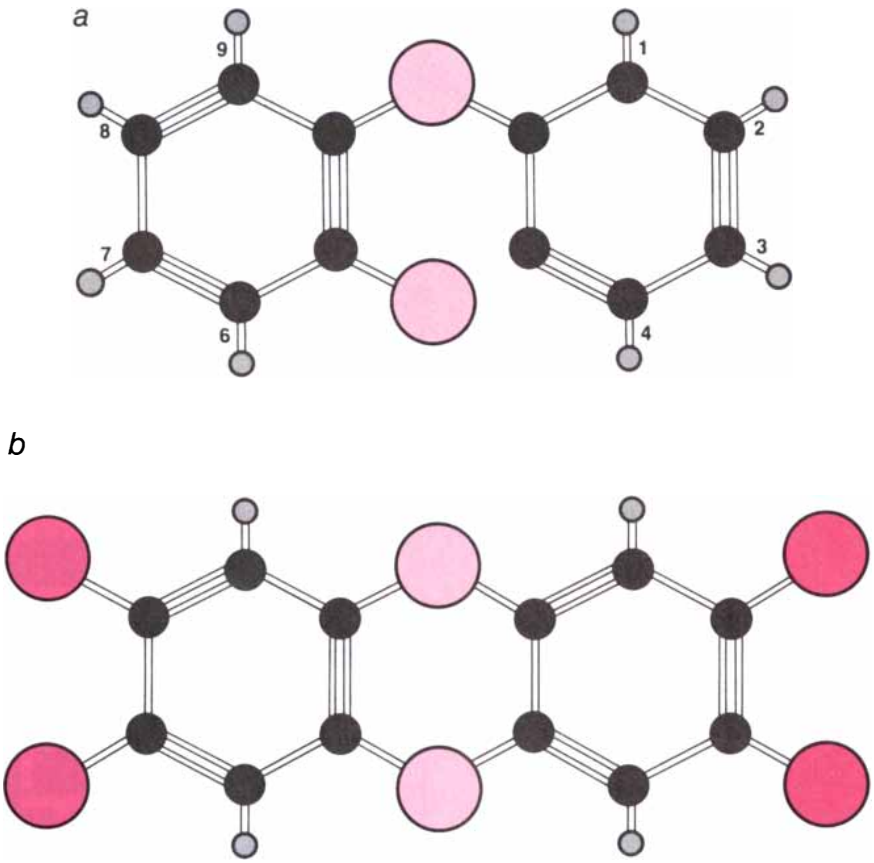
One might then ask why it was not detected sooner. For one thing, no one looked for it seriously before the 1970's. At the time it would not have been found anyway, except in unusual circumstances, because analytic chemists were only able to detect concentrations of a few parts per million. Since then the detecting equipment has improved at least a millionfold, so that concentrations of a few parts per trillion are now detected routinely. In addition dilution and the destruction of TCDD by light may reduce concentrations to undetectable levels. As analytic technology improves, one can expect that TCDD will be found in many more sites than are currently known.

In places where TCDD is protected from light it is an extremely persistent material. In the early 1970's it was thought the half-life of TCDD (the time required for half of a given amount to be degraded) was about a year; later studies in the U.S. suggested the half-life might be as long as three years. Recent reports from Italy raise the possibility that the half-life of TCDD in soil might be 10 years or even more.

Although precise data are meager, TCDD is known to be strongly held by most soils. The strength of the binding is inferred from the fact that known concentrations of TCDD applied to soil have remained near the surface. Even in the sandy soil examined in Florida the concentration of TCDD in the upper 15 centimeters was as high as 1,500 nanograms per kilogram 10 to 12 years after application.

In the most heavily contaminated zones near Seveso TCDD has been found at a depth of 136 centimeters. Moreover, the concentrations well below the surface were slightly but significantly higher in 1977 than they were in 1976, soon after the accident took place. The presence of soil fissures does not adequately explain this unexpected vertical distribution.

The processes that degrade TCDD in soil are poorly known. Microorganisms do degrade the substance, but at a low rate. S. D. Aust of Michigan State University found a wood-decaying fungus (the white mold *Phanerochaete chryosporium*) that breaks down TCDD without observable mortality of the organism. The rate of degradation is low, but it is conceivable that contaminated sites could be inoculated with the mold to speed up the degradation process.



CHEMICAL STRUCTURE of a dioxin and of TCDD is depicted. The dibenzo-p-dioxin molecule (*a*) consists of two benzene rings joined by two oxygen atoms (*light color*). From one chlorine atom to eight atoms can substitute for hydrogen atoms (*gray*) attached to carbons (*black*) at the numbered positions to form any of the 75 chlorinated dioxins. The chlorinated dioxin known as TCDD (*b*) is shown with its four chlorine atoms (*dark color*).

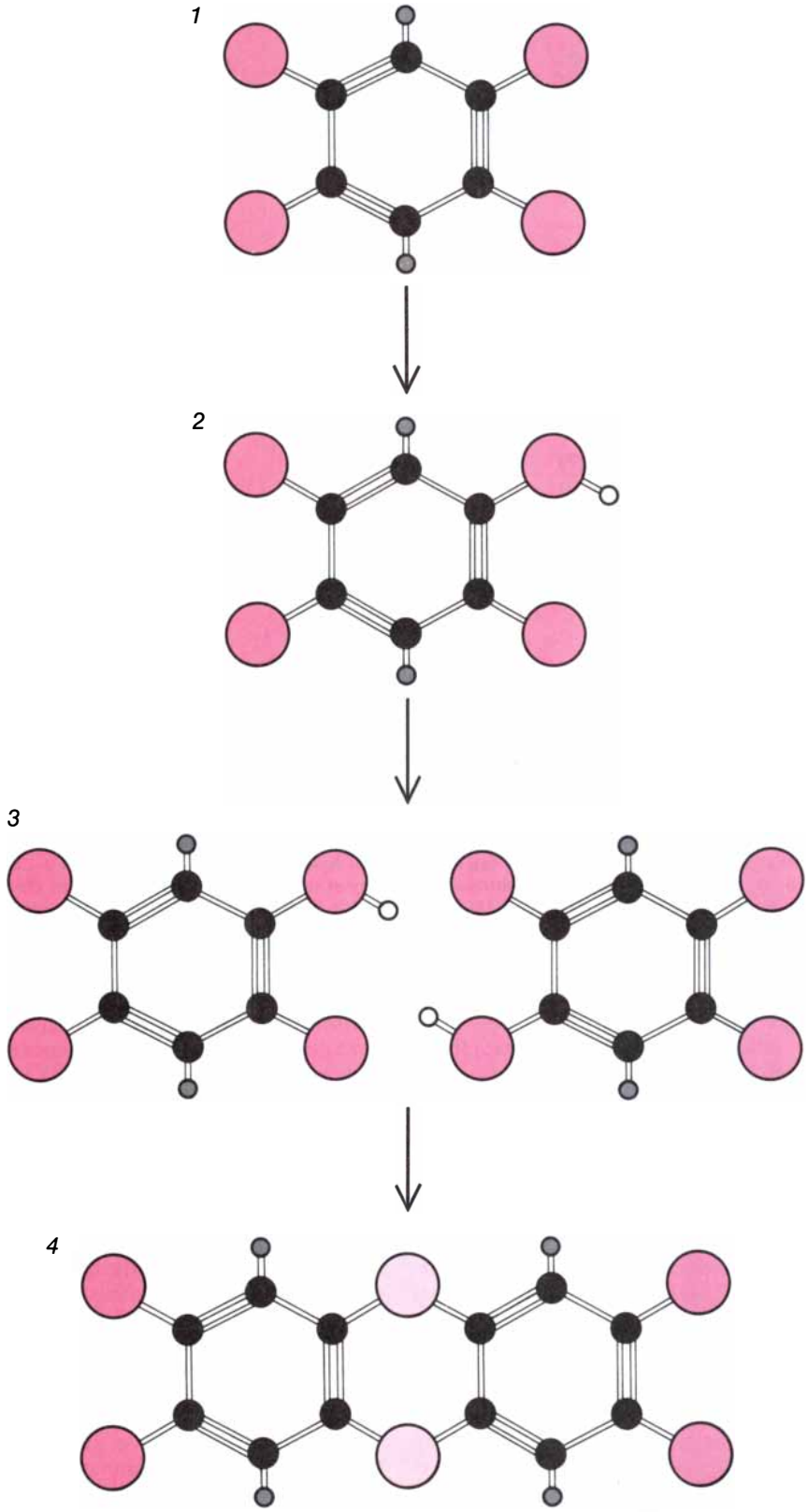
Sunlight degrades TCDD rapidly by splitting off the chlorine atoms. The reaction requires a hydrogen donor, which is usually available in water or in the wax on leaves. Experiments by Donald G. Crosby of the University of California at Davis showed that 40 percent of the TCDD layered on a glass plate remained after six hours in sunlight; the amount was from 25 percent to negligible when the material was applied as drops on leaves of the rubber plant, but on a loam soil the figure was 85 percent.

TCDD is a highly toxic chemical in experimental animals. The first animal-toxicity test performed is usually a determination of the LD_{50}: the dose that kills half of a test population. Between 1973 and 1978 the LD_{50} of TCDD was determined for eight species. The guinea pig was by far the most sensitive species tested: its LD_{50} for an oral dose was .6 microgram per kilogram of body weight. The hamster was the least sensitive animal tested: its oral-dose LD_{50} was about 1,900 times as high as the guinea pig's. Its intraperitoneal level, 3,000 micrograms per kilogram, was about 5,000 times as high as the oral dose for the guinea pig.

The reasons for the extreme range of acute (short-term) toxicity may relate to the relative speed of clearance from the body. In the hamster half of a dose is removed within 15 days, compared with 30 days for the other species tested. Even though the hamster is much less sensitive than the guinea pig, an intraperitoneal LD_{50} of only 3,000 micrograms per kilogram signifies an extremely toxic material: the toxicity is comparable to that of the insecticide parathion.

Many essentially acute symptoms have been observed in human beings. They include chloracne, digestive disorders, effects on some essential enzyme systems, aches and pains of muscles and joints, effects on the nervous system and psychiatric effects. These symptoms have been transitory except for a few severe cases of chloracne.

Additional tests have measured the chronic, or long-term, effects of



STEPS IN FORMATION of TCDD during the manufacture of 2,4,5-trichlorophenol (TCP) are shown. (TCP is used to make the herbicides 2,4,5-T and Silvex and the antibacterial agent hexachlorophene.) In the process of making TCP a molecule of 1,2,4,5-tetrachlorobenzene is hydrolyzed (*1*) in the presence of caustic soda to form 2,4,5-trichlorophenate (*2*). Two molecules of the phenate (*3*) combine to form TCDD (*4*). They lose atoms of sodium (*white*) and chlorine. The higher the temperature and the alkalinity are, the greater is the amount of TCDD that is formed. When the mixture temperature is above 180 degrees Celsius, a heat-forming reaction occurs; it has been the cause of industrial accidents.

TCDD in rodents, rabbits and nonhuman primates. Chloracne, the most sensitive indicator of human exposure to TCDD, has also appeared in rabbits, nonhuman primates and hairless mice. The skin and body can become dry and scaly. A few species lose hair. Some nonhuman primates lose fingernails and toenails without apparent evidence of pain.

TCDD also causes reproductive effects in experimental animals. Cleft palate and abnormalities of the kidney were caused in the offspring of exposed mice at dosage levels of from one nanogram to three nanograms per kilogram of body weight per day. Similar doses in rats caused the death of fetuses. In monkeys a dose of 1.7 nanograms per kilogram per day for two years caused abortions in four out of seven pregnancies.

TCDD is a proved carcinogen in rats and mice. Although the test results vary somewhat, there is fairly good agreement among tests by different investigators. The liver is the primary target in both rats and mice, although the brain, the respiratory system and the thyroid gland have also been involved in a few studies. It is important to recognize that an oncogenic response was reported only after the animal had ingested high doses of TCDD over a long period of time. Rats showed no oncogenic response at dosage levels of from one nanogram to 1.4 nanograms per kilogram per day; mice showed none at dosage levels ranging from one nanogram to 30 nanograms. Moreover, a study by R. J. Kociba of Dow Chemical and his co-workers showed that rats tolerated a daily dose of one nanogram per kilogram per day for two years without showing toxicologic effects.

Nevertheless, the findings from the tests on animals intensified concern about the effects of TCDD on people. On several occasions people have been exposed to "high" levels of TCDD. In this context "high" is a relative term because with one exception—a group of prisoners who volunteered for tests with TCDD—the amount of the material to which a person was exposed is not accurately known. The criterion I employ here to distinguish high from low exposure is whether or not the exposure results in chloracne.

The case of the prisoners is important because known amounts of TCDD were applied to their skin. In the first experiment 60 volunteers were treated with concentrations ranging from 200 to 8,000 nanograms (from three to 114 nanograms per kilogram for a 70-kilogram person), and the

dose was repeated two weeks later. The dosages chosen were those that had caused chloracne when they were applied to the ears of rabbits. None of the volunteers developed chloracne, and no other symptoms were observed. The second experiment involved 10 volunteer prisoners who were treated with 107,000 nanograms of TCDD per kilogram. Eight of them developed chloracne, but no other symptoms were noted.

From these experiments one can conclude only that TCDD does cause chloracne in humans when the dose is sufficiently high but that people are less sensitive than rabbits. The tests did not identify a threshold for the development of chloracne in human beings—a piece of information that would be of great value.

The number of people who have been exposed to high levels of TCDD cannot be determined accurately, but it must be in the thousands. Alistair Hay of the University of Leeds has estimated that in the chemical industry alone about 2,000 workers have had high exposure. Low levels of exposure have undoubtedly been experienced by people who handle the herbicides 2,4,5-T and Silvex, in which TCDD was a contaminant; by Vietnam veterans exposed to Agent Orange (50 percent of which was 2,4,5-T); by residents of Times Beach, Mo., where waste oil that contained TCDD was spread on the ground in several places; by chemical-industry workers making the products that include the material, and by many thousands of people who have eaten food (notably fish) containing trace amounts of TCDD or have been exposed to fallout from combustion processes that form TCDD. The total number of individuals with such low exposures probably runs well into the millions.

SPECIES	ROUTE	LD_{50} (micrograms per kilogram)
GUINEA PIG (MALE)	ORAL	.6
GUINEA PIG (FEMALE)	ORAL	2.1
RABBIT (MALE,FEMALE)	ORAL	115
RABBIT (MALE,FEMALE)	DERMAL	275
RABBIT (MALE,FEMALE)	INTRAPERITONEAL	252–500
MONKEY(FEMALE)	ORAL	<70
RAT (MALE)	ORAL	22
RAT (FEMALE)	ORAL	45–500
MOUSE (MALE)	ORAL	<150
MOUSE (MALE)	INTRAPERITONEAL	120
DOG (MALE)	ORAL	30–300
DOG (FEMALE)	ORAL	>100
FROG	ORAL	1,000
HAMSTER (MALE,FEMALE)	ORAL	1,157
HAMSTER (MALE,FEMALE)	INTRAPERITONEAL	3,000

ACUTE TOXICITY of TCDD in experimental animals is ascertained on the basis of the LD_{50}: the dose (in micrograms per kilogram of body weight) that kills half of a test group.

The possibility of chronic effects from exposure to TCDD causes far greater public concern than that of acute effects. An aspect of this problem about which little is known is the effect of protracted exposure to low levels of the material, as might occur in an occupational setting or from incidental exposures to, for example, the fallout of combustion effluents or to fish that contain low levels of TCDD.

A look at some of the major exposures to TCDD, approximately in order of their severity, reveals few if any unambiguous chronic effects. The industrial accident at Seveso in 1976 exposed some 37,000 people of all ages to considerable amounts of TCDD. A relatively small number of them showed transient effects such as chloracne (184 cases, 164 of them children under the age of 15), headaches and digestive upsets, but no long-term effects such as birth defects and chromosomal damage have been identified. It is too early to tell whether the incidence of cancer is abnormal.

An accident in a Monsanto plant in Nitro, W.Va., in 1949 exposed more than 200 workers to TCDD. Of 122 who developed chloracne, 121 were monitored for the next 30 years. The total number of deaths in that group did not differ significantly from that expected in the population at large, and there were no excess deaths due to cancer or diseases of the circulatory system. Similar findings have been made after other industrial accidents, except for two in which an excess of deaths from cancer was found in small groups of the people exposed.

A particular type of cancer (soft-tissue sarcoma, a generic term for more than 100 different types of rare cancer) has become a focus of concern because of a survey of Swedish forestry workers by Lennart Hardell of the University of Umea. He concluded that their exposure to 2,4,5-T (and thus to TCDD) had caused six times the normal incidence of soft-tissue sarcoma.

This study led to an investigation of chemical-plant workers in the U.S. who had been exposed to 2,4,5-T and other chemicals. Seven apparent cases of soft-tissue sarcoma were discovered, raising the level of concern substantially. Subsequent events have emphasized the difficulties in accurate diagnosis of soft-tissue sarcoma and in accurate identification of exposed individuals. At a conference in 1983 Marilyn A. Fingerhut of the National Institute of Occupational Safety and Health reported that two of the seven people had in fact died of cancers other than soft-tissue sarcoma. Moreover, the exposure of three others to TCDD could not be documented. Such findings fall far short of being hard evidence for the proposition that TCDD causes soft-tissue sarcoma.

Other studies also fail to support Hardell's hypothesis. In the state of Washington no consistent pattern of death due to soft-tissue sarcoma was found among occupations in which workers would have been exposed to TCDD. A study in Finland found no cases of the disease among 1,900 people who applied herbicides, nor was their death rate from any natural cause different from that of the total male population in Finland. The U.S. Air Force, in its Ranch Hand study of about 1,200 military personnel who sprayed Agent Orange in Vietnam, found no cases of soft-tissue sarcoma. Finally, examinations by the Veterans Administration of 85,000 self-selected veterans showed fewer cases of these cancers than the national average would suggest.

Reproductive effects are also a subject of concern because of the animal findings. The most celebrated case alleging such effects in humans is commonly known as the Alsea II study, made by the U.S. Environmental Protection Agency. The study reported a link between the spraying of 2,4,5-T on foliage and spontaneous abortion among pregnant women in Alsea, Ore.

This study has come in for much criticism, notably by an interdisciplinary group at Oregon State University. The group concluded that an asso-

ciation between herbicide spraying and spontaneous abortion could not be shown from the data relied on by the agency. Other studies—in Australia, Hungary, New Zealand and the U.S.—failed to find a link between the use of 2,4,5-T and birth defects.

Because of the extreme acute toxicity and the multiple chronic effects of TCDD in animals, regulatory agencies have had to consider what to do in order to protect people from exposure to the material. Such agencies must extrapolate animal data to human beings in all but a few instances, in spite of the fact that the validity of this type of extrapolation has not been ascertained. Compounding the difficulty is the lack of a simple, accurate method for determining whether and at what level TCDD occurs in the tissues of exposed individuals. (The present test requires a surgical procedure to obtain samples of the liver and fat tissues where TCDD resides.) Without such information a dose-response relation cannot be established.

TCDD has been called the most toxic synthetic chemical known to man. If its acute toxicity to the guinea pig, and even the rat and the mouse, is the criterion, the statement is probably correct. If its considerably lower toxicity to the hamster is the criterion, however, the statement would surely not be true. Yet there is no need to quibble: TCDD is unquestionably a chemical of supreme toxicity to experimental animals. Moreover, severe chronic effects from low dosages have also been demonstrated in experimental animals. Therefore the concern about its effects on human health and the environment is understandable.

When toxic chemicals are at issue, a regulatory agency has few options beyond extrapolating animal data to humans. Yet health effects on humans are rarely proved in the case of environmental chemicals to which the public is variably exposed at subacute levels that can only be estimated (and then only in the crudest approximation). A case in point is aflatoxin, the product of a mold that develops commonly in stored oilseed crops such as peanuts. In animal tests aflatoxin is one of the most potent carcinogens known, but it has not yet been proved to have this effect in human beings.

Diversity in reaction to stimuli is a hallmark of biological organisms. Reactions to toxins are no exception to the general rule. People may be more, less or equally sensitive to a given toxin than an experimental animal is. Extrapolation is neither art nor science; it is simply the most rational way to assess a hazard in the absence of definitive data. Hence regulatory actions continue to be based on the animal data even when the human data, although they are not definitive, may be sufficiently compelling to allow a scientific judgment that the hazard to people has been overestimated.

That appears to be the case with TCDD. Investigators are in general agreement that TCDD is less toxic to humans than it is to experimental animals, but the available information is not sufficiently compelling to stimulate a change in regulatory posture toward either more or less restriction of exposure to the material. I suspect that the direct evidence of TCDD's effects on humans will never be either more or less compelling than it now is.

The public's perception of a toxin is an important determinant of the posture taken by a regulatory agency. The public has heard a great deal about both the acute and the chronic effects of TCDD on experimental animals but little about the substantial body of data showing that human beings are less sensitive. The initial reports of TCDD's acute toxicity, followed by reports of its carcogenicity and reproductive effects, have instilled a public fear that probably cannot be dispelled even by adequate information about the countervailing experience with human beings. The regulatory agency is therefore left in the position of having to deal with not only the available evidence but also the public's fear.

The U.S. Environmental Protection Agency has responded to the public's fear with a number of regulations intended to control the formation and release of TCDD and to limit individual exposure to it. Those regulations could be made stronger or weaker on the basis of new evidence. What the

DATE	WORKERS EXPOSED	LOCATION OF ACCIDENT	REMARKS
1949	250	Monsanto plant in Nitro, W.Va.	122 cases of chloracne being studied; 32 deaths v. 46.4 expected; no excess deaths from malignant neoplasms or circulatory disease
1953	75	BASF plant in Ludwigshafen	55 cases of chloracne, 42 severe; 17 deaths v. 11 to 25 expected (four gastrointestinal cancers and two oat-cell lung cancers); most common injuries were impaired senses and liver damage
1956	?	Rhone-Poulenc plant in Grenoble	17 cases of chloracne, also elevated lipid and cholesterol levels in the blood
1963	106	NV Philips plant in Amsterdam	44 chloracne cases (42 severe), of whom 21 also had internal damage or central-nervous-system disturbances; eight deaths (six possible myocardial infarctions); some symptoms of fatigue
1964	61	Dow Chemical plant in Midland, Mich.	49 cases of chloracne; four v. 7.8 expected deaths; three cancer deaths v. 1.5 expected, one a soft-tissue sarcoma
1965–69	78	Continuing leaks in Spolana plant near Prague	78 cases of chloracne; five deaths; many of the 50 workers studied for more than 10 years have hypertension, elevated blood levels of lipid and cholesterol, prediabetes; significant amounts of severe liver and neurologic damage
1966	?	Rhone-Poulenc plant in Grenoble	21 chloracne cases
1968	90	Coalite & Chemical plant in Derbyshire	79 chloracne cases; one death from coronary thrombosis
1976	156	ICMESA plant in Seveso, Italy	Workers are being studied along with exposed townspeople; more than 500 residents treated for presumed toxic symptoms; 134 confirmed chloracne cases; overall mortality rate normal

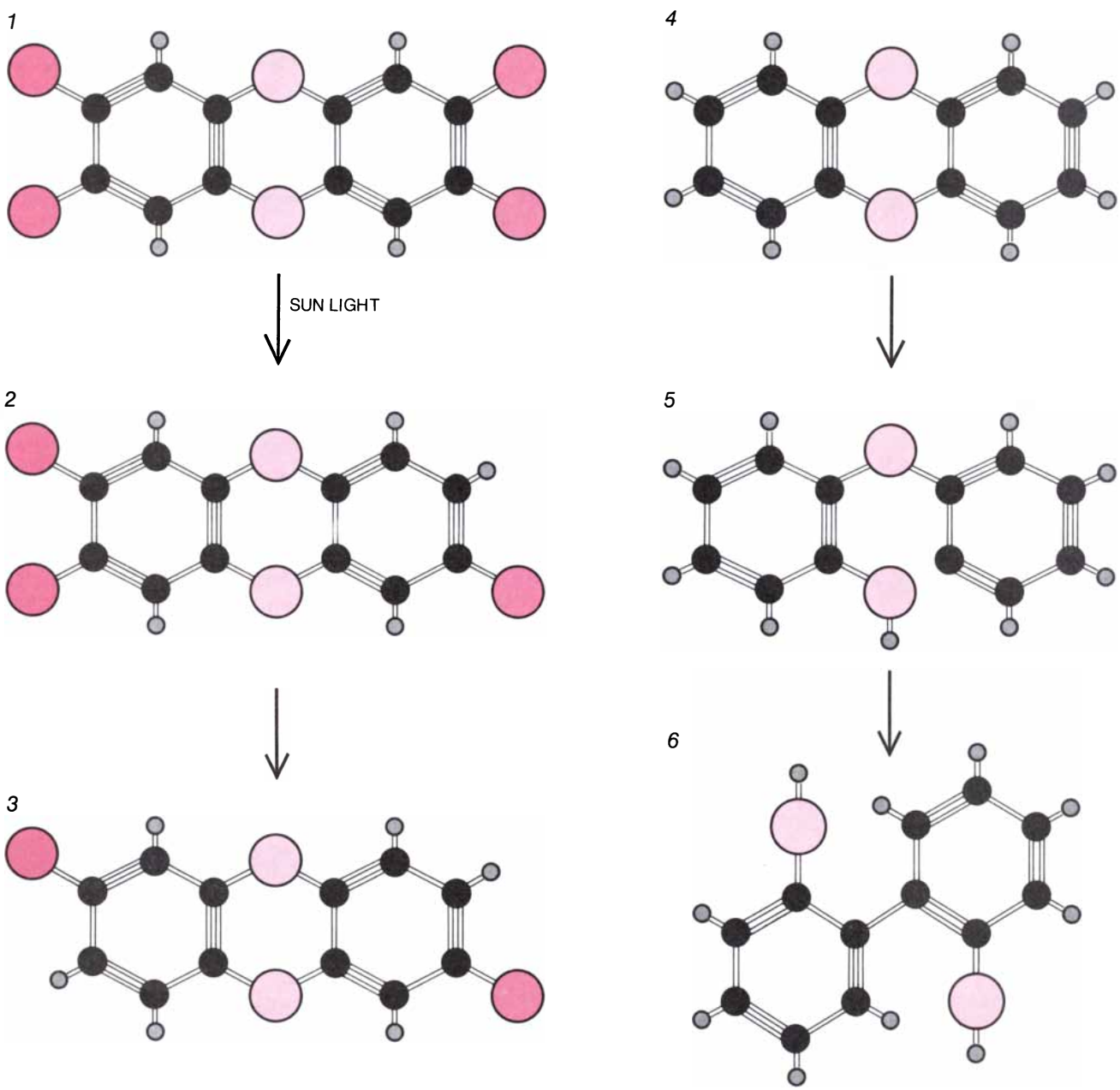
INDUSTRIAL ACCIDENTS have exposed more than 800 workers to significant amounts of TCDD. The accident at Seveso in 1976 also exposed some 37,000 residents of nearby communities. The data are based on a study made by the American Medical Association.

agency has not done—and might be said to have a responsibility to do—is to try to dispel the public's fear on the basis of the evidence that exposure to low concentrations of TCDD in the environment appears not to have serious chronic effects on human beings.

The TCDD case is further exacerbated by its relation to the defoliation program in Vietnam, an unpopular program in an unpopular war. The many and diverse health effects alleged by Vietnam veterans to have been caused by exposure to Agent Orange have been widely publicized. The public is generally aware that the complaints were settled out of court for $180 million, and many people believe the settlement was an admission of guilt by the chemical companies that manufactured Agent Orange. Apparently few people know of Federal Judge Jack B. Weinstein's statement to the attorneys for the plaintiffs that "in no case have you shown causality for the health effects alleged."

A troublesome matter exemplified by the TCDD issue is the appropriate utilization of scientific resources. A. L. Young of the Office of Science and Technology Policy has calculated that more than a billion dollars will have been spent by the Federal Government for research and other dioxin-related matters before all the major studies now in progress have been completed. Additional expenditures of both time and money have been made by chemical companies, private organizations and government agencies. The total outlay is a tremendous amount for an issue of questionable importance.

Two years ago a conference on dioxin at Michigan State concluded that the TCDD case is relatively less important than a number of other issues and that the nation's limited scientific resources should be devoted to the issues posing a greater threat. On the basis of the evidence turned up so far, the conclusion is still valid.



DECOMPOSITION OF TCDD takes place when the ultraviolet in sunlight splits off the molecule's chlorine atoms. Six steps in the reaction are shown. In soil, where sunlight cannot get at the material, TCDD tends to be quite persistent, enduring as long as 10 years.

The *T* Cell and Its Receptor

The cell plays a key role in the body's capacity to fight viral infection, but it also acts to reject grafted tissue. Experiments have now identified the molecule that underlies this behavior

by Philippa Marrack and John Kappler

The cells of the immune system are responsible for the ability of vertebrate animals to recognize that antigens, or foreign materials, have invaded their bodies. The immune response that follows is remarkable for its specificity. A person immunized by vaccination against smallpox, for example, can resist infection by the smallpox virus but not by, say, the influenza virus. In the past decade immunologists have come to recognize that the most important factor in the ability of the immune system to react specifically to viruses is a class of small cells called *T* lymphocytes, or *T* cells. The *T* cells also play an essential auxiliary role in the immune response to bacterial infection.

It has long been accepted that the trigger for this activity is a molecule embedded in the membrane of the *T* cell called the *T*-cell receptor. A specific antigen is assumed to fit and bind to the receptor as a key fits into a lock, thereby setting in motion the complex series of biochemical events that constitute the immune response. For a variety of reasons the isolation of the *T*-cell receptor has proved to be exceedingly difficult, and until quite recently its properties have had to be inferred indirectly, without the guidance afforded by a knowledge of its structure. That structure is now rapidly coming into focus.

The structural clarification of the *T*-cell receptor is bringing about a far better understanding of the complex interactions of *T* cells with other elements of the immune system. In particular, it is becoming apparent that the *T* cell is specially suited for dealing with infections associated with the cells of the host, rather than with infections that circulate freely in the host's bodily fluids. To carry out this role the *T*-cell receptor must not only recognize a specific antigen but also recognize certain membrane proteins of the host cell itself. Such a recognition mechanism must be kept under tight control, for if a *T* cell were to be activated by the host proteins alone, it could readily turn against the healthy cells of the organism. The consequent delicacy of the *T* cell's recognition system leads to its many fascinating and medically important properties. For example, *T* cells will rapidly act to reject foreign tissue that is surgically grafted or transplanted into the body. Hence the investigation of the *T*-cell receptor is an issue of major interest to surgery.

There are actually two kinds of lymphocyte responsible for the recognition of specific antigens: the second kind is the *B* cell. Both the *B* cell and the *T* cell are derived from the bone marrow, but the *T* cell undergoes further development in the thymus gland, just under the upper part of the breastbone in man. Both *B* and *T* cells circulate in the blood and the lymph and are concentrated in the major lymphatic organs: the lymph nodes and the spleen in higher vertebrates. They can be quite long-lived; in man they can persist for many years without dividing. In response to an antigen, however, lymphocytes enlarge considerably, divide rapidly and secrete a number of protein factors that contribute to the elimination of the invading organism or foreign material.

It has been known for some time that the initial response of the *B* cell to an antigen is mediated by a receptor protein displayed on the cell membrane. The antigen binds to the receptor, and the binding leads the *B* cell to divide and differentiate into a clone of plasma cells. These cells secrete antibodies that have the same antigen-binding properties as the receptor molecules embedded in the surface of the parent *B* cell. Indeed, the antibody is identical with the *B*-cell receptor to which the antigen was originally bound, except that the end of the chain of amino acids anchoring the receptor protein in the membrane of the *B* cell is not found on the soluble antibody. Both *B*-cell receptors and antibodies are also called immunoglobulins.

Once the antibodies are secreted into the blood or the lymph, they bind to free antigen and mark it for destruction by other components of the immune system. This general picture of how a *B* cell is "selected" by an antigen for the capacity of the cell to clonally expand and secrete antibody to the antigen is called the clonal selection theory; it was developed in the 1960's by Sir Macfarlane Burnet of the Walter and Eliza Hall Institute for Medical Research in Melbourne, David W. Talmage, then at the University of Chicago, and Niels Kaj Jerne, then at the World Health Organization.

In their role as *B*-cell receptors the antibodies are found in quite small quantities. When the *B* cell is challenged with antigen, however, the antibodies appear in serum in large and soluble quantities. Antibodies are also secreted at high levels by certain types of *B*-cell tumor such as plasmacytomas. The ready availability of such high concentrations of soluble antibody and the fact that each antibody molecule can bind to antigen made it possible to isolate antibodies relatively easily and learn a great deal about the structure of the *B*-cell receptor.

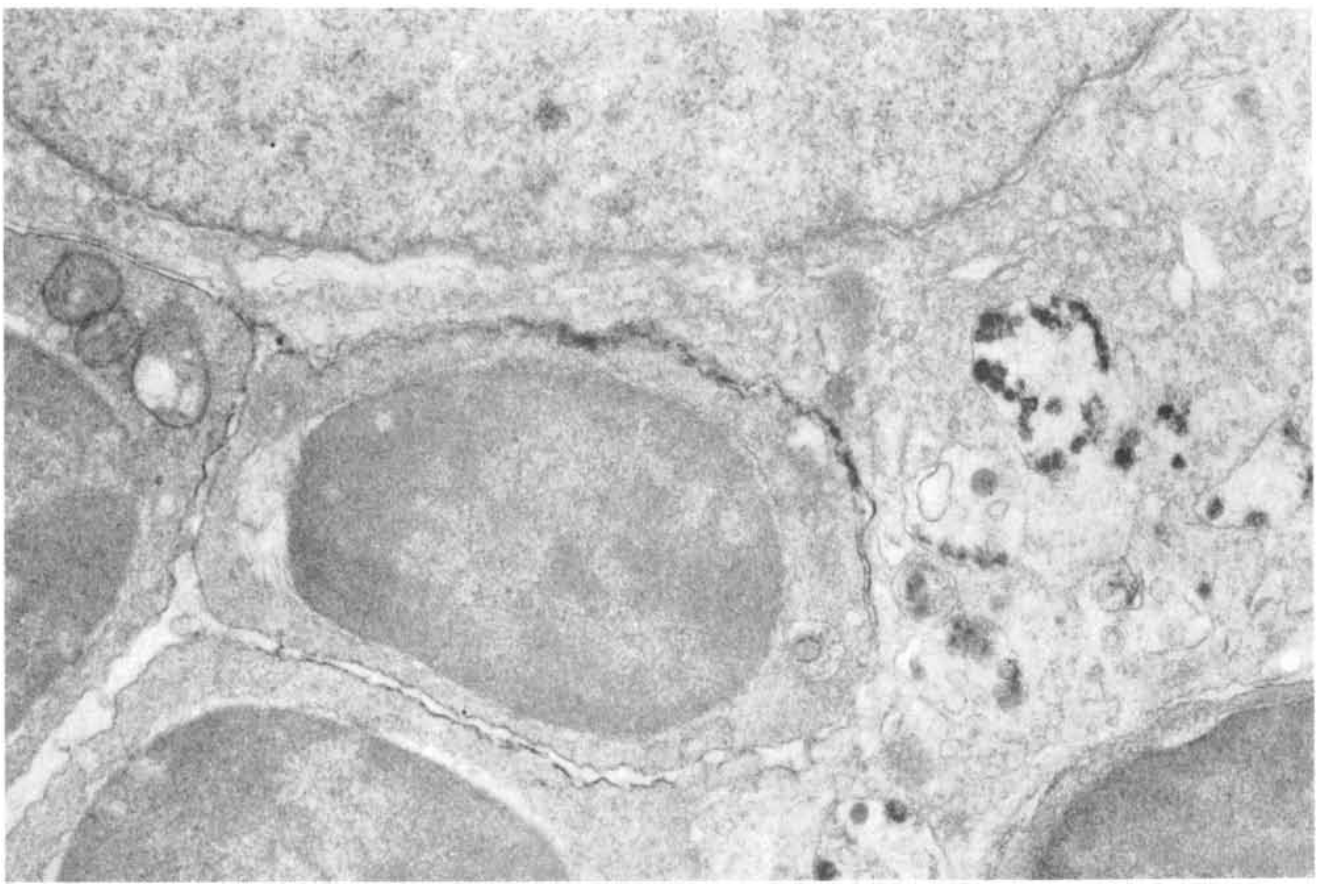
The *T*-cell receptor has been much more elusive. The *T* cell, like the *B* cell, responds to an antigen by clonally dividing and differentiating into one of several kinds of *T* cell specific to the antigen. Cytotoxic *T* cells bind to viral antigen displayed on the surface of an infected cell and kill the cell. Suppressor *T* cells act to inhibit the immune response to an antigen some time after the response has been set in motion. Helper *T* cells bind to antigen on the surface of a *B* cell that has already

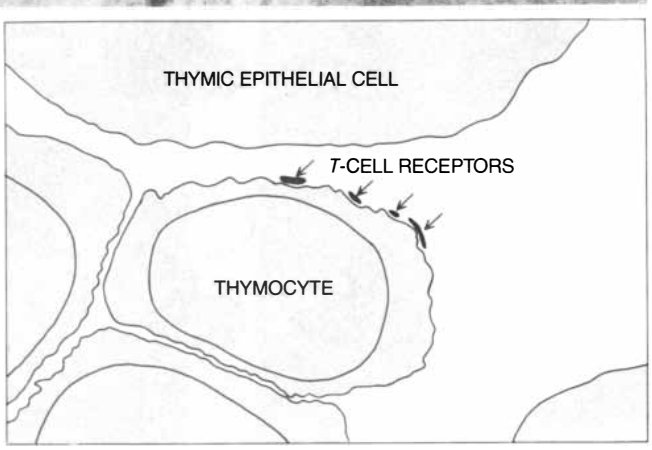
bound itself to the antigen. Each helper *T* cell then releases hormonelike molecules called lymphokines that enable the *B* cell to multiply and differentiate. Thus there is a two-key system for releasing the enormous destructive potential of a *B* cell: one key (the free antigen) for the *B*-cell receptor and the other key (the antigen on the surface of the *B* cell) for the *T*-cell receptor.

T cells themselves never differentiate into cells that secrete antibodies. Hence unlike the *B*-cell receptor, the *T*-cell receptor is not readily available in the quantities of purified, soluble chemical that are needed for convenient analysis. Because antibodies are so elegantly constructed and so efficient in their capacity to recognize antigens, it was assumed for many years that the *T* cell would rely on the same molecules as the *B* cell does in order to bind and respond to antigen.

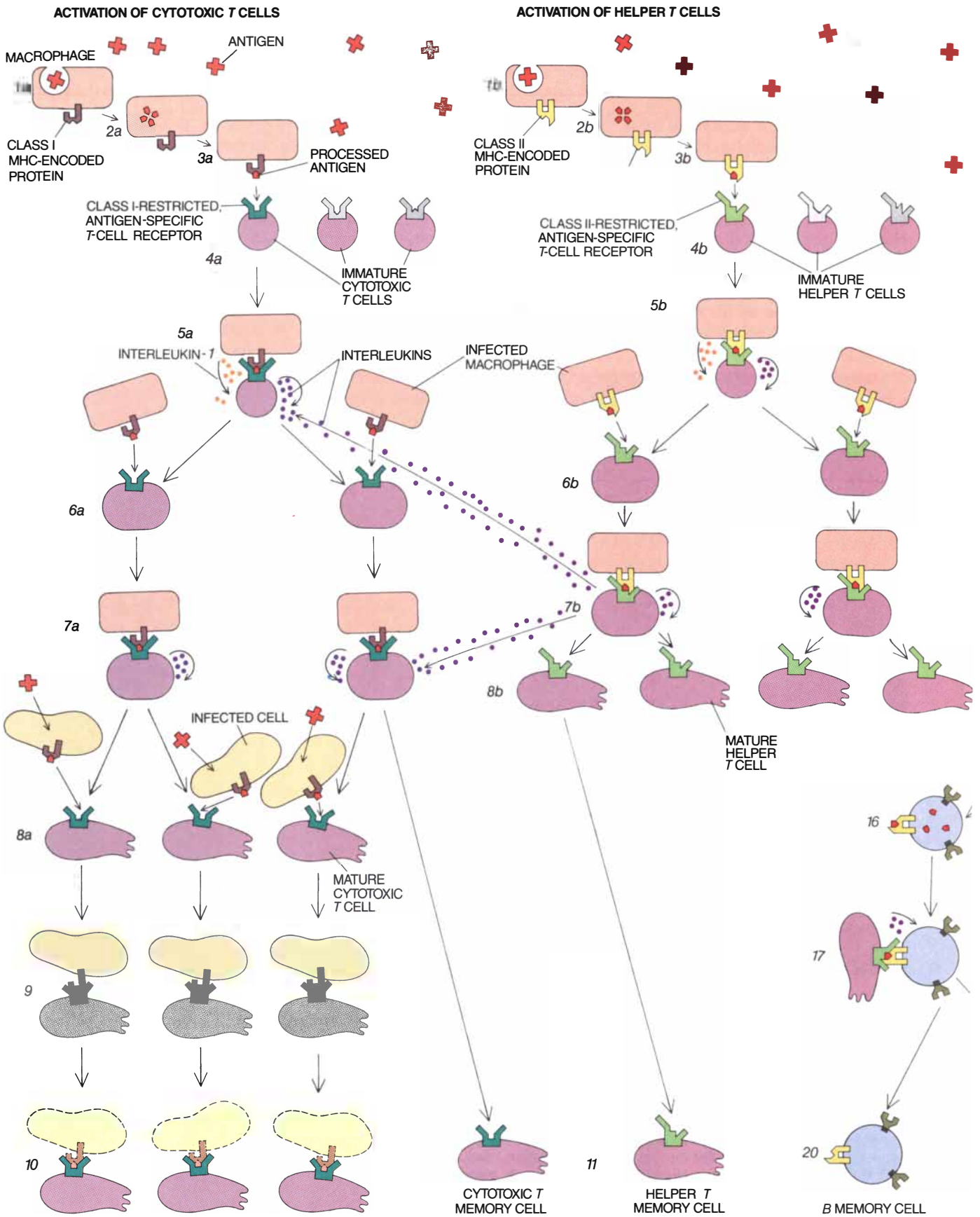
Many investigators spent years examining the surfaces of *T* cells and their secretions for immunoglobulins. Although extensive searches have suggested that suppressor *T* cells may bind antigen with molecules similar to immunoglobulin, many experiments have shown that most *T* cells are not associated with immunoglobulins. Not only are immunoglobulins not secreted by *T* cells but also they are not found either on the membranes of *T* cells or in their cytoplasm. Other experiments show that *T* cells do not express messenger RNA transcribed from immunoglobulin genes; moreover, when one examines the immunoglobulin genes in the *T* cell, one finds that the *T* cell does not usually rearrange them the way the *B* cell does.

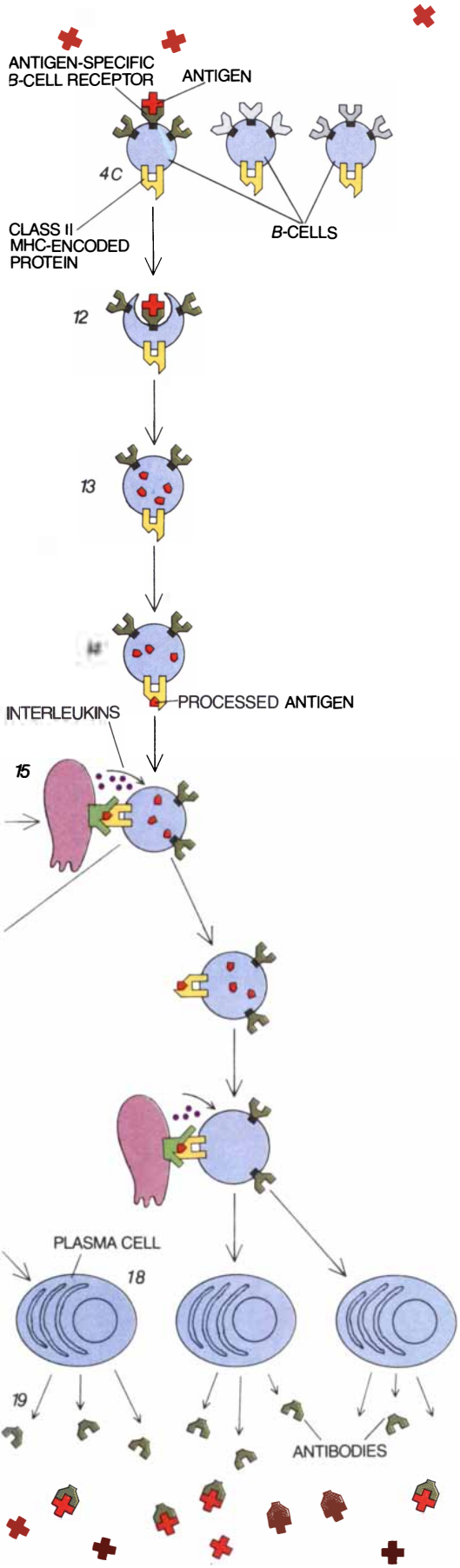
Such negative results accumulated in the 1970's and in the early 1980's. Although the results of such experiments are not usually very persuasive when they are considered one at a time, the sheer number of failures gradually led molecular immunologists to change their approach. If immunoglobulins could not be found in association with *T* cells, one would





"EDUCATION" OF A *T* CELL in the thymus confers on the cell the ability to distinguish other cells in the animal from "nonself" cells. That ability is a critical element in the specificity of the immune response. The thymocyte, or immature *T* cell, in the photomicrograph is bound to an epithelial cell that makes up part of the lining of the thymus (*see map at right*). The dark binding sites in the photomicrograph are receptor molecules on the surface of the thymocyte, which are probably bound to "self"-defining proteins displayed on the epithelial cell. The proteins are encoded in the genome by a region of DNA called the major histocompatibility complex (MHC). The binding may stimulate the education, or maturation, of the thymocyte. Mature *T* cells acquire a low affinity for self-MHC-encoded proteins, but they have a high affinity for such proteins in association with antigens. The *T* cells trigger an immune response only to infected cells of the self or to nonself cells from another animal. The photomicrograph was made by Andrew G. Farr of the University of Washington; magnification is 20,900 diameters.





have to look elsewhere for the *T*-cell receptor for antigen.

The study of foreign tissue grafts and their rejection by a host animal had led cellular immunologists to the same conclusion some years earlier. For example, mouse cells taken from one strain of mice are rejected rapidly after being injected into a mouse of a different genetic strain. Beginning in the early 1930's Peter Gorer and other workers showed that such rejection is caused by antigenic molecules on the surfaces of the foreign mouse cells.

The proteins that mark every cell as "foreign" or "self" are encoded by genes linked close together in a region of DNA called the major histocompatibility complex (MHC), after the prefix "histo-," meaning tissue. The proteins themselves are called MHC-encoded proteins. One of their most remarkable properties is their extreme polymorphism: there are millions of alleles, or variants, for the MHC genes encoding each protein. Hence the likelihood that two unrelated individuals have identical MHC-encoded proteins is small.

Following the work of Gorer it became increasingly clear that graft rejection and the MHC-encoded proteins are closely related to the immune response. Sir Peter Medawar and others demonstrated that lymphocytes are responsible for recognizing the antigenic molecules on a graft of foreign tissue; later work showed that the lymphocytes central to the graft rejection are *T* cells rather than *B* cells. Graft transplantation is not an experiment of nature, however, and so immunologists were still in doubt about the functions of the MHC-encoded proteins.

Some clues about such functions began to emerge in the mid-1960's. Hugh O. McDevitt, then at the National Institute for Medical Research in England and Michael Sela of the Weizmann Institute of Science in Israel, and Baruj Benacerraf, then at the National Institute of Allergy and Infectious Diseases, and his colleagues, studied the response of various strains of animals to antigens. They found that the response of a mouse to certain synthetic antigens can depend on the genetic strain of the mouse. For example, when the polymer known as TGAL is injected into mice, the mice bearing *b* alleles in their MHC make antibodies to the polymer, but the mice bearing *k* alleles in their MHC do not. It soon became apparent that these genes act by affecting the functions of *T* cells and not some other cell type in the immune system.

Some immunologists then suggested the *T*-cell receptor might be encoded by the MHC. If it were, one would naturally expect different kinds of MHC to encode different *T*-cell receptors, and the latter would account for differences in the response to TGAL from one mouse strain to another. The theory had to be discarded when it was discovered that some of the MHC-encoded proteins that affect the binding of TGAL are not even expressed on the surface of the *T* cell in mice. Hence the source of the observed differences could not be the *T*-cell receptor alone, and there was no further point in supposing the receptor was encoded by

ANTIGEN-SPECIFIC RESPONSE of the immune system is portrayed in this highly schematic diagram. Three major kinds of response are shown: that of the cytotoxic *T* cell (*left*), that of the helper *T* cell (*middle*) and that of the *B* cell (*right*). Antigen that invades the body (*top*) is engulfed by a macrophage (*1a, 1b*), processed, or digested, into shorter pieces of protein (*2a, 2b*) and presented in partially digested form on the surface of the macrophage (*3a, 3b*). There the antigen is bound to one of two classes of MHC-encoded proteins. A *T* cell whose receptor fits the antigen-protein complex binds to it; the *T* cell is thereby "selected" for clonal propagation by the antigen-protein complex (*4a, 4b*). Similarly, free antigen selects a *B* cell whose receptor fits the antigen and the two become bound (*4c*). *T* cells destined to become cytotoxic bind antigen in association with class I MHC-encoded proteins (*5a*), whereas future helper *T* cells bind antigen in association with class II proteins (*5b*). The *T*-cell bond stimulates the release of the hormonelike molecule interleukin-1 by the macrophage, and that in turn stimulates the *T* cell to divide and differentiate (*6a, 6b*). Cell division continues as long as it is stimulated by cells that present surface antigen (*7a, 7b*). A mature *T* cell (*8a, 8b*) can then fulfill its role. If it is cytotoxic, it can either bind to an antigen-presenting infected cell (*9*) and kill the cell (*10*), or it can remain circulating in the blood and lymph as a memory cell that deals more rapidly with any future antigen of the same kind (*11*); a mature helper *T* cell can also become a memory cell. The function of the helper *T* cell is to stimulate the proliferation of activated *B* cells. A *B* cell, having engulfed its bound antigen (*12*) and processed it (*13*), also presents a piece of antigen on its surface bound to a class II MHC-encoded protein (*14*). The mature helper *T* cell can then bind to the antigen-protein complex on the *B* cell (*15*). The bond releases interleukins, which enable the *B* cell to divide and differentiate (*16*); the cell division continues as long as it is stimulated by helper *T* cells (*17*). Mature plasma cells (*18*) then release their antigen-specific receptors as antibodies, which bind to free antigen and mark it for destruction (*19*). Other mature *B* cells in the clone remain in circulation as memory cells (*20*). The diagram incorporates the results of the experiments described here by the authors as well as the results of many other experiments, but certain details of the model are still controversial.

the MHC. The experiment did lend support to a subtler point: the differences in the antibody response to TGAL indicated that MHC-encoded proteins affect the way TGAL is recognized by the *T*-cell receptor. This finding was the first clue that some interaction between foreign antigens and MHC-encoded proteins was a prerequisite for the action of the receptor.

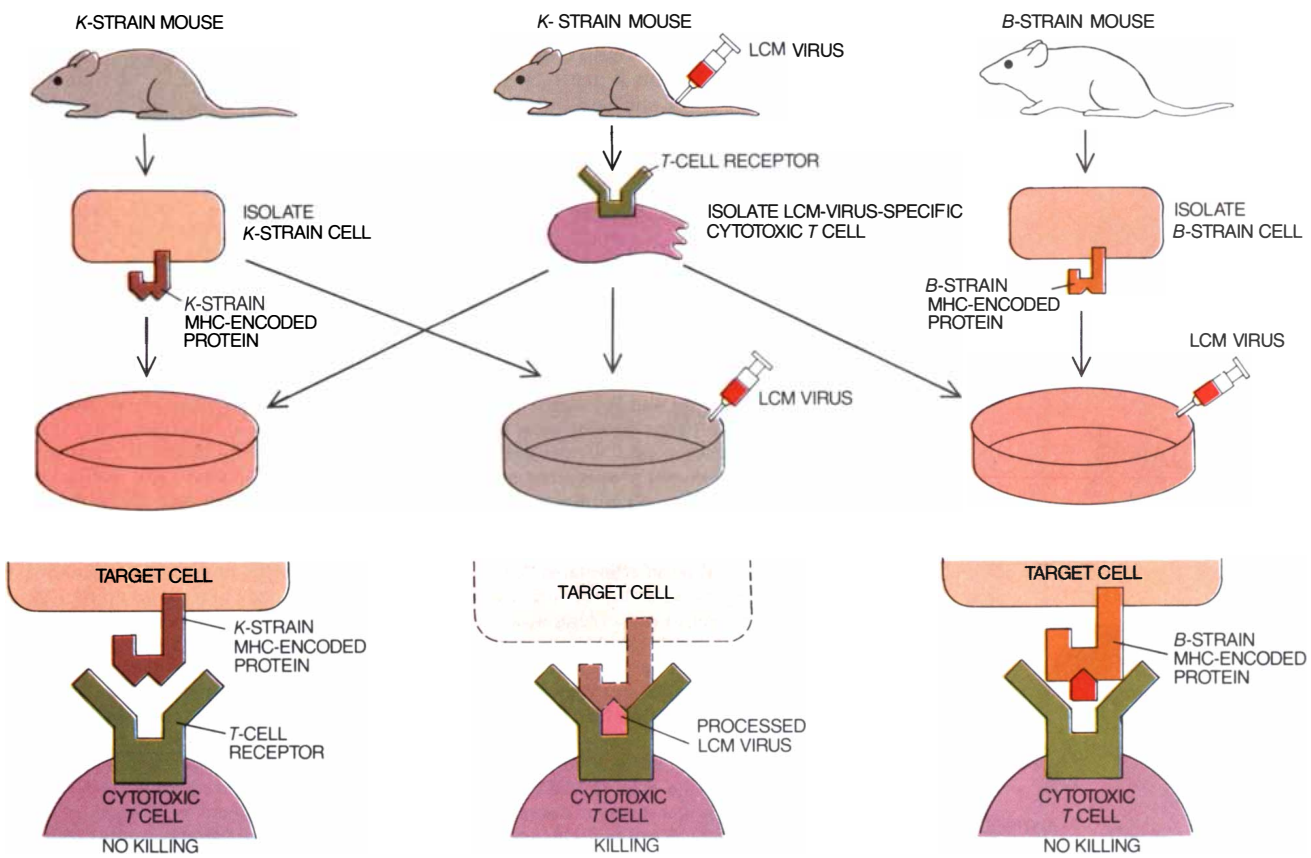
How could one account for such a strange collection of results? The answers came from the pioneering work of Ethan M. Shevach and Alan S. Rosenthal at the National Institute of Allergy and Infectious Diseases, David H. Katz and Benacerraf at the Harvard Medical School, Bernice Kindred at the University of Constance in West Germany, Donald Shreffler of the University of Michigan School of Medicine and, most clearly, from the remarkable experiments of Rolf Zinkernagel and Peter C. Doherty of the Australian National University.

In one experiment Zinkernagel and Doherty injected sublethal doses of a virus known as LCM (for lymphocytic choriomeningitis) into mice and isolated LCM-specific, cytotoxic *T* cells from the immunized animals [*see illustration below*]. The usual function of such *T* cells is to recognize viral antigens on infected cells and destroy the cells. Most strains of mice make antibodies to LCM virus, and so there was little chance that LCM virus would fail to induce an immune response in some strains of mice the way TGAL does. Nevertheless, Zinkernagel and Doherty found to their surprise that the *T* cells from *k*-strain, LCM-immunized mice are able to kill cells infected with the LCM virus only if the infected cells bear at least one of several kinds of *k*-strain, MHC-encoded proteins. In other words, the *T* cells that were lethal to infected cells bearing certain *k*-strain, MHC-encoded proteins could not kill cells from closely related mice, infected with the same virus but bearing, say, the *d* strain of the same MHC-encoded proteins.

This experiment and others demonstrated that cytotoxic *T* cells were paying attention not only to viral antigens on the infected cell but also to the amino acid sequences of MHC-encoded proteins on the same infected cell. Other investigators quickly showed that the same was true for cytotoxic *T* cells specific for other antigens. The phenomenon is now called MHC restriction. In general the MHC-encoded proteins recognized by the cytotoxic *T* cells belong to a single structural class called the class I proteins; they are found on the surfaces of all nucleated cells in the body.

We and several other investigators then found that helper *T* cells are also MHC-restricted; in general, however, the helper *T* cell recognizes a different class of MHC-encoded proteins called the class II proteins. The expression of class II proteins is considerably less widespread in the body than the expression of class I proteins: in people the class II proteins are found only on *T* cells, *B* cells, macrophages and certain cells of other tissues.

As soon as MHC restriction was discovered it was clear that two kinds of theory could account for the facts. One, the associated-recognition theory, suggests in its extremest form that each *T* cell bears receptors of a single kind. Each such receptor would somehow bind a complex of the antigen and a particular MHC-encoded protein that appear on the surface of



T CELL responds to antigen only if the cell binds both the antigen and an MHC-encoded protein on the target cell. *T* cells taken from a *k*-strain mouse immunized with lymphocytic choriomeningitis (LCM) virus can kill cells bearing *k*-strain, MHC-encoded proteins after the cells have been infected with LCM virus (*middle*). The *T* cells cannot kill cells infected with the virus that do not bear the *k*-strain, MHC-encoded proteins (*right*), and of course they do not kill uninfected mouse cells that bear the *k*-strain proteins (*left*).

the antigen-presenting cell. The second theory, which is called the dual-recognition theory, suggests in its extremest form that each *T* cell bears two kinds of receptor: one would bind a specific antigen and the other would bind a specific MHC-encoded protein.

There are advantages and disadvantages to each kind of theory. The associated-recognition theory accounts for the observation that *T* cells only rarely bind to antigen or to MHC-encoded proteins without binding to both molecules at the same time. Moreover, the theory accounts nicely for the findings of McDevitt, Sela and Benacerraf on the immunogenic properties of TGAL in mice. If a single *T*-cell receptor must simultaneously bind a complex of antigen and MHC-encoded protein, the failure of *k*-strain mice to respond to TGAL might be caused by the failure of the TGAL antigen to form a complex with the *k*-strain, MHC-encoded protein. On the other hand, the dual-recognition theory does not share a disadvantage of the associated-recognition theory: it need not postulate an interaction between every form of antigen and the invariant forms of MHC-encoded protein that are found in any individual animal.

Although the final word is not yet in, it is probably fair to say that the current data favor the associated-recognition theory, with its model of a single *T*-cell receptor for antigen and MHC-encoded protein. For example, we have studied the properties of hybrid *T* cells that express *T*-cell receptors inherited from two different parents [*see illustration on next page*]. We first isolated helper *T* cells that specifically recognize an antigen, chicken ovalbumin (cOVA), when it is associated with certain *k*-strain, class II proteins. We fused the *T* cells to a *T*-cell tumor in order to make a *T*-cell hybridoma, a hybrid cell that grows rapidly and clones readily in tissue culture.

In order to construct a cell having two different sets of *T*-cell receptors we fused one of our *T*-cell-hybridoma cells with a second set of antigen-specific *T* cells. The second set of *T* cells could respond only to the keyhole-limpet hemocyanin (KLH) antigen in the presence of *f*-strain, class II proteins.

We tested the response of the newly fused *T* cells to cOVA or KLH antigens in association with *k*-strain or *f*-strain antigen-presenting cells. If the dual-recognition theory were correct, one would expect that the newly fused *T* cells would bear four types of receptor, one for each antigen and one for each class II, MHC-encoded protein. Hence they would respond to both antigens presented on a *k*-strain cell and to both antigens presented on an *f*-strain cell. Instead we found they would respond to the antigens only as their parents did: to cOVA associated with *k*-strain cells and to KLH associated with *f*-strain cells. There was no sign of a response to cOVA on *f*-strain cells or to KLH on *k*-strain cells.

If the associated-recognition theory is correct, the antigen and the MHC-encoded protein must somehow form a complex before a *T* cell can bind to them. The results of several experiments suggest there is some kind of interaction between MHC-encoded proteins and antigens, but only a few experiments have been able to detect the interaction directly. Perhaps the most striking recent demonstration of such an interaction comes from the work of Emil R. Unanue and his colleagues at Washington University School of Medicine.

Unanue and his colleagues identified a small antigen that can be recognized by *T* cells when it is associated with *k*-strain, MHC-encoded proteins but not when it is associated with *d*-strain proteins. They placed equal concentrations of the antigen in solution on both sides of a semipermeable membrane, and they placed one or the other strain of MHC-encoded proteins on one side of the membrane. Because the antigen molecule is significantly smaller than either kind of protein, the membrane allowed free passage of the antigens but restricted the proteins to their initial compartment.

Unanue and his group found that when *k*-strain, MHC-encoded proteins were added to one compartment and the antigen concentration was allowed to equilibrate, the concentration of the antigen became higher in the compartment holding the *k*-strain proteins. The concentrations of the antigen in the two compartments did not change when the *d*-strain, MHC-encoded proteins were added to one compartment. The result suggests that antigen does bind, at least sometimes, to the MHC-encoded protein with which it is recognized, although the process may not take place very efficiently.

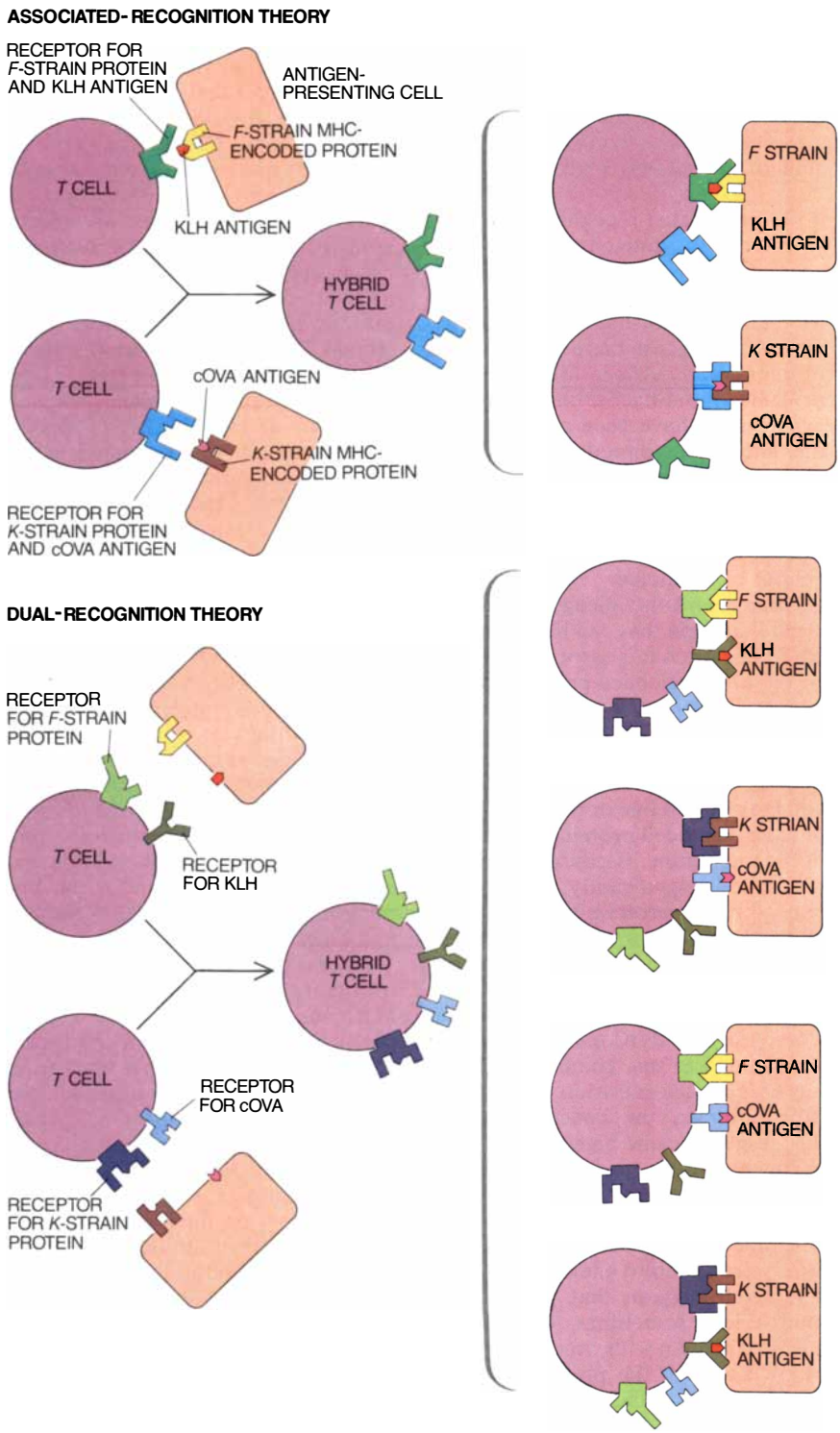
Even after all the foregoing properties of the *T*-cell receptor had been established, the identity of the protein responsible for its activity was still a mystery. It was clear by then, however, that the *T*-cell receptor differs in at least one important way from the antibody molecule. Whether the *T*-cell receptor is one distinct protein or two, at least some component of the receptor has a strong tendency to recognize part of an MHC-encoded protein. The antibody molecule does not.

There were two major technological breakthroughs that enabled investigators to begin to discern the structure of the *T*-cell receptor. One was a discovery made in the early 1970's by George Köhler and Cesar Milstein at the Medical Research Council's Laboratory of Molecular Biology in Cambridge; the two workers found a way to produce *B*-cell hybridomas that can secrete large quantities of selected kinds of antibody in vitro. Such antibodies are called monoclonal antibodies, and they can be produced and purified in enormous quantities. The second breakthrough was the development of methods whereby *T*-cell clones or *T*-cell hybridomas for specific antigens and specific MHC-encoded proteins could be propagated in culture. We and a number of other investigators have exploited both breakthroughs to create antibodies to *T*-cell receptors and thereby identify structural properties of the *T*-cell receptor.

To make antibodies to a *T*-cell receptor we first built a *T*-cell hybridoma that bore a receptor for recognizing cOVA antigen in association with a *d*-strain, class II protein. When the receptors on such hybridomas are engaged with the right antigen and the right MHC-encoded protein, they respond rapidly by secreting lymphokines. Such hybridomas are convenient experimentally because the production of lymphokines can be induced and quickly measured in culture.

To make an antibody of known specificity one must immunize an animal against a known antigen, but in the case of the *T*-cell receptor the precise identity of the antigen was not known. Our strategy was to immunize mice against the *T*-cell hybridomas, in the hope of inducing the production of antibodies that would interfere only with functions presumably carried out by the receptors on the hybridomas. We reasoned that an antibody to the receptor would bind to the receptor and block its ability to engage cOVA and *d*-strain, MHC-encoded protein. A decrease in the production of lymphokines would indicate the blockage.

We immunized many mice against large numbers of *T*-cell hybridomas. We then drew serum from each mouse at various times after the immunizations and tested the serums in culture for their ability to block the response of the *T*-cell hybridomas to *d*-strain cells and cOVA. Eventually we identified several mice that made the blocking antibody. The antiserums of these mice had an additional encouraging property: when *T*-cell hybridomas of a different specificity were challenged in culture with the antigen and MHC-encoded protein to which they were specific, the mouse antiserums did not



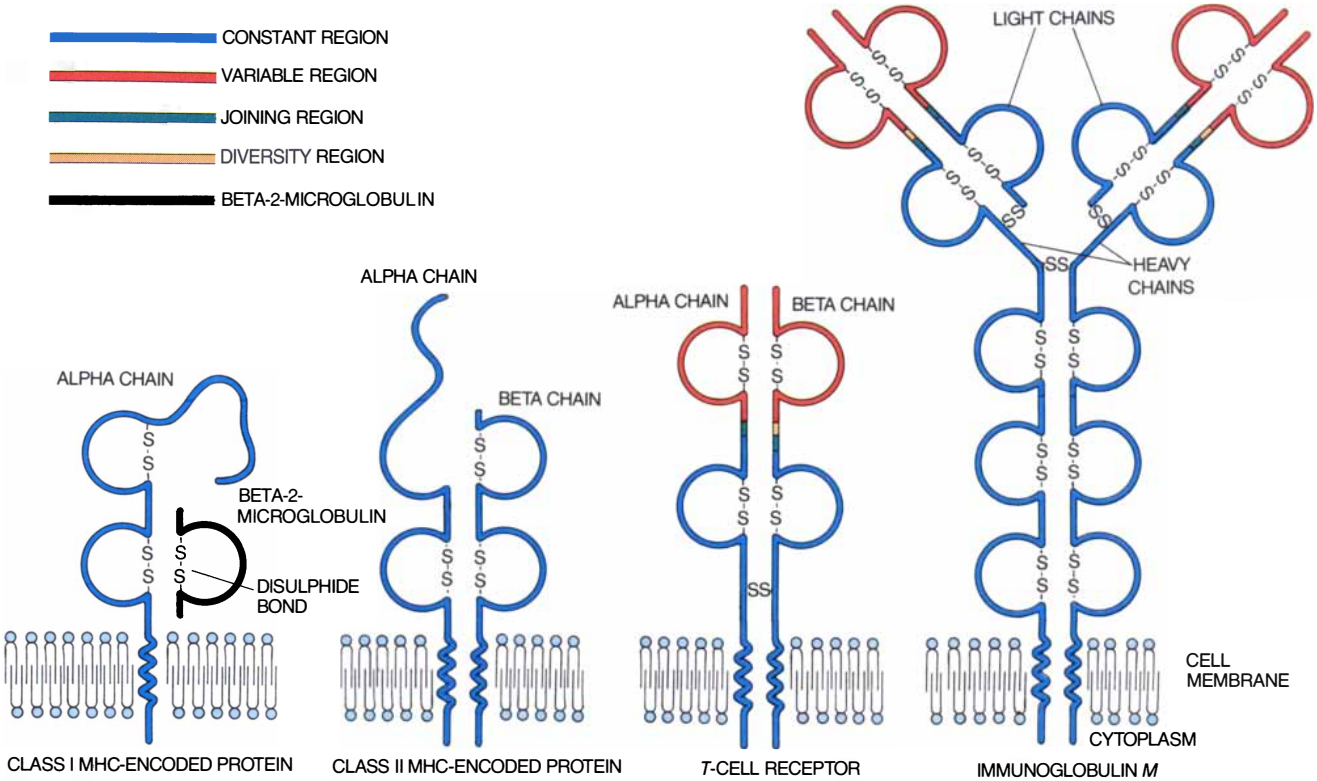
TWO INCOMPATIBLE THEORIES can explain the recognition of both antigen and MHC-encoded proteins by *T* cells. The associated-recognition theory asserts that each *T* cell bears a single receptor, which can bind a combination of antigen and MHC-encoded protein (*upper left*). The dual-recognition theory asserts that each *T* cell bears two different receptors, one receptor for antigen and one for the MHC-encoded protein (*lower left*). To test the theories the authors constructed a hybrid *T* cell by fusing two normal *T* cells. One parent *T* cell was isolated from animals having *f*-strain, MHC-encoded proteins, after the animals were immunized against the antigen keyhole-limpet hemocyanin (KLH). The second parent *T* cell was isolated from *k*-strain animals immunized against the antigen chicken ovalbumin (cOVA). According to an extreme version of the associated-recognition theory, the hybrid cell should recognize KLH only on *f*-strain cells and cOVA only on *k*-strain cells (*upper right*). An extreme version of the dual-recognition theory asserts the hybrid should recognize KLH or cOVA on both strains of cell (*lower right*). Actually the hybrid cell behaved in accordance with the assertion made by the associated-recognition theory.

block the *T*-cell response. We repeated the experiment with other *T*-cell hybridomas in the role of the original hybridoma. In each case the blocking antibodies developed by the mice were effective only against the *T*-cell hybridoma used for the immunization.

The specificity of such antiserums for a particular *T*-cell hybridoma led us to believe we had indeed produced antibodies to a particular *T*-cell receptor. The *T*-cell receptor is the only structure on the surface of the *T*-cell hybridoma that one would expect to be blocked by an antibody in such a specific way, because it alone should vary from one *T*-cell clone to another. Nevertheless, our task was not yet finished. Because only small amounts of the antibodies could be isolated from each mouse, we applied the method of Köhler and Milstein to immortalize the antibody-secreting cells.

Kathryn Haskins and Janice White of our laboratory extracted plasma *B* cells from one of the mice immune to the cOVA *T*-cell hybridoma. The plasma cells were fused with tumor cells to make plasma-cell hybridomas that would readily grow in culture. The plasma-cell hybridomas secreted antibodies, which were then screened for their ability to block the recognition of cOVA and *d*-strain, class II protein by the cOVA *T*-cell hybridoma. One plasma-cell hybridoma had this property. Similar antibodies were obtained at about the same time by James P. Allison, then at the University of Texas Cancer Center in Smithville, Tex., and by Stefan C. W. Meuer and Ellis Reinherz of the Harvard Medical School.

It is primarily the successful propagation of antibodies for specific *T*-cell receptors that has made it possible to build up a picture of the *T*-cell-receptor protein. The tight bond between the receptor and the antibody has enabled investigators to purify the receptor in large enough quantities to reveal its basic molecular properties. Perhaps surprisingly, it turns out there is a close structural resemblance between the antibody molecule and the *T*-cell receptor. Both are made up of two polypeptide chains encoded by distinct genes in the DNA and held together by strong, covalent bonds connecting two sulfur atoms [*see illustration on opposite page*]. In antibodies the two chains differ in size; they are called the heavy chain and the light chain. Each chain includes a sequence of amino acids that is relatively constant within an animal, even for antibodies that bind to distinct antigens. In addition each chain has a long stretch of amino acids that varies considerably for antibodies of different antigen specificity.



MOLECULAR STRUCTURES of the class I MHC-encoded protein, the class II protein, the *T*-cell receptor and the immunoglobulin, or antibody, molecule are similar, and the molecules also share similar sequences of amino acids. The molecules are characterized by loops made up of about 70 amino acids within each chain; sulfur atoms at each end of the loop are joined by covalent bonds. Class I proteins are expressed on the surface of every nucleated cell in higher vertebrates, in association with the non-MHC-encoded protein beta-2-microglobulin. Class II proteins are expressed only on the surface of selected cells, such as the *B* cells. Variable, joining and diversity regions have been identified in the *T*-cell receptor and in the immunoglobulin. Each such region is encoded by a sequence of DNA selected at random from among a number of distinct sequences in the genome. Such regions give rise to substantial combinatorial variability in the molecules expressed by different cells in an animal. The highly schematic diagrams are not drawn to scale.

The variability of antibody within an animal arises from the genome. The basic idea is that each polypeptide chain of the antibody is made up of three or four regions, each of which can be encoded by one of several randomly selected pieces of DNA. The combinatorial variability arising from the construction leads to a large number of distinct antibodies.

For example, a heavy chain is made up of four regions of amino acids: the constant region, a joining region, a diversity region and the variable region. The constant region is encoded by only one piece of DNA, but the strand of DNA coding for the joining region can be drawn at random from one of four distinct segments. Similarly, the diversity region is encoded by one of more than 10 pieces of DNA and the variable region is encoded by one of more than 100. The total number of different amino acid sequences that can arise from the different combinations is thus more than $4 \times 10 \times 100$, or 4,000. The light chain is also made up of different combinations of DNA, and additional variability arises because of imprecise points of contact between the joining, diversity and variable regions. According to Susumu Tonegawa of the Massachusetts Institute of Technology, the number of distinct antibody molecules may be as great as one billion [see "The Molecules of the Immune System," by Susumu Tonegawa; SCIENTIFIC AMERICAN, October, 1985].

In the *T*-cell receptor the two chains are called the alpha chain and the beta chain. In the mouse both have a molecular weight of about 43,000 atomic mass units (a.m.u.); in people the alpha chain weighs about 50,000 a.m.u. and the beta chain weighs about 39,000. By comparing the alpha chains and the beta chains from different *T*-cell clones with one another we found that certain fixed sequences of amino acids appear on each chain from clone to clone. Other sequences can vary from clone to clone.

With the discovery of the proteins that make up the *T*-cell receptor, the techniques of molecular biology could be called into play. Such techniques enable workers to analyze the genes that encode a protein of interest, and that analysis can disclose the structure of a protein much faster than biochemical methods can. Moreover, given the similar roles of *T*-cell receptors and immunoglobulins in the immune system and the emerging structural similarities between them, it seemed likely that the segments of DNA encoding the *T*-cell receptor would be rearranged before its constituent proteins are expressed on the surface of the *T* cell, much as the segments of DNA encoding the *B*-cell receptor are. The race was soon on to find the *T*-cell-receptor genes.

The first serious candidates for the genes were reported simultaneously by two groups: Stephen M. Hedrick, Mark M. Davis and their collaborators at the National Institutes of Health, the University of California at San Diego and Stanford University, and Tak W. Mak and his group at the Ontario Cancer Institute. Both groups, the first group working with mouse genes and the second working with human genes, reasoned that *T*-cell-receptor proteins would be found only in *T* cells and not, for example, in *B* cells. They adopted elegant experimental techniques to exploit this assumption,

and they were soon able to identify genes expressed only in *B* cells or only in *T* cells but not in both kinds of lymphocyte. The genes encoding the beta chain of the receptor were identified first, and within another year the genes for the alpha chain were isolated as well. Much detail about the structure of the *T*-cell receptor has been derived from the analysis of these genes.

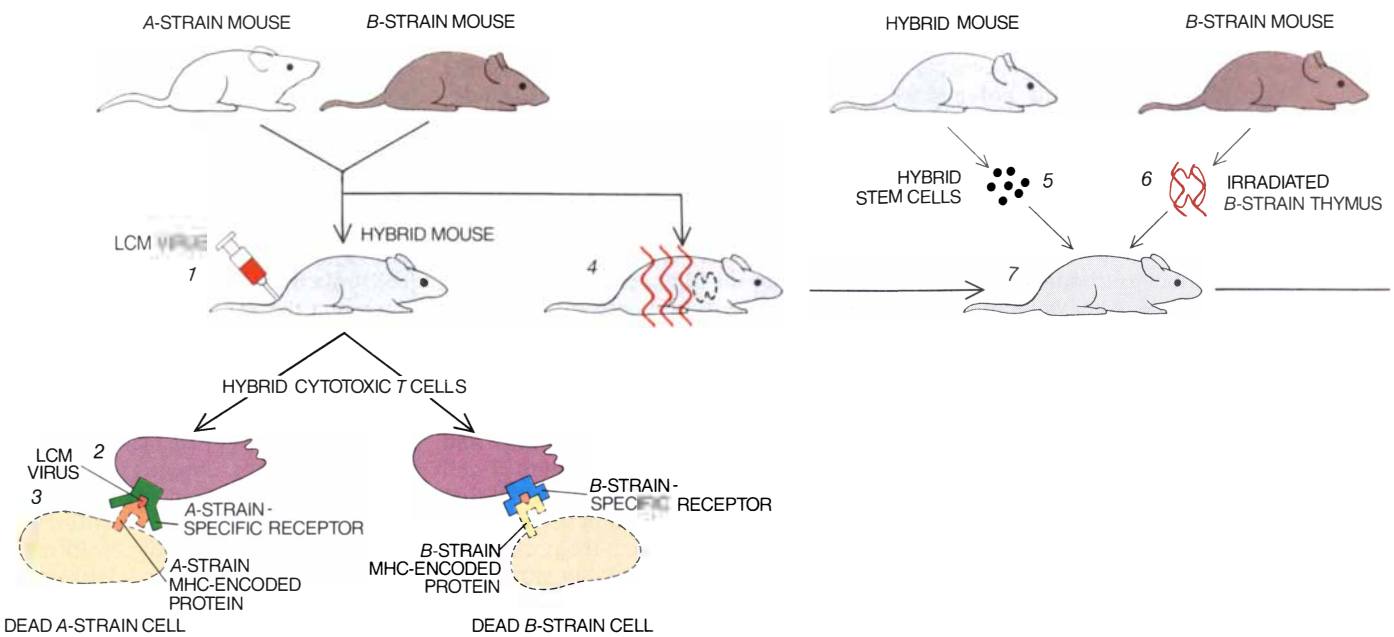
Both the alpha and the beta chain of the *T*-cell receptor have variable, constant and joining regions. In addition investigators have confirmed a diversity region in the beta chain, and the alpha chain may have one too. The amino acid sequences of each of these regions are similar to their analogues in the immunoglobulins, but they are by no means identical.

Davis and his collaborators and Leroy Hood and his colleagues at the California Institute of Technology have studied the organization of the DNA sequences that encode the beta chain. So far they have found 12 joining regions, two diversity regions (each of which can be read in any one of three transcription frames) and about 20 variable regions. The number of possible amino acid combinations in the beta chain is therefore at least $12 \times 2 \times 3 \times 20$, or 1,440. Even that number is much too small: a realistic estimate of the variability must also take account of mutations and imprecise joining, which affect the *T*-cell receptor just as they affect the antibody molecule. The variability in the alpha chain may be even greater: although no diversity regions have yet been identified, there appear to be many more joining and variable regions than there are for the beta chain—perhaps as many as 100 of each kind. The alpha and beta chains can therefore combine to form on the order of 10 million different kinds of *T*-cell receptor, which is enough to account for the known repertory of *T* cells in an animal.

The new structural information on the *T*-cell receptor makes it possible to recast many of the traditional questions about the immune system in much sharper terms. There are essentially three kinds of observed recognition event for which one would now like to give a structural account. First, the *T* cell does not respond to the MHC-encoded proteins of the self, or in other words it tolerates the self. Second, the *T* cell responds when it is confronted simultaneously with an antigen and a self-MHC-encoded protein, but usually not when it is confronted with an antigen in association with an MHC-encoded protein of another strain of animal. Third, the *T* cell also responds to an MHC-encoded protein from another individual in the absence of antigen; it is this effect that accounts for the rejection of grafted or transplanted tissue. There is a developmental question associated with these observations: How do precursor *T* cells become differentiated in the thymus into cells that have such properties?

The most straightforward explanation for tolerance is that *T*-cell clones reacting to self-MHC-encoded proteins are somehow eliminated in the thymus. There is not yet any clear account of how the screening might take place. One suggestion is that at some stage in their development *T* cells die if their antigen-specific receptors bind any molecule expressed by the cells of the organism. This idea is known as the clonal-abortion theory; it asserts that as *T* cells develop, all *T* cells that react specifically to self-MHC-encoded proteins or to other self antigens would be killed because they are continuously bombarded by such molecules. *T* cells reacting specifically to self-MHC-encoded proteins associated with foreign antigen would also die if they were to bind such complexes, but in an uninfected animal such *T* cells would develop to full maturity. They could then survive until an invading foreign antigen triggered their response.

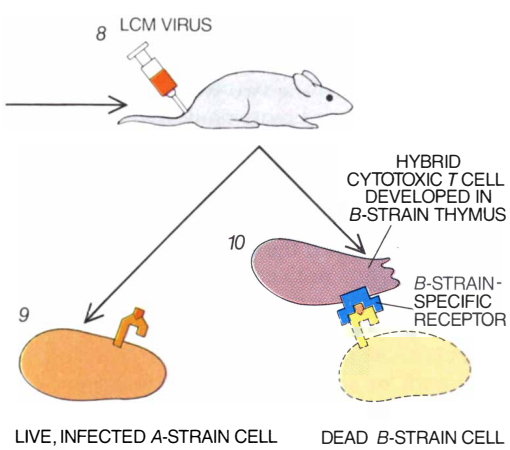
The response of the *T* cell to an MHC-encoded protein associated with an antigen raises its own puzzles. Rolf Zinkernagel and his colleagues and Michael J. Bevan, then at M.I.T., first



EFFECT OF THYMUS on the *T* cell was demonstrated in an experiment by Rolf Zinkernagel, then at the Scripps Clinic and Research Foundation. Hybrid mouse progeny of *a*-strain and *b*-strain parents served as experimental controls. They were immunized against LCM virus (*1*), and cytotoxic *T* cells specific to the virus were isolated from the mice (*2*). The *T* cells were able to kill cells infected with the virus from both *a*-strain and *b*-strain mice (*3*). To determine the role of the thymus, the thymus was removed from another hybrid mouse, and all preexisting *T* cells and other lymphoid cells were killed by irradiating the animal (*4*). New bone-marrow stem cells, from which *T* cells normally develop, were then obtained from another hybrid mouse (*5*), and the irradiated thymus of a *b*-strain mouse was transplanted into the thymectomized hybrid (*6*). The hybrid stem cells were then allowed to develop in the hy-

recognized the dimensions of the problem in their study of the "education" of the *T* cells in the thymus [*see illustration on these two pages*]. They crossed *a*-strain and *b*-strain mice to generate hybrid progeny that carried both strains of MHC-encoded protein throughout their bodies. The thymus of each hybrid mouse was then removed and the animal was irradiated to kill all its *B* and *T* cells. A new thymus from a *b*-strain mouse was transplanted into the hybrid mouse, and new bone-marrow stem cells were grafted into the mouse from another hybrid animal of the same kind.

In an ordinary hybrid mouse the hybrid stem cells would develop into mature *T* cells that respond collectively to antigen in association either with *a*-strain, MHC-encoded protein or with *b*-strain protein. In the experimental animals, however, only *b*-strain proteins were present on the nonlymphatic cells in the thymus, although both strains were still present throughout the rest of the body. Surprisingly, the investigators found that the mature *T* cells from such animals could respond to antigen only in association with *b*-strain, MHC-encoded protein, not to antigen in association with *a*-strain protein. Thus as the *T* cells developed in the *b*-strain thymus their receptors were apparently selected to recognize antigen only in association with the MHC-encoded proteins found in that organ. The *T* cells apparently fail to recognize antigen associated with MHC-encoded protein that is foreign to the host's thymus.



brid mouse with the *b*-strain thymus (*7*). When the experimentally constructed hybrid mouse was immunized against LCM virus (*8*), the *T* cells isolated from it were not able to kill virus-infected cells from *a*-strain mice (*9*), although they were able to kill such infected cells from *b*-strain mice (*10*).

Many hypotheses have been put forward to account for this unexpected finding, none of them yet completely satisfactory. Perhaps *T*-cell receptors must bind weakly to self-MHC-encoded proteins in the thymus before the *T* cells can mature and become functional. Subsequently the *T* cells whose receptors have the highest affinity for self-MHC-encoded proteins might undergo clonal abortion; the remaining *T* cells, with a low but still positive affinity for self-MHC-encoded proteins, would then be released into circulation. When antigen becomes bound to such a protein, however, the affinity of the receptor on the circulating *T* cells for the antigen might be substantially increased.

What about the strong response of *T* cells to foreign tissue graft? The favored explanation is that to the *T* cell a foreign-MHC-encoded protein looks much the same chemically as a complex of a self-MHC-encoded protein bound to an antigen. The explanation accounts for several observations. For example, the *T*-cell receptor that binds the self-MHC-encoded protein appears to be the same receptor that also binds the foreign protein. Furthermore, the receptor seems to have a predisposition for binding to antigen in the presence of either the class I protein or the class II protein, but not both, whether or not the MHC-encoded protein is foreign.

There has been much work in the past year seeking some basis for these receptor predispositions. So far, however, no one has observed any obvious differences between *T* cells restricted to one or the other class of MHC-encoded proteins. In 1984 Tonegawa and his colleagues did find a third kind of gene, called the gamma gene, that is rearranged by *T* cells. The properties of the genetic sequence suggest the protein it encodes is not part of the *T*-cell receptor as the receptor is currently understood. Nevertheless, the protein is expressed only in cytotoxic *T* cells, and so perhaps it contributes in some unsuspected way to the MHC-specificity of the cell that bears it.

If there is a chemical similarity between foreign-MHC-encoded protein and a complex of self-MHC-encoded protein and antigen, one might expect the antigen in the complex to be quite small. A small piece of antigen would also fit more readily into the binding cleft of the *T*-cell receptor than a large piece. Recent work by a number of groups, including Emil Unanue and his colleagues, Ronald H. Schwartz and his colleagues at the National Institute of Allergy and Infectious Diseases and our own group, in collaboration with Howard M. Grey of the National Jewish Center for Immunology and Respiratory Medicine, has shown that the bound antigen is indeed small.

For example, we found that *T* cells respond to live cells incubated with an antigen, but they do not respond to the same antigen when it is added to the cells after they have been fixed with a chemical. When only a small fragment of the antigen is added to the chemically fixed cells, the *T* cell responds once again. The work confirms that the antigen-presenting cells to which a *T* cell can bind have already processed, or digested, the antigen in some way.

We must raise one final point. The existence of the *B* cell makes it clear that a system can be developed in which a receptor recognizes and binds free, native antigen. The binding eventually leads to clonal expansion and differentiation of the *B* cell, which gives rise to antibodies that are quite effective in marking the antigen for destruction. Why then has the *T* cell evolved such an elaborate system for recognizing antigen only in association with the products of the MHC?

There is a teleological answer to the question: The *T* cell is intended to react solely to antigen on a cellular target, and not to free antigen. A cytotoxic *T* cell, for example, is designed to kill virus-infected cells and thereby inhibit the growth and spread of the virus. Such a cell cannot kill a free virus particle, and so the resources of the *T* cell could be squandered if there were no means of directing its attention to a virus growing inside a nucleated cell. The immune system has therefore designed the *T*-cell receptor in such a way that it can bind a viral antigen only when it can also bind a self-MHC-encoded class I protein, which is present on the surface of every nucleated cell in the body.

Even more craftily, the immune system has designed receptors on helper *T* cells that are destined to interact primarily with the *B* cell and with other cells of the immune system. These receptors bind antigen only when it is associated with self-MHC-encoded class II proteins, which are expressed only on the surfaces of the *B* cells and the other immune cells. The immune system has thus evolved in such a way that both classes of MHC-encoded proteins serve as signposts for the *T* cell. They guide the *T* cell to antigen in precisely the places where the action of the *T* cell can be effective.

Quantum Chemical Reactions in the Deep Cold

Quantum-mechanical effects allow some classically forbidden reactions to take place near absolute zero. This suggests that cold, dark clouds of galactic dust could contain seeds of life

by Vitalii I. Goldanskii

From our daily experience we know that chemical reactions slow down at low temperatures. Meat, for instance, can be stored by freezing, a fact that even ancient peoples seemed to realize. About a century ago the Swedish physical chemist Svante Arrhenius proposed a law of classical chemistry that relates chemical reaction rate to temperature. According to the Arrhenius equation, at absolute zero (zero degrees Kelvin, or minus 273 degrees Celsius) the rate of all chemical reactions must be zero, which is another way of saying that the reactions must come to a stop.

Experimental evidence, however, reveals that although in general the Arrhenius equation accurately describes the rate of chemical reactions at relatively high temperatures, the equation fails at low temperatures. In such a domain a quantum-mechanical effect known as tunneling comes into play that allows classically forbidden chemical reactions to take place. Specifically, entire atoms can tunnel through barriers represented by the repulsive forces of other atoms and form complex molecules even though the atoms do not have the energy required by classical chemistry to overcome the repulsion. Of course, the rate at which complex molecules can be formed is extremely low, but the tunneling process could play a significant role.

In this connection I have suggested the possibility of a cold prehistory of life: the formation of rather complex organic molecules in the deep cold of outer space, where temperatures usually reach only a few degrees Kelvin. Cosmic rays (high-energy protons and other particles) might trigger the synthesis of such molecules in dark clouds of interstellar dust. Afterward the reactions would proceed, slowly but surely, by means of tunneling. Within a year of my proposal (1973) two British astrophysicists, Fred Hoyle and N. C. Wickramasinghe, argued that molecules of interstellar formaldehyde have indeed evolved into stable polysaccharides such as cellulose and starch. Their conclusions, although strongly disputed, have generated excitement among investigators who would like to consider the galactic clouds as the places where the prebiological evolution of compounds necessary to life occurred.

It is not my intention to support or refute the claims of Hoyle and Wickramasinghe. I do want to point out, however, that if interstellar formaldehyde has evolved into complex molecules, it can have done so only by tunneling. In speaking of the role that cold chemical reactions might play in the evolution of prebiological compounds, it is fitting to recall the words with which the Russian poet Vladimir Mayakovski addressed the poet and novelist Aleksandr Pushkin: "The eternity is ours, so why not spend a couple of hours?"

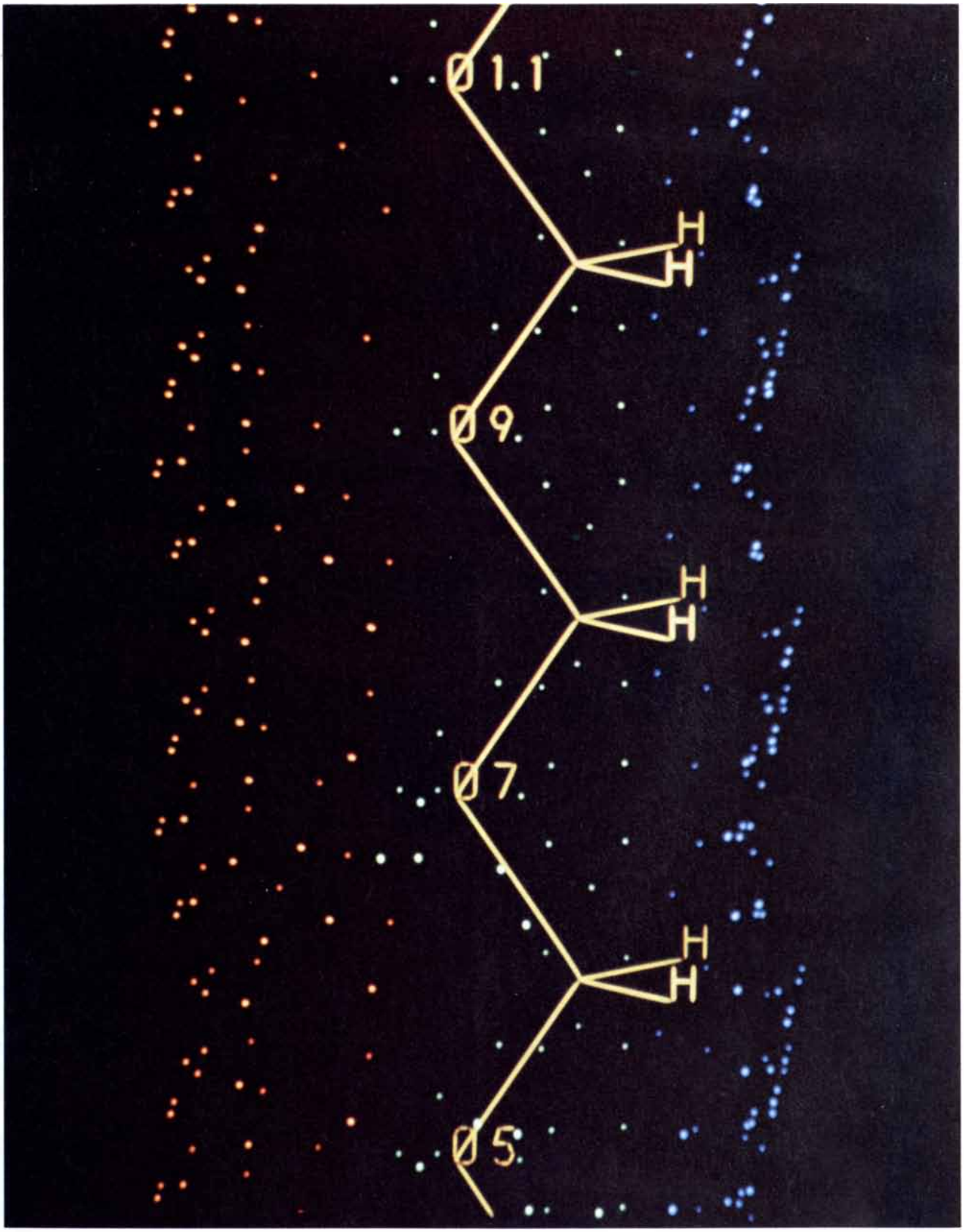
The rate of any chemical reaction, hot or cold, depends on a very important characteristic known as its activation energy, which can be thought of as follows. Any molecule can be imagined to reside at the bottom of a so-called potential well of energy. The depth and width of the well depend on the spatial arrangement of the constituent atoms of the molecule, which changes in the course of a chemical reaction. A chemical reaction therefore corresponds to the transition of a molecule from the bottom of one potential well to the bottom of another. Classically such a transition can be accomplished only by going over an activation, or potential, barrier between the wells, the height of which is called the activation energy of the reaction. The greater the activation energy, the slower the reaction rate.

Activation energies are typically tenfold or more greater than the thermal energies of molecules in our everyday world, where the temperature is roughly 300 degrees K. At the temperature of liquid helium, 4.2 degrees, activation energies are a thousandfold or more greater than the thermal energies of molecules. According to classical chemistry, therefore, at both room and liquid-helium temperatures many reactions will not take place unless the molecules are given additional energy. In a way the situation is like a lottery: before you can win anything, you must raise enough money in order to buy a lottery ticket.

You could, however, win without even bothering to buy a ticket, so to speak, if the reacting molecules could tunnel from the bottom of one potential well to a deeper well without having to rise over the barrier (activation energy) between the two wells.

Tunneling is a direct result of an underlying tenet of quantum theory: the particle-wave duality of all matter and radiation. In the familiar macroworld the duality, although it is not abrogated, is pushed into the background and is usually not noticeable. It is in the microworld, the domain of chemical reactions, that quantum-mechanical effects come into play. Specifically, the behavior of such "particles" as atoms and molecules is often dominated by their wavelike characteristics.

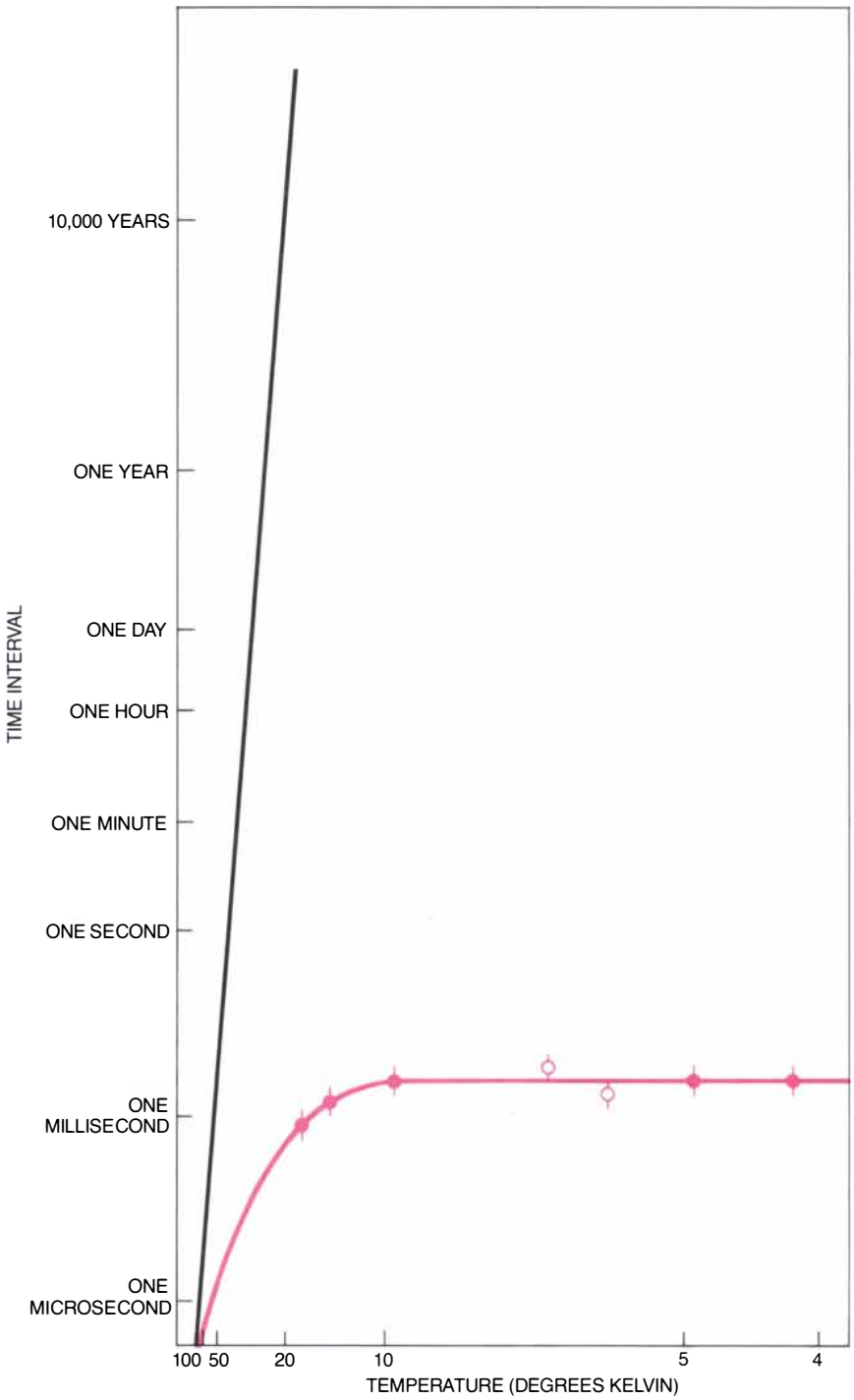
The quantum tunneling of a particle through an activation barrier has a high probability of occurring whenever the width of the barrier is smaller than the characteristic wavelength of the particle involved. The phenome-



FORMALDEHYDE POLYMER, a long chain of individual formaldehyde molecules, can be synthesized at liquid-helium temperatures (4.2 degrees Kelvin, or minus 269 degrees Celsius). According to classical laws, most chemical reactions should stop at such low temperatures; the synthesis of formaldehyde polymers results from a quantum-mechanical effect known as tunneling. In the deep cold of space formaldehyde may even evolve into complex organic compounds such as starch and cellulose. In this computer image, prepared by Paul Weiner and his colleagues at Rutgers University, the red dots around the oxygen atoms (*O*) indicate points of negative potential, and the blue dots around the hydrogen atoms (*H*) and carbon atoms (*unlabeled vertexes*) indicate points of positive potential.

non is analogous to what happens when a light wave strikes the surface of an opaque material. Some of the wave is reflected and some is absorbed by the medium. If the opaque medium is thin enough, the penetrating part of the wave will appear on the other side.

One of the early triumphs of tunneling took place in 1928 when George Gamow applied the theory to explain quantitatively the observed decay rates of certain radioactive elements. Within a year David G. Bourgin of the University of Illinois hypothesized that the activation barrier of chemical reactions could be penetrated by means of a tunneling mechanism. While working at the University of Oxford, Ronald P. Bell made a number of valuable contributions to the understanding of tunneling in chemical reactions. In particular he derived equations for the probability of particles penetrating through activation barriers of varying shape. His predictions were subsequently verified for reactions occurring at the relatively high temperatures characteristic of the liquid phase of materials.



POLYMER SYNTHESIS should require an increasingly long time at low temperatures (*black line*), according to a law of classical chemistry known as the Arrhenius equation. At absolute zero (zero degrees K.) the time should approach infinity and the reaction should come to a stop. Actually, however, the amount of time (*colored curve*) is finite; in other words, the reaction takes place even though it proceeds slowly. As is shown here, the time to add a molecule to the growing chain holds steady at about 10 microseconds for temperatures below roughly 10 degrees. The measurements were made in the author's laboratory.

More recently investigators have turned to searching for the manifestations of tunneling in chemical reactions at low temperatures. Under such conditions the number of quantum tunneling transitions through an activation barrier outweighs the number of classical Arrhenius transitions over the barrier. In a paper written in 1959 I introduced the concept of tunneling temperature: the temperature below which the underbarrier, or tunneling, transitions strongly outnumber the overbarrier, or Arrhenius, transitions, and classical mechanics gives way to its quantum counterpart.

As an example, suppose that the height of an activation barrier is .4 electron volt (which is a typical activation energy of many chemical reactions) and that its width is two angstrom units, or roughly the diameter of an atom. (An angstrom is one ten-billionth of a meter.) Then the calculated tunneling temperature for a hydrogen atom is close to 160 degrees K., and for deuterium, the heavy isotope of hydrogen, it is about 120 degrees. For heavier atoms or molecules the tunneling temperature is still lower because it falls off in proportion to the square root of the mass of the particle. This explains why in order to make significant and valid observations of tunneling in chemical reactions one should work in the cryogenic region.

Another effect also occurs at low temperatures. Near absolute zero all molecules move to the zero-point energy level; that is, they fall to the base of the activation barrier, where the distance between potential wells is maximum. Since the distance remains finite, however, as the temperature goes down the probability of tunneling reaches a finite level that is independent of temperature. When this happens, the reaction rate has reached its quantum low-temperature limit. I also deduced and analyzed the existence of such a limit in the same paper in 1959. Tunneling through the activation energy barrier does not necessarily lead by itself to the appearance of a low-

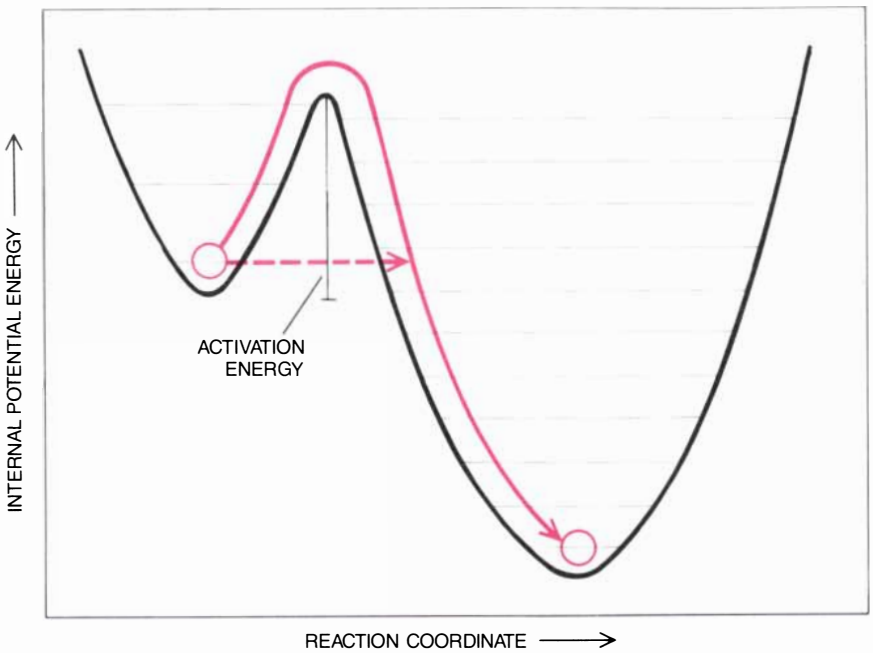
temperature limit for a reaction rate. The existence of the limit does, however, indicate that at low temperatures quantum-mechanical effects prevail over classical effects and that tunneling is playing a decisive role.

What effect does tunneling have on the important class of chemical processes known as redox reactions? Such reactions, which form the oldest branch of chemistry, can be schematically depicted as the transfer of an electron from one chemical species to another. An atom or a molecule that gives up electrons is known as a reducer, or an electron donor, and an atom or a molecule that receives electrons is known as an oxidizer, or electron acceptor. (In a broad sense, an acid is an oxidizer, and a base is a reducer.)

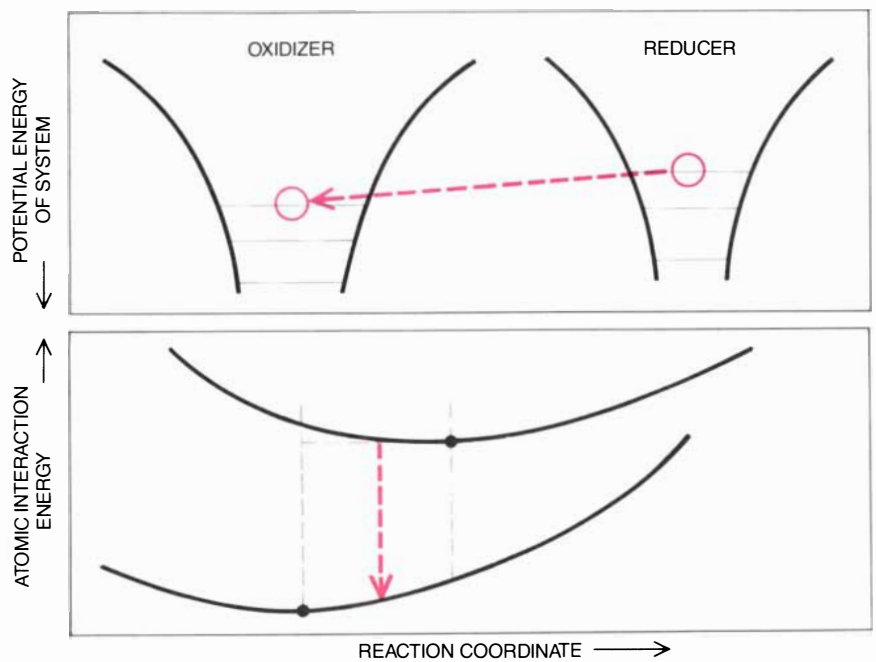
Because the tunneling temperature falls off in proportion to the square root of the mass of the particle doing the tunneling and the mass of an electron is about 2,000 times less than that of a proton, one might expect the tunneling temperature of an electron to be close to 7,000 degrees K. One might therefore conclude that all redox reactions should proceed solely by tunneling. That they do not do so has a simple explanation. In contrast to the free motion of electrons in metals, superconductors and other conductors, the motion of electrons in redox reactions is loosely bound to the motion of (much heavier) atomic nuclei.

The displacement of particles in this process can be likened to the motion of a train consisting of a lightweight locomotive attached to a heavy car by an elastic and easily stretched coupling. The electron is the locomotive and the nucleus is the car: although the electron can travel some distance before the nucleus begins to move, the two particles are linked. The tunneling temperature of the electron therefore depends on the mass of both the electron and the nucleus and as such is typically below 200 degrees K.

The first unquestionable demonstration that tunneling plays an important role in redox reactions at low temperatures was given by Britton Chance and Don C. DeVault of the University of Pennsylvania in 1966. They oxidized (added electrons to) cytochrome *c* by reducing (removing electrons from) chlorophyll. Cytochrome *c* is an enzyme that transports electrons to molecular oxygen by undergoing alternate oxidation and reduction. The investigators found that the reaction rate leveled off at about 120 degrees K. and remained constant through the lowest temperature they examined, 4.2 degrees, the temperature of liquid helium. They calculated that the activa-



TUNNELING between potential wells can allow chemical reactions to take place that would otherwise be prohibited. The potential wells correspond to the various spatial arrangements of atoms in a molecule. The internal energy of each arrangement (*vertical axis*) varies with the reaction coordinate, or the distance between the atoms in the molecule (*horizontal axis*). The molecule itself can be imagined as residing at the bottom of a well. Since the internal energy of a molecule changes in the course of a chemical reaction, a reaction can be represented as the transition of a molecule from the bottom of one potential well to the bottom of another. Such a transition is impeded by a barrier known as the activation energy. Classically the molecule can move from well to well only by going over the activation barrier (*solid colored line*). Quantum mechanically, however, the molecule can tunnel through the barrier (*broken colored line*), even at extremely low temperatures. The gray lines indicate the quantum-mechanically allowed energy levels of the molecule.



REDOX REACTION consists of the transfer of an electron from a reducer (an atom or a molecule that gives up electrons) to an oxidizer (an atom or a molecule that accepts electrons). The top illustration shows the tunneling of a valence, or outer, electron from a reducer to an oxidizer. The potential wells result from the attraction of the negatively charged electron to the positively charged nuclei of the two chemical species. The bottom illustration schematically depicts the atomic interaction in the reducer before it loses an electron (*top curve*) and after it has lost one (*bottom curve*). The transfer of the electron is represented by the broken colored line. The reaction, an exothermic one, releases heat.

tion energy was .14 electron volt and that the tunneling distance of the electron was roughly 30 angstroms. Such a distance corresponds to the linear dimension of a protein globule.

At the Institute of Chemical Physics in Moscow my colleagues Kiril I. Zamarayev, Alpha I. Mikhailov and Ravil F. Khairutdinov and I initiated an investigation of a wide range of redox reactions that involve the tunneling of electrons between ions separated by several tens of angstroms. We studied the recombination of electrons and oxygen anions, O^-, resulting in the formation of O^{2-} ions. A frozen solution of ionized sodium hydroxide (NaOH) donated the electrons. We triggered the reactions by irradiating the solution and monitored the formation of O^{2-} by applying electron paramagnetic resonance.

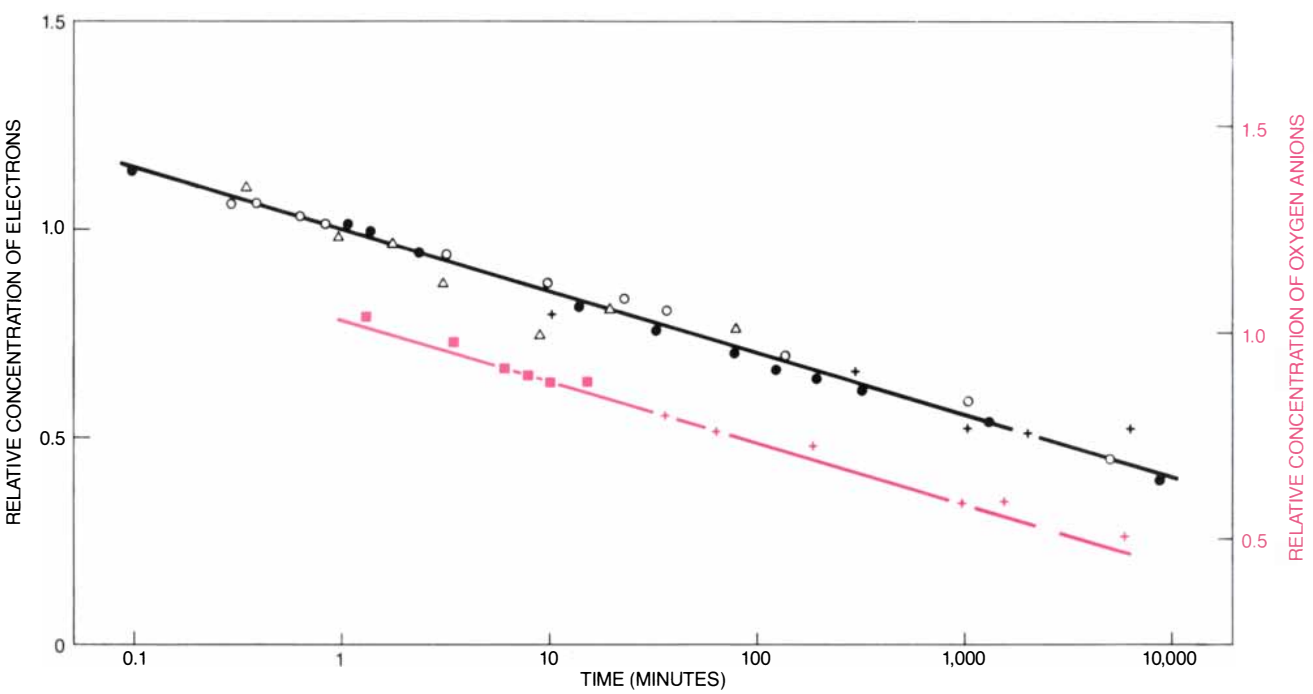
In this technique, which is analogous to nuclear magnetic resonance, the response of electrons in a magnetic field to discrete radiation frequencies is monitored. We observed that the rate of formation of O^{2-} ions remained constant in the temperature interval from 120 to 4.2 degrees K. The constancy of the rate served to show that the reaction was proceeding by tunneling. We estimated that the tunneling distance was between 30 and 40 angstroms. We have also investigated other and more peculiar examples of tunneling in redox reactions.

So far I have discussed only the tunnel transfer of electrons. A chemical reaction in the full sense of the words is, however, a process that both rearranges atoms and changes the nature, length and angles of the valence bonds that hold them together.

The existence of a quantum low-temperature limit for the rate of chemical reactions in such a full sense was discovered by my colleagues Igor M. Barkalov, Anatolii M. Kaplan, Dmitrii P. Kiryukhin and me at the Institute of Chemical Physics in Moscow. We studied the formation of long chains of molecules from individual molecules of formaldehyde (CH_2O). We triggered the growth of these chains both with beams of high-energy electrons from a particle accelerator and with gamma rays (high-energy radiation) emitted by atoms of radioactive cobalt 60. Once it is initiated, the growth of the chains proceeds spontaneously and does not require a further supply of heat because the reactions release heat.

We worked over a wide range of cryogenic temperatures, from 140 to 4.2 degrees K. By measuring the amount of heat produced by the reaction with a device known as a calorimeter we determined the reaction rate. We found we could produce very long molecular chains from the formaldehyde monomers, or individual molecules. At 140 degrees the chains consisted of about 10 million monomer units; at 77 degrees they were 100,000 units long, and near 4.2 degrees they were between 1,000 and 2,000 units long. At temperatures above approximately 140 degrees the mean time required for the addition of a new link to a growing chain increased with decreasing temperature in accordance with the classical Arrhenius equation. At temperatures well below 140 degrees, however, the increase in this mean time slowed progressively until at 12 degrees its quantum low-temperature limit, a hundredth of a second, was reached. At 4.2 degrees the reaction rate was 110 orders of magnitude greater than the Arrhenius equation would lead one to expect.

Our results fit nicely with the picture that the reaction proceeds by the tunneling of individual formaldehyde molecules. A molecule tunnels from the relatively shallow potential well produced by its interaction with neighboring molecules into an adjacent position associated with a deeper potential well. The tunneling molecule forms the end link of the formaldehyde chain, where it emerges in an energetically more stable position, ac-



REDOX REACTION RATE was measured for the case of the recombination of electrons with oxygen anions, O^-, to form O^{2-} anions (doubly negatively charged species). A frozen solution of ionized sodium hydroxide (NaOH) donated the electrons. It was concluded that O^{2-} anions were being formed because the rate of decrease in concentration of the electrons (*black line*) equaled the rate of decrease in concentration of the O^- anions (*colored line*). (The reactant concentrations are normalized on the basis of their initial values; the vertical scales are somewhat shifted to avoid overlapping.) Different symbols correspond to measurements made at different temperatures, ranging from 120 to 4.2 degrees K. The fact that the recombination rate remained constant over that temperature range indicates the reaction proceeded by tunneling; in a classical reaction one would expect the rate of the reaction to decrease with decreasing temperature.

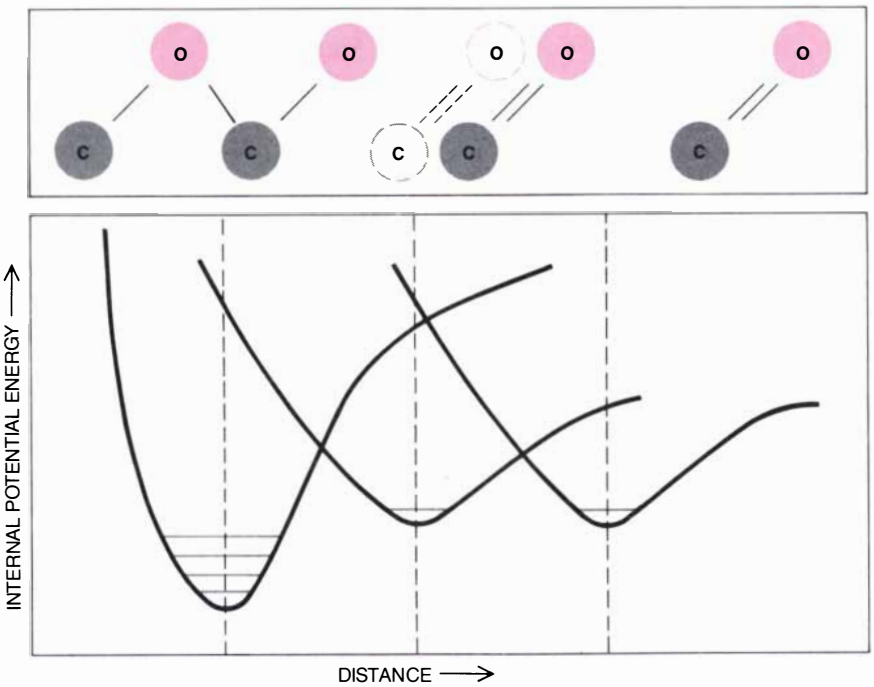
companied by the release and subsequent dissipation of .4 electron volt of energy.

The growth of formaldehyde chains by means of tunneling can be likened to the spontaneous fission of nuclei. As theoretical analysis of tunneling reveals, the probability of barrier penetration is proportional to the product of the barrier width, the square root of the barrier height and the square root of the mass of the particle doing the tunneling. The mass of a chemically reacting species is comparable to the mass of a fission fragment; the barrier height for a typical chemical reaction is about 10 million times less than that for nuclear fission, and the barrier width for a chemical reaction is nearly 3,000 times as great as that for nuclear fission. Height and width effects therefore compensate for each other (the square root of 10 million is approximately 3,000), and the probability of formaldehyde tunneling is comparable to the probability of nuclear fission, unexpectedly enough.

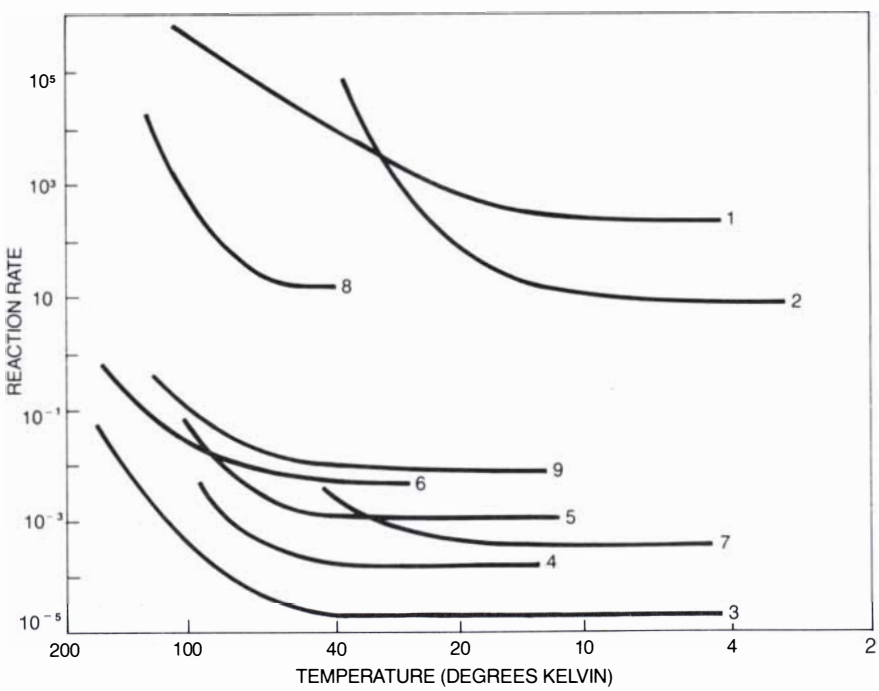
In reality the picture is more complicated than the one I have portrayed: the position of an individual formaldehyde molecule next to the open end of a growing chain is not strictly fixed and the molecule can oscillate between its nearest neighbors. The addition of a new molecule to a chain will most likely occur at the instant when the penultimate molecule in the chain is as close to the end as possible. At that time the height and width of the activation barrier are at their minimum values. When the penultimate molecule takes up its position far from the end of the chain, in contrast, the probability of adding a new link is extremely low. The height and width of the activation barrier are then very large, and the frequencies of both overbarrier and tunneling transitions are much smaller.

The effects of such relatively slow intermolecular oscillations were clearly seen in experiments my colleagues and I carried out at the Institute of Chemical Physics. We studied the radiation-triggered reaction of hydrogen bromide and ethylene. We found that the substitution of deuterium for hydrogen caused a fourfold reduction in the reaction rate at temperatures around the tunneling temperature. If the reaction had proceeded between stationary molecules, however, the decrease should have been by a factor of about 100,000.

Tunneling mitigated by intermolecular motions may also play an important role in the functioning of proteins and the nucleic acid DNA. The consequences may be felt not only at cryogenic temperatures but also at temper-



MOLECULAR TUNNELING of a formaldehyde monomer (*top right*) can lead to the formation of a polymer chain (*top left*). Formaldehyde consists of an oxygen atom (*O*) and two hydrogen atoms (*H*), all three of which are bound to a carbon atom (*C*). (The two hydrogen atoms are not shown here.) The complex in the middle of the top sequence schematically depicts the intermediate step of the process: the breaking of the double bond between the oxygen atoms and the carbon atoms to allow for the forming of the single –C–O–C–O– bonds of the polymer. The bottom part of the illustration shows the potential-energy wells of each stage in the reaction. Repetition of the process increases the length of the polymer.



LOW-TEMPERATURE REACTION-RATE LIMITS indicate that tunneling has occurred. Such limits have been observed for many chemical reactions: (*1*) the growth of formaldehyde polymer chains (1973); (*2*) the rebinding of carbon monoxide to hemoglobin (1975); (*3*) the isomerization, or structural rearrangement, of radical pairs in dimethylglyoxime after irradiation with gamma rays (1977); (*4, 5*) the abstraction of hydrogen atoms by methyl radicals from frozen methanol (*4*) and ethanol (*5*) (1977); (*6*) the transfer of hydrogen atoms during the isomerization of certain radicals (1978); (*7*) the formation of carbon-carbon bonds in certain radicals (1979); (*8*) the hydrobromination of (addition of hydrogen bromide to) ethylene (1978), and (*9*) the chlorination of butyl chloride by molecular chlorine (1980). The observations are from the U.S.S.R., the U.S.A., Japan and Canada.

atures above 200 degrees K. and even up to physiological temperatures. Specifically, tunneling may significantly contribute to transitions between so-called conformational substates in biopolymer molecules.

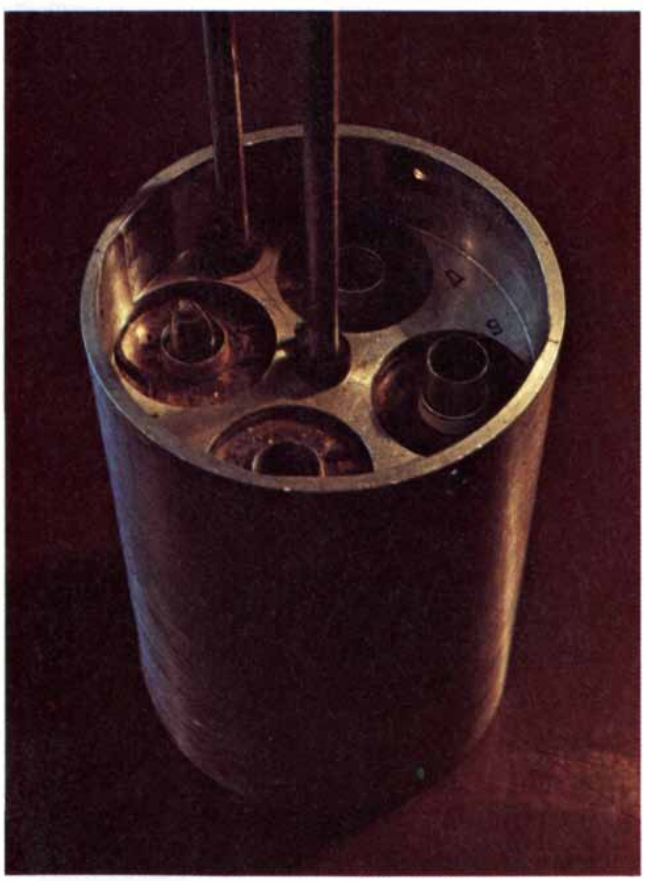
Conformational substates correspond to slightly different configurations of a molecule that have roughly equivalent energies. An atom or a group of atoms may, for instance, be rotated about a single bond. Transitions between substates account for the mobility of biopolymer molecules and may also be involved in their functioning. Hans Frauenfelder and his colleagues at the University of Illinois at Urbana-Champaign and Wayne State University and David Phillips and his colleagues at Oxford discovered and investigated the properties of conformational substates. To aid in their work both groups employed such techniques as X-ray-diffraction analysis. In collaboration with Rudolf Mössbauer and Fritz Parak of the Technical University of Munich my colleagues and I have also investigated the substates.

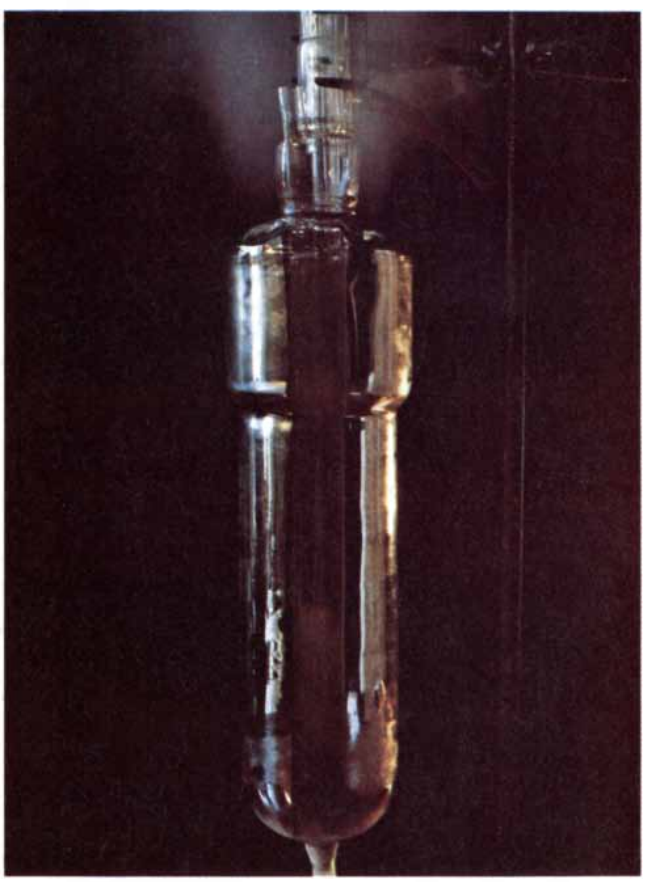
The study of transitions between conformational substates has led to new dynamical models of proteins and other biopolymers. Some 40 years ago the eminent physicist Erwin Schrödinger postulated in his book *What Is Life?* that a protein is an aperiodic, or irregular, crystal. About 10 years ago investigators advanced the view that a protein consists of a quasi-solid core surrounded by a liquidlike shell. According to that view, the shell accounts for most of the volume of the protein globule. The third and most recent proposal holds that proteins and DNA can be thought of as heterogeneous glasses, and that the transitions between their substates at cryogenic temperatures and even perhaps above 200 K. are really instances of tunneling.

Two final remarks are in order about some of the other effects of tunneling. Although tunneling plays a role in the formation of chemical bonds, it can also take part in the destruction of bonds, and therefore of compounds. This is because of tunneling penetration through activation barriers that tend to oppose the development of destructive processes.

The fact that tunneling allows chemical reactions to occur at extremely low temperatures adds yet another reason for doubting the possibility of reviving complex organisms that have been frozen for a long time. At the very least the organisms would have to be shielded from events that trigger chemical reactions such as penetrating radiation.

The experimental data and theoretical considerations regarding the manifestations of tunneling in chemical reactions, along with the many applications of quantum chemistry in conditions of deep cold, testify to the fact that chemistry has not degenerated into a servant of physics and biology. Chemistry is, as it has always been, a solvent partner in the exchange of new, fundamental ideas.





CALORIMETERS, devices that measure the amount of heat released or absorbed during a reaction, were employed in the author's laboratory to determine the rate of polymerization of formaldehyde at cryogenic temperatures. The instrument at the left proved to be effective over a temperature range from 350 to 80 degrees K. A glass vessel containing formaldehyde monomers protrudes from the leftmost chamber of the calorimeter. During the experiments the entire unit was placed in a larger vessel and chilled with liquid nitrogen. The calorimeter at the right was employed at temperatures down to 4.2 degrees K. An outer region of that instrument was filled with liquid nitrogen and an inner region was filled with liquid helium. Igor M. Barkalov, Anatolii M. Kaplan and Dmitrii P. Kiryukhin of the Institute of Chemical Physics in Moscow developed both calorimeters. The author supervised the experimental procedure.

Give in to an overwhelming drive.

Who'd blame you? After all, the new Volkswagen Jetta GLI with its sleek aero-dynamic lines and its gleaming body has pumped a lot of adrenaline into a lot of hearts. And its looks aren't deceiving: The GLI's high performance fuel-injected engine and close-ratio 5-speed transmission move it from 0-50 in 6.8 seconds. Its independent sport suspension gives you crisp overall handling with improved tracking on curves thanks to its patented rear axle.

On the other hand, the carrying capacity of the GLI (5 people and 14 standard size pieces of luggage) makes it unbelievably practical. And its remarkably low price makes it practically unbelievable.

The 1986 Jetta GLI. **It's not a car. It's a Volkswagen.**

Seatbelts save lives. All Volkswagens are covered by a 2-year Unlimited-mileage Protection Plan and a 6-year limited warranty on corrosion perforation. See dealer for details about these limited warranties. © 1985 Volkswagen

SCIENCE AND THE CITIZEN

Potemkin Wars

During the past several months the Strategic Defense Initiative Organization (SDIO) has treated the public to a series of demonstrations apparently designed to confirm the validity of the basic missile-tracking and -killing technologies the Reagan Administration hopes the U.S. will someday deploy. A laser beam shone from a mountaintop on Hawaii was reflected back from an eight-inch-diameter retroreflector on the space shuttle *Discovery;* an experimental missile has shot down a working satellite; here on the earth, both a chemical laser and a projectile of the kind that would be fired by an electromagnetic gun have destroyed mock-ups of missile boosters, one of which was designed to telegenically disintegrate. Lieutenant General James A. Abrahamson, Jr., the director of the "Star Wars" program (as the defense is popularly known), has underlined the message, declaring that progress is achieving "an incredible pace." Is it?

Two types of weapons for destroying ballistic missiles in flight are under consideration: directed-energy weapons and kinetic-energy weapons. Directed-energy weapons are lasers and particle beams; kinetic-energy weapons are chemical guns (rockets) and electromagnetic guns. What requirements must be met to make such weapons practical?

Wavelength and brightness determine whether a laser can serve as a weapon. Typical wavelengths produced by most lasers are absorbed by the atmosphere; laser wavelengths between .3 and one micrometer are in general the most easily transmitted. (A micrometer is a millionth of a meter.) To kill a missile a laser must also be able to achieve a certain level of brightness, or intensity. SDI workers estimate that if a laser firing over a 3,000-kilometer engagement distance is to burn through a missile skin in one second, it must deliver 10,000 joules of energy per square centimeter. These requirements correspond to a brightness of 10^{21} watts per steradian, or unit of solid angle. A laser having the desired wavelength and brightness would require a beam power of about 100 megawatts. (A typical nuclear power plant has an output of about 1,000 megawatts; a coal power plant has an output of about 700 megawatts.)

Four types of laser are currently under development: chemical, excimer, free-electron [illegible] X-ray. Chemical lasers, which [illegible]erate radiation by means of chem[illegible] reactions between two gases, repre[illegible]t the most mature technology. The [illegible]ghtest of these is the mid-infrared [illegible]vanced chemical laser (MIRACL), a l[illegible]rogen-fluoride laser, which is at the [illegible]ite Sands Missile Test Range in New [illegible]xico. It has destroyed a mock-up of Soviet missile standing about half a [illegible]le away.

The brightness of MIRACL, roughly 10^{17} watts per steradian, falls short of the SDI goal by a factor of about 10,000. Moreover, the wavelength of MIRACL, 2.8 micrometers, is too long to be transmitted great distances through the atmosphere. Indeed, the wavelengths generated by most chemical lasers, typically between one micrometer and four micrometers, not only are too long to be transmitted through the atmosphere but also are subject to divergence even in the near-vacuum of outer space.

Two lasers in relatively early stages of development that offer promise of delivering radiation within the range of wanted wavelengths are the excimer and the free-electron lasers. Excimer stands for excited dimer: an unstable compound composed of two molecules. An electric discharge excites the molecules into forming the dimer; as the dimer breaks down it gives off radiation, triggering a cascade of such reactions that produces the laser beam. The radiation in the beam has a wavelength that is generally between .2 and .4 micrometer. The most powerful excimer laser is the krypton-fluoride laser at the Los Alamos National Laboratory. Operating at a wavelength of .25 micrometer, it can now deliver 10,000 joules of energy in a 380-nanosecond pulse. (A nanosecond is a billionth of a second.) Although the energy matches the SDI goal, the pulse duration needs to be increased by a factor of about three million.

In a free-electron laser a beam of electrons passes by a series of "wiggler" magnets that cause the electrons to vibrate and emit radiation. By changing the distance between the magnets the laser can be tuned to radiate at theoretically any wavelength between about .1 micrometer and 20 micrometers. In general smaller wavelengths require greater energy. The largest free-electron laser, which is at Los Alamos, has been operated at wavelengths down to 10 micrometers. Workers hope to upgrade the device to produce one-micrometer radiation of 100-microsecond pulses containing 30 kilowatts of power.

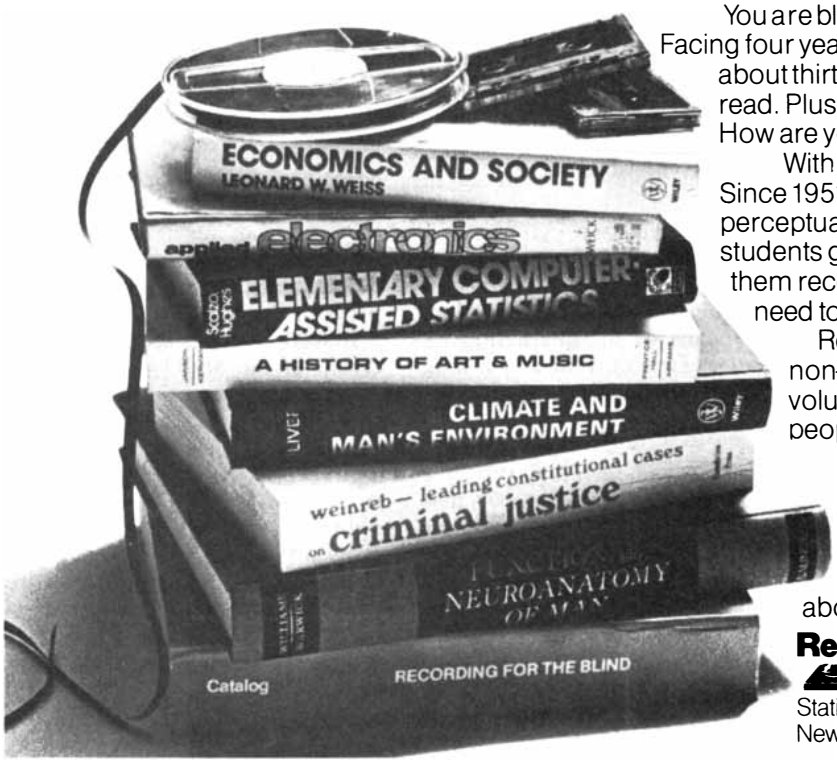
A serious drawback to both excimer and free-electron lasers is their relatively poor efficiency in converting electric energy into beam energy: 2 to 3 percent compared with the 10 percent efficiency of chemical lasers. Low efficiency translates into high power requirements and massive power supplies, making it impractical to base excimer and free-electron lasers in space. Beams generated by ground-based lasers are, however, distorted by atmospheric turbulence, the phenomenon responsible for the twinkling of stars. Although the distortions might be removed by special mirrors known as modular mirrors or by a technique known as optical phase conjugation, the technologies are still young and need further development.

Perhaps the most exotic laser under development is the X-ray laser. The device consists of a nuclear explosive surrounded by a cylindrical array of thin metal fibers. X rays emitted during the nuclear explosion stimulate the emission of a beam of X rays from the fibers in the microsecond before the device immolates itself. Several months ago news leaked out that workers had succeeded in intensifying and focusing the beam during an underground test. More recent reports indicate that monitoring equipment had been improperly calibrated, thereby rendering the results uncertain.

The other class of directed-energy weapons consists of particle beams, which can be subdivided into charged-particle beams and neutral-particle beams. According to Stephen Rockwood, head of SDI work at Los Alamos, in order to make an effective weapon a particle beam would need an energy of 250 million electron volts. He notes: "If one assumes an accelerating gradient of 10 million electron volts per meter, then it follows that the structure must be at least 25 meters in length. When accounting is made of the mass of the power supply and its fuel, the weight of the weapon is found to be 50 to 100 tons." Such a bulky device would be expensive to put in space; currently typical payloads weigh a few tons.

Significant hurdles would also have to be surmounted if particle beams were based on the ground. Weapon specialists generally agree that a particle beam would not penetrate the atmosphere. In addition the magnetic field of the earth would bend

Close your eyes. Now have someone read this to you.



You are blind. A student. Facing four years of college. With about thirty-two textbooks to read. Plus fifty supplemental texts. How are you going to manage?

With Recording for the Blind. Since 1951, we've helped over 53,000 blind, perceptually and physically handicapped students get through school. By sending them recordings of the books they need to read. Free.

Recording for the Blind is non-profit, and supported by volunteers and contributions from people like you who can imagine what it's like to be blind.

Your tax-deductible donation will help our students meet their educational goals. We'd all be grateful.

If you want to know more about us, write:

Recording for the Blind, Inc.
an educational lifeline.
Station E, 215 East 58th Street
New York, New York 10022, (212) 751-0860.

Now, from Europe . . .

Biofutur, Europe's foremost magazine of biotechnology, comes to you in English!

Read about important French advances on the most challenging frontier of science and industry—biotechnological research and development.

A valuable addition to the library of every scientist, technician and corporate manager in the field, ***BIOFUTUR***'s special issue looks with particular emphasis at the practical uses and industrial applications of biotechnology in agriculture and agro-food, medicine and pharmaceuticals and chemistry and energy conversion.

We are pleased to introduce you to ***BIOFUTUR*** with this exciting and authoritative 150-page issue on biotechnology in France.

sa

Please send _____ copy/copies of ***BIOFUTUR*** at $10.00 (or 79 FF) each.
Make check or money order payable to ***BIOFUTUR***, or use credit card.

VISA card Account No. _____ Expiration Date _____
MasterCard Account No. _____ Expiration Date _____
American Express No. _____ Expiration Date _____
Name _____
Company _____
Address _____
_____
Country _____
Signature _____

For U.S. and Canadian residents
Mail coupon
BIOFUTUR
P.O. Box 22576, Kansas City, Mo. 64113-2576 U.S.A.

For other countries
BIOFUTUR
29, rue Buffon
75005 Paris, France

a charged-particle beam. Preliminary work has been done to guide charged-particle beams through the atmosphere with laser beams. (The approach also offers promise as a way to counteract beam instabilities arising from the repulsive force between charged particles that have the same sign.) The basic idea is to create in a gas an ionized channel that guides a high-energy electron beam just as a metallic pipe would. Such a channel is formed by directing a laser beam through the gas, which strips electrons from the atoms of the gas. By exploiting variations of this technique workers at both the Sandia National Laboratories and the Lawrence Livermore National Laboratory have guided high-energy electron beams over distances of several meters. The solution would seem to offer little benefit in outer space, however, where there is no gas to ionize and where the effects of the earth's magnetic field are still felt.

Because neutral-particle beams are not affected by magnetic fields, investigators have directed attention to them. The largest program for developing this technology is the Accelerator Test Stand at Los Alamos. Neutral-particle beams with energies of three million electron volts have been achieved; the target for this year is seven million electron volts, several orders of magnitude less than Rockwood's goal of 250 million electron volts. A drawback of neutral-particle beams is that they are rapidly ionized: the atmosphere strips them of their outer electrons. As a result a neutral-particle beam can become a charged-particle beam and inherit the latter's limitations.

What is the forecast for kinetic-energy weapons (electromagnetic guns and chemical guns)? Electromagnetic guns accelerate projectiles to hypervelocities by means of an interaction between an electric current and a magnetic field. SDI researchers would like to be able to fire three-kilogram projectiles at speeds of from 15 to 30 kilometers per second; the greatest projectile speed attained so far has been 11 kilometers per second for a three-gram plastic bullet. Although the development of chemical guns is far more mature than the development of electromagnetic guns, the problems associated with each type of kinetic-energy weapon are not insignificant. As Gerold Yonas, chief scientist of the SDIO, notes in *Physics Today:* "There is little question of the lethality of an eight-kilogram projectile, say, hitting a target at ten kilometers per second.... But the development of a low-mass guidance and control system for such a homing projectile, capable of surviving initial acceleration, and of low-cost chemical rockets or a space-based power supply to deliver hundreds of megawatts of electricity for electromagnetic guns may be among the most difficult jobs before the SDI research program."

Indeed, it would seem that the leadership of the SDIO is cautious in its approach to some of the more complex problems entailed by the proposed shield. Following the announcement of congressional budget cuts (the Administration had asked that $3.7 billion be appropriated for the SDI program this fiscal year, but Congress approved $2.7 billion), Abrahamson stated that less emphasis will be placed on the development of space-based lasers and electromagnetic guns. Workers will focus instead on land-based lasers and "smart rocks," or killer rockets, because, Abrahamson remarks, "it appears right now to be the simplest way to proceed."

LASER-BEAM SPOT is seen as it appeared from the space shuttle *Discovery* during a tracking experiment conducted last June. An argon laser on a mountaintop in Hawaii projected the beam to the shuttle as the craft passed over the islands at an altitude of 230 miles. Initially pencil-thin, the beam widened to a diameter of 15 feet as it passed through the atmosphere. For more than two minutes the beam locked onto a retroreflector on the shuttle.

The Bad Screed

Although Sir Cyril Burt was exposed as a fraud more than a decade ago, his scientific legacy is alive and well in the world of textbook publishing. Burt, the eminent British psychologist who died in 1971 at the age of 88, devoted much of his career to demonstrating the effects of genes on human intelligence. In the late 1970's it was found that Burt had fabricated some of his data, disguised the sources of other information and even invented two of his purported coauthors. Following those revelations Burt's studies were generally rejected as unreliable. Yet according to Diane B. Paul of the University of Massachusetts at Boston, Burt's work continues to appear in genetics textbooks, often without attribution to him.

Paul concludes that the anonymous citation of Burt's results and of other discredited or outdated studies is part of a pattern of poor scholarship with regard to genetics and human intelligence that prevails in college-level genetics textbooks. Textbooks often make claims about the influence of genetic inheritance on intelligence that are not justified by the available empir-

ical findings. Paul's findings, published in a recent issue of *The Quarterly Review of Biology,* are based on an examination of 28 of the 31 introductory texts in human or general genetics published in the U.S. between January, 1978, and March, 1984.

Of the 28 texts, 19 included a substantial discussion of the heritability of human intelligence. Most of the 19 books stated that the heritability of human I.Q. is high, and many offered numerical estimates of the degree of heritability. According to Paul, however, the studies most often cited in support of the hypothesis that human intelligence is highly heritable are fraudulent or are unreliable because of methodological inadequacies. Among them is an analysis by Burt of identical twins who were reared in different homes. That work, carried out over many years, was thoroughly discredited in the 1970's when it was discovered that the findings Burt reported had been based on spurious data.

In 10 of the 19 textbooks, however, Burt's analysis of identical twins appeared as evidence for a high degree of heritability. His results were generally not referred to explicitly. Instead, they were incorporated in a figure summarizing the findings of 52 studies of I.Q. and genetic relatedness. The figure is from a review published in *Science* in 1963 by L. Erlenmeyer-Kimling and Lissy F. Jarvik of the Columbia University College of Physicians and Surgeons. In addition to including Burt's fabricated data the review has many methodological shortcomings and is widely perceived by workers in the field as being outdated, Paul says. Nevertheless, she found it was the most frequently cited evidence for a high level of heritability of intelligence. Indeed, in some of the texts it was the only evidence.

Textbooks propagate invalid hypotheses partly because their authors copy from one another, Paul suggests. The writer of an introductory text is rarely expert in every field covered by his or her book. Consequently the author tends to rely on material presented in other texts, often copying passages almost verbatim. The result is that a finding can persist in textbooks long after it has come to be rejected by knowledgeable investigators.

The idea that genes have a strong influence on intellectual performance also may be consistent with the biases of the textbook authors. "Whatever the causes of the current situation," Paul writes, "the consequence is clear. A majority of genetics students are being taught that intelligence is highly heritable . . . on the basis of evidence from studies that more properly belong in histories of science, or pseudoscience, than in contemporary textbooks."

Gloomy Groves

College teachers in the U.S. are so dissatisfied with their pay and their chances for advancement that nearly 40 percent of them say that within five years they may seek other kinds of work. The finding emerges from a survey conducted among 5,000 faculty members at four-year and two-year colleges by the Carnegie Foundation for the Advancement of Teaching.

The survey shows that 60 percent of the teachers regard their salaries as either fair or poor. About 75 percent say their college salaries have not kept pace with inflation. In fact, the typical teacher—a composite of medians calculated from the survey data—is a 46-year-old white male who has a tenured professorship at a four-year public institution and earns $29,000 a year.

A related source of dissatisfaction is the disparity in earning power between teachers of science and engineering and teachers of subjects in the liberal arts, due to consulting work and other off-campus opportunities. Although the survey did not explore this issue, officials of the foundation acknowledge that it "is there."

The prospects for advancement also contribute to the sense of dissatisfaction. Two-thirds of the teachers said tenure is harder to win than it was five years ago. Nearly 30 percent said they feel trapped in their jobs because opportunities for promotion or moving to a different institution are limited.

Self-Defense

Media reaction was overblown late last year when National Cancer Institute (NCI) investigators reported that treatment with a chemical messenger playing an important role in the immune system reduced tumors by at least half in 11 out of 25 patients with advanced cancer. The therapy is in fact one of several therapeutic strategies, sometimes called biological-response modifiers, that differ from conventional surgery, radiation or chemotherapy.

Biological-response modifiers recruit elements of the natural immune system to attack tumor cells while, it is hoped, leaving healthy cells intact. The new approach has developed from improved understanding of the interplay among such immunologic elements as lymphocytes (white blood cells, some of which attack diseased cells directly), antibodies (molecules that bind to specific target cells and can trigger their destruction), macrophages (scavengers that digest foreign particles) and lymphokines (molecules, secreted by lymphocytes, that stimulate attack-cell activity). Recent advances in technology have made possible the isolation, cloning and systematic study of these elements and others.

The December results were reported by Steven A. Rosenberg and his colleagues at the NCI. They removed lymphocytes from the blood of patients and activated the cells with interleukin-2 (IL-2), a lymphokine. The resulting cells, designated lymphokine-activated killer (LAK) cells, were then infused back into patients, along with extra IL-2 to enhance the cells' tumor-destroying ability.

Cancer experts are encouraged by the results, which included total remission of melanoma (a virulent skin cancer) in one person and tumor regression in 10 other patients with skin, colon, kidney or lung cancer. Yet many experts, including Rosenberg, insist that the data are too sketchy to indicate whether the new treatment will prove to be practical and safe. Among the obstacles to widespread future use are cost (tens of thousands of dollars per person), complexity and side effects. In two patients the most serious side effect—fluid retention—brought on severe breathing problems. Rosenberg has also noted that one person in a later group of subjects died soon after treatment.

Clinical trials are also being conducted to explore the therapeutic potential of interferons, which stimulate macrophages and tumor-fighting lymphocytes. The form designated alpha interferon was once touted as a "magic bullet" against cancer. Jordan U. Gutterman of the M. D. Anderson Hospital and Tumor Institute in Houston has found it is effective in varying degrees against some rare malignancies, notably one called hairy-cell leukemia. It has, however, been disappointing against lung, colon, pancreas and prostate cancers.

Interferons have been shown to be more effective when they are administered together or in conjunction with tumor necrosis factor (TNF). TNF, discovered by Lloyd J. Old of the Memorial Sloan-Kettering Cancer Center, is a monokine (a protein released by activated macrophages) that is toxic to tumor cells. Old thinks TNF may have had a role in the Rosenberg successes, inasmuch as peripheral blood cells undergoing treatment with IL-2 produce TNF.

Another line of research focuses on monoclonal antibodies, which are engineered to be specific for selected

antigens. A monoclonal antibody can, at least in theory, deliver a toxin, drug or radioactive isotope selectively to tumor cells, leaving normal cells unharmed. Monoclonals administered without "foreign" substances are also being studied. Hilary Koprowski of the Wistar Institute of Anatomy and Biology says that monoclonals have activated macrophages taken from patients and then directed the reintroduced macrophages to cancer sites, reducing tumors in about half of some 300 patients with metastasized cancer, including cancer of the pancreas, stomach and large bowel. Monoclonals have also been tested as anticancer weapons in their own right in small clinical trials against lymphoma, leukemia and melanoma, with varying degrees of success. Antibodies do not kill cells directly but appear to activate such immune-system components as nearby lymphocytes, macrophages or complement (a cascade of proteins that destroy antibody-bound cells).

Among other potential immunotherapies are BCG, a tuberculosis vaccine that has been studied for some time and has been shown to have some effect on melanoma and the recurrence of bladder cancer; interleukin-1, which stimulates lymphocytes to produce IL-2, and colony-stimulating factor, which activates macrophages and other cells derived from bone marrow. Genes encoding the last two have now been cloned, making clinical trials possible. Old has great hope for yet another approach: vaccination to prevent the recurrence of a cancer after the primary tumor has been removed.

Whether novel immunological approaches will one day replace conventional therapies is an open question, but many investigators are confident that new immunotherapies will at least augment established treatments, in some cases completing tumor reduction begun by surgery, radiation or chemotherapy. Ultimately combination immunotherapies are expected to be most effective. As Gutterman points out, immune-system elements normally work in concert, and the same synergy is now being seen among experimental biological agents.

New Nerves

The possibility that the ends of a severed nerve can be induced to grow together and resume function is suggested by work recently reported to the Society for Neuroscience. Two groups of workers have done experiments in which the severed ends of nerves from the peripheral nervous system of a rat were inserted into a silicone tube filled with a polymeric matrix that seems to facilitate nerve growth. Nerve fibers were apparently induced to regenerate over a gap as wide as 15 millimeters—which is a considerable distance in nerve-injury cases.

One of the groups is headed by Ioannis Yannas of the Massachusetts Institute of Technology. Yannas' group packs a silicone tube with collagen and a polysaccharide; the materials are cross-linked so that they form a porous network. Another group, led by Richard Sidman of the Harvard Medical School, uses either a collagen matrix or a gel containing a glycoprotein called laminin.

At present a damaged peripheral nerve that is crucial to bodily function can be repaired by surgically grafting a piece of a nerve from another part of the body. The approach has limitations. The area from which the nerve was removed is left without sensation. Moreover, the amount of nerve tissue that can reasonably be removed for such grafts is limited.

Why does the tube technique work? "When the polymer wasn't used," said Jerry Silver of the Case Western Reserve University School of Medicine, a member of the Yannas group, "there was either no nerve growth or only growth of connective tissue." With the polymer, blood vessels re-formed and a tremendous amount of growth occurred. In addition a large number of Schwann cells—nonneuronal cellular elements that provide structural support and insulation to nerve endings—were observed with the electron microscope. "Apparently," Silver observes, "the polymer-filled tube allows the Schwann cells to travel farther along the gap than [they would] in the empty tube. Our guess is that the polymer's surface promotes cell migration." The next step, according to Silver, is to try to increase the efficiency of the polymer by seeding it with Schwann cells before the tube is implanted.

Yannas and Sidman and their colleagues worked with tissues from the peripheral nervous system (the nerves outside the brain and the spinal cord). Albert Aguayo of McGill University and his colleagues have found that axons in the central nervous system (the brain and the spinal cord) can regenerate significantly in the presence of grafted peripheral-nerve tissue.

Hands of Clay

It seems to defy intuition that nature should prefer right-handedness or left-handedness, yet there are several examples of such fundamental asymmetry. G. E. Tranter of the University of Oxford, writing in *Nature,* has proposed a mechanism by which two of the most basic examples, one from subatomic physics and the other from biology, could be causally related.

Tranter notes that the weak nuclear force, which is responsible for such phenomena as beta decay (the decay of a neutron into a proton, an electron and an antineutrino), is asymmetric. For example, in beta decay the electron always emerges from the nucleus with "left-handed" spin. The mirror-image case, in which the electron emerges with right-handed spin, never occurs.

Tranter also points out that such substances as amino acids and sugars, which rotate the direction of polarization of linearly polarized light, do so in part because they are asymmetric molecules. In principle both mirror-image forms of any asymmetric molecule could exist, yet all the amino acids found in living systems are of the form that rotates linearly polarized light counterclockwise; the sugars rotate it clockwise. Why should one form of a molecule outnumber its mirror image to such an extent?

It has often been suggested that interactions due to the weak force might make one molecule more energetically favorable than its mirror image. The weak force can indeed have an effect on an atom's outer electrons (which are most responsible for its interactions with other atoms) and in principle it could therefore have an effect on the formation of molecules. Outside the atomic nucleus the energy of events that depend on the weak force is almost vanishingly small, however, and so it is hard to see how the weak force could be responsible for the great predominance of certain asymmetric molecules over their mirror images.

Tranter proposes a way the effect of the weak force could be amplified: the small energy differences between mirror-image molecules could have been greatly magnified in crystals or clays containing large numbers of identical molecules. He notes that a detectable difference in the quantities of left- and right-handed forms of one particular crystal, quartz, has already been documented.

Such asymmetric minerals may in turn have acted as catalysts in the formation of the first biotic molecules, according to Tranter. Asymmetric catalysts may well have selectively produced one mirror-image form of a molecule. Tranter adds that his hypothesis represents an argument in favor of such theories as that of A. G. Cairns-Smith of the University of Glasgow, who has hypothesized

that layered silicate clays may have been the chemical precursors of such biotic molecules as RNA.

Retrenching

In a series of deep trenches bordering the Pacific ocean basin from the Aleutian Islands to New Zealand the Pacific plate, the piece of the earth's crust bearing that ocean's floor, descends at a rate of a few centimeters a year into the mantle and is consumed. The process, called subduction, generates earthquakes and, in island arcs behind the trenches, gives rise to volcanic activity.

At the December meeting of the American Geophysical Union in San Francisco, Loren W. Kroenke and Daniel A. Walker of the Hawaii Institute of Geophysics presented evidence that a new subduction zone may be developing along a crescent of sea floor stretching from the southern end of the Mariana Trench to Western Samoa, some 3,000 miles to the southeast at the northern end of the Tonga Trench.

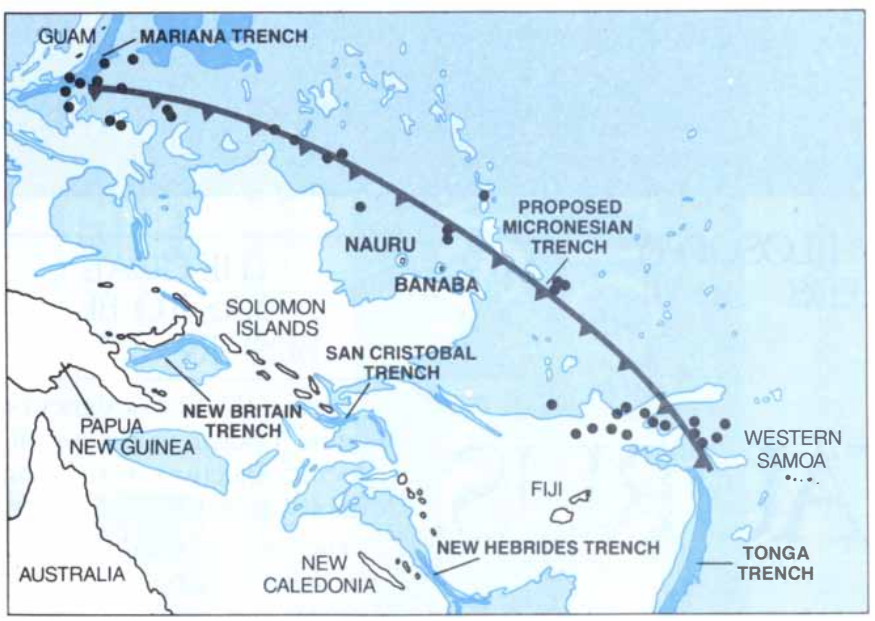
The evidence comes, the workers said, from an array of hydrophones operated since 1982 by the Hawaii Institute of Geophysics. The hydrophones are moored in the middle depths of the ocean and on the ocean floor near Wake Island. The array has detected numerous shallow earthquakes originating in the Pacific floor that have gone unrecorded by conventional seismic stations based on land.

A natural waveguide that carries seismic waves through the deep ocean floor may account for the discrepancy in the observations, according to Walker. Much of the energy of shallow earthquakes on the ocean floor is trapped in the waveguide, which probably results from variations with depth of seismic-wave velocity in the oceanic crust and uppermost mantle. Instead of generating waves that dive into the mantle and are refracted upward, making them detectable at distant land-based stations, the quakes produce waves that can be recorded only on the ocean floor.

Comparisons of data from different hydrophones in the array made it possible to determine the locations of the earthquakes. Many fell within established regions of seismic activity, such as the Mariana Trench. Other earthquake epicenters delineated an arc-shaped zone never before recognized as seismically active. Kroenke and Walker believe these earthquakes signal the initial stage in the development of a new subduction zone: the fracturing of the ocean floor as one crustal block begins to underthrust another.

Kroenke points out that the new subduction zone makes tectonic sense. In the long gap between the Mariana Trench and the Tonga Trench, the convergence of the Pacific plate and the Indian-Australian plate to the southwest has for the past 10 million years been accommodated by a complex of three small subduction zones: the New Britain, San Cristóbal and New Hebrides trenches. All three have become obstructed: islands and oceanic rises on opposite sides of the trenches have collided as oceanic crust was consumed. It is reasonable to expect a new subduction zone to develop elsewhere to absorb the plate motions.

In contrast to the deep trenches elsewhere in the Pacific, no chasm on the sea floor marks the site of the proposed Micronesian Trench, although the sea floor near two islands, Banaba and Nauru, deepens slightly at the trench site. A seismic reflection profile across the region of seismicity reveals something more: faulting in the basement rock and an indication that the oceanic crust to the northeast of the fractures is beginning to dip downward. Such evidence suggests the youth of the Micronesian Trench. Kroenke and Walker estimate it is no more than between 500,000 and a million years old.



PROPOSED NEW SUBDUCTION ZONE extends 3,000 miles across the western Pacific from the southern end of the Mariana Trench to the northern end of the Tonga Trench. The New Britain, San Cristóbal and New Hebrides trenches, where the Pacific Ocean floor was subducted in the past, have become obstructed by islands and oceanic rises; earthquake epicenters (*dots*) now suggest that subduction is beginning in the new Micronesian Trench. The trench's youth would explain why the sea floor does not deepen markedly at the site.

Quick Change

Sirius, in the constellation Canis Major, is the brightest star in the night sky. It is seen as a point of cold blue-white light. Yet astrological and astronomical texts of the ancient Babylonians, Greeks and Romans consistently attribute a distinct reddish color to what appears to have been the same star. An early medieval manuscript has now been found to give additional and independent evidence that Sirius may have had a ruddy complexion at least up to the end of the sixth century.

The historic appearance of Sirius was deduced by Wolfhard Schlosser and Werner Bergmann of the Ruhr University in Bochum, West Germany. The text they studied was the only extant copy of an astronomical almanac that was compiled by Gregory of Tours in about A.D. 580 and was preserved 200 years later in an eighth-century Lombardian manuscript. Gregory wrote the almanac to give monasteries clear instructions on when to schedule nocturnal ecclesiastical services. He therefore included month-by-month lists of the times when certain constellations would appear above the horizon.

To compare Gregory's descriptions with the actual positions of the stars in the heavens above central France, the investigators went to a planetarium, where computer reconstructions of the night skies as they must have appeared more than 1,000 years ago could be projected. It soon became clear that, if Gregory was to be taken at his word, the star he called Rubeola or Robeola

FORD'S NEW DESIGN PHILOSOPHY:
TO ANSWER YOUR NEEDS
BY LISTENING TO
WHAT YOU WANT.

FORD TAURUS.

THE GOAL WAS TO BE BEST IN CLASS.

Even as the first designer or engineer began work, over 400 specific objectives were being established for Taurus.

The list was based on how you drive a wagon...what you want from it...what you need from it. The list included: How easy it is to fold down the rear seat; the effort required to open and close the doors; how comfortable the seat belts are to wear. The idea was this: If we

design each part, assembly and feature to make Taurus better to use, then we will have designed a better wagon overall.

THE RESULT IS TAURUS. *MOTOR TREND* CAR OF THE YEAR.

Motor Trend's Car of the Year award offers proof that Taurus' philosophy works.

Taurus is a front-wheel-drive wagon that is a delight rather than a chore to drive. It has a standard roof rack and optional seating for eight. It offers thoughtful features to driver and passenger alike. In total, it is a unique wagon where design and engineering come together in one functional unit.

BEST-BUILT AMERICAN CARS.

At Ford Quality is Job 1. A 1985 survey established that Ford makes the best-built American cars. This is based on an average of problems reported by owners in a six-month period on 1981-1984 models designed and built in the U.S.

3-YEAR UNLIMITED MILEAGE WARRANTY.

The new 3-year unlimited mileage warranty covers major powertrain components on 1986 Ford cars. Warranty is limited and certain deductibles apply. Ask to see the 3-year unlimited mileage warranty when you see your Ford Dealer.

Ford

All do-it-yourself fluid checks (oil level, power steering fluid, etc.) are arranged in front of the engine. They are also color-coded to make them easy to find, easy to check.

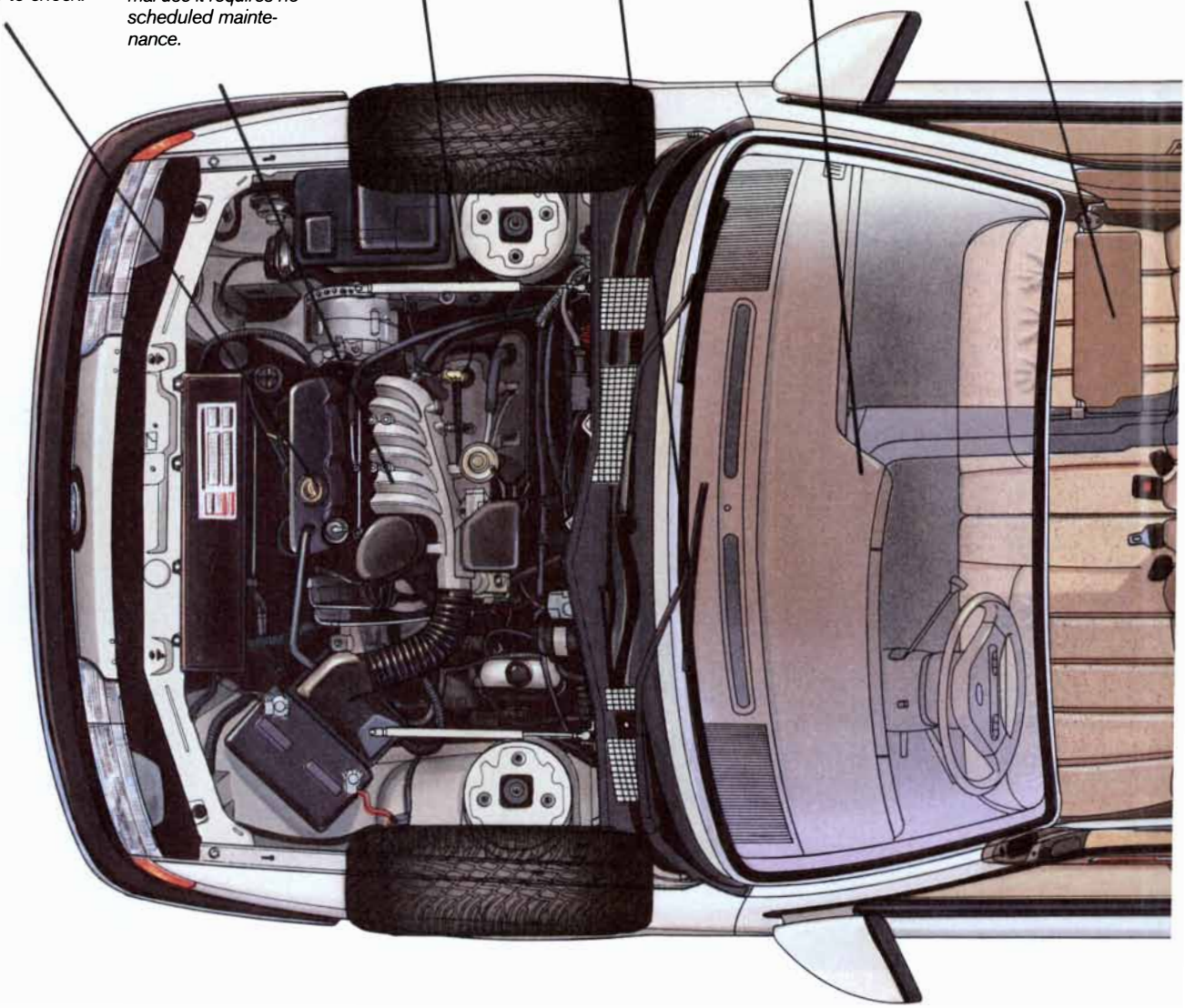
A new multiple port fuel injected V-6 engine designed for simplicity and responsiveness. Plus a new four-speed automatic transaxle so well designed, under normal use it requires no scheduled maintenance.

The hood features gas-charged struts which assist you to raise the hood and hold it open without a prop rod.

20" windshield wipers clear water nearly all the way to the driver's roof pillar to aid visibility.

Optional electrically heated Insta-Clear™ windshield clears itself of mist and frost.

Optional dual sun visors provide simultaneous front and side protection from the sun. Built-in extensions provide protection when the sun is low in the sky.



FORD'S NEW CONCEPT IN ENGINEERING BEGAN WITH YOUR NEEDS FIRST.

FORD TAURUS.

ENGINEERING... FOR PEOPLE.

The Taurus wagon was designed and engineered with a clear understanding of the special needs of a wagon buyer. It is designed to carry many different combinations of people and cargo.

Taurus seats feature a lower back support for comfort mile after mile. The seats are made of a high-density foam for pro-

Rear seat passengers have their own heat and fresh air ducts for balanced comfort.

Child proof rear door locks. Plus 37 other Lifeguard Design Safety Features.

Front seat head rests adjust up and down, forward and back. Seat belts are simple to use, comfortable to wear.

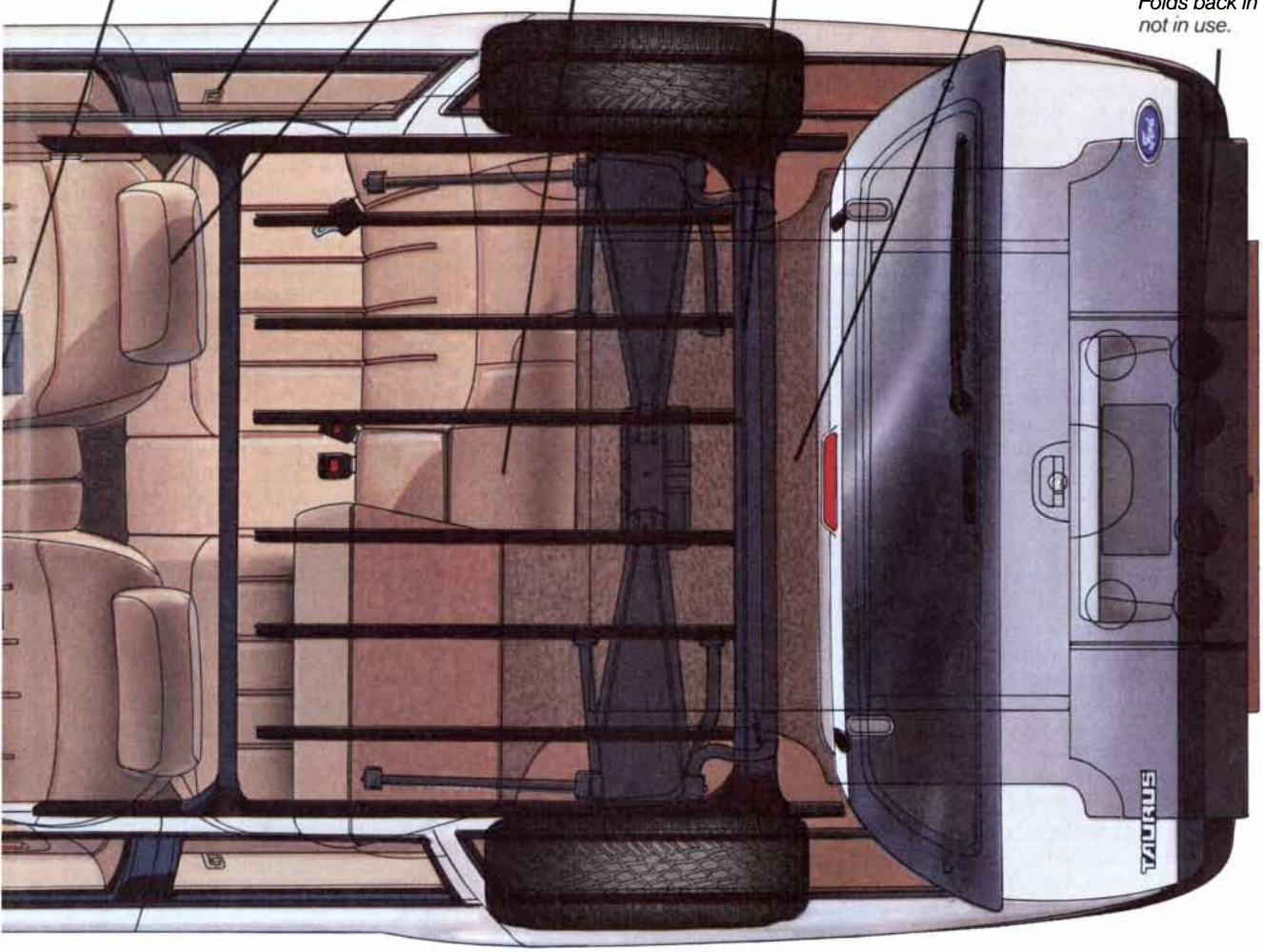
Optional seating for eight.

The standard roof rack adjusts for different length loads. Special brackets are available for skis or bicycles.

Room for up to 19 full-size grocery bags with the rear seat up. In addition, the load floor is low to make loading easier, and to provide rear visibility over normal loads.

The cargo floor features four rugged tie-downs to help secure smaller loads. An elasticized cargo net is also available.

An optional picnic tray folds out of the rear of the wagon. Folds back in when not in use.



Buckle up—together we can save lives.

longed comfort...less fatigue.

All controls are placed where your hands can easily find them. And the switches feature a "user friendly" design that allows you to read their function by touch, even at night.

Above, you will find a list of just a few of the function and comfort features which have been designed into the Taurus wagon.

There are more. Many more. But we hope these few will give you a sense of just how well the Taurus wagon is made. Have you driven a Ford... lately?

Ford

(meaning red or rusty) could not have been the currently orange-tinted star Arcturus, as had long been assumed. It had to have been Sirius.

Granting that Sirius shone red more than a millennium ago, what could have caused it to change? In a letter to *Nature* Schlosser and Bergmann suggest that a fundamental astrophysical transformation may have been responsible. Sirius is actually a binary star, consisting of a bright blue-white star (Sirius A) and a faint white-dwarf companion (Sirius B). White dwarfs are thought to be the remnants of collapsed red giants (such as Arcturus): old stars that increase in luminosity and size as they begin to burn helium and heavier nuclei instead of hydrogen. Could Sirius B have been a red giant as recently as 1,500 years ago?

Assuming that Sirius A had not changed in brightness since antiquity, Schlosser and Bergmann calculated how red and how bright Sirius B would have had to be to give the binary system a definite reddish color visible from the earth. The possible color-and-magnitude combinations place the old Sirius B well within the evolutionary track of normal red-giant stars.

Red giants do not become white dwarfs suddenly, however. The transition is thought to take very much longer than 1,500 years. Moreover, the transition is generally accompanied by at least one cataclysmic explosion. The only possible trace of such an event is the slightly higher than usual concentration of metals reported for Sirius A. It is conceivable that the metallic elements might indeed be the remnants of a nearby explosion related to the metamorphosis of Sirius B, but the apparent speed of the transition remains to be explained.

Ballistic Electrons

Can electrons pass between the atoms of a thin piece of semiconducting material as easily as they travel through a vacuum, so that they do not scatter? Some theorists think they can, and experiments now provide the first direct proof of such "ballistic," or collision-free, transport through a semiconductor. The result is of considerable interest because the transport mechanism could eventually be applied in the design of computer transistors that switch several times faster than present-day transistors.

The evidence for ballistic transport was provided by an experimental semiconductor device called a tunneling-hot-electron-transfer amplifier, or THETA. It was designed and built by Mordehai Heiblum, Marshall I. Nathan, David C. Thomas and Christina M. Knoedler of the IBM Thomas J. Watson Research Center.

As the investigators explain in *Physical Review Letters,* the device consists of three sequential layers of gallium arsenide: the emitter, the base and the collector. The emitter injects "hot" electrons (electrons that are not in thermal equilibrium with the device) into the base, which is made thinner than the other two layers to facilitate the ballistic travel of the injected electrons. The electrons then traverse the base to the collector, where their final energy can be measured.

Sandwiched between the emitter and the base is an extremely thin layer of aluminum gallium arsenide that serves as an electric-potential "barrier": electrons of a given energy can get into the base only by tunneling through the barrier. (Tunneling is a quantum-mechanical effect: a subatomic particle, in the guise of a wave, can "leak" through what would be considered an impenetrable barrier in classical physics.) Another layer of aluminum gallium arsenide separates the base from the collector. This barrier is thick enough to prevent any electrons at thermal equilibrium from penetrating the collector; only hot electrons have sufficient energy to surmount the barrier. The effective "height" of the barrier can be varied, however, by adjusting a bias voltage applied to the collector so that hot electrons whose energies are below a certain level are excluded. The device can therefore function as an electron-energy spectrometer. The number of electrons that make it through the device can be plotted as a function of their energy—or, equivalently, of the collector's bias voltage.

When the hot electrons were injected into the device, the IBM group observed a final energy distribution that peaked sharply at precisely the energy with which the electrons had been injected. This implied that many of the electrons had traversed the base and the collector barrier of the THETA device ballistically, and indeed calculations revealed that about half of the hot electrons had not lost any energy in collisions. Most of the other electrons had lost substantial energy in the base, however. The investigators are puzzled by this finding that one out of every two hot electrons in the device was practically stopped cold.

Chameleon Conductor

When they are cooled to sufficiently low temperatures, some materials such as mercury and aluminum become superconducting: they conduct electricity with no resistance. Now a team of workers has synthesized a material whose transition to the superconducting state can be induced by an externally applied high magnetic field.

The material, which is known as cerium-lead-3 ($CePb_3$), is a member of a small class of superconductors called heavy-fermion superconductors. The term heavy fermion reflects the fact that the electrons in these superconductors behave as if they were 100 to 1,000 times more massive than ordinary electrons. (Although an electron is a particular kind of fermion, or type of elementary particle, the materials would perhaps have been more aptly named heavy-electron superconductors.) Heavy electrons are not actually more massive than ordinary electrons. The apparent increase in mass is a consequence of the fact that the electrons are closely bound to atomic nuclei and therefore move as if they had additional mass.

The discovery is the work of seven investigators. They are C. L. Lin, Joseph Teter, Jack E. Crow and Ted W. Mihalisin of Temple University, James Brooks and Aly I. Abou-Aly of Boston University and Gregory R. Stewart of the Los Alamos National Laboratory. They describe their findings concerning cerium-lead-3 in *Physical Review Letters.*

The investigators found that the material has another interesting property. When it is cooled to temperatures slightly above absolute zero (zero degrees Kelvin, or minus 273 degrees Celsius), it becomes an antiferromagnet. In such a material groups of magnetic ions are polarized in opposite directions in a highly ordered way. The pattern of opposite polarizations could, for instance, take on the arrangement of a checkerboard.

The workers discovered that they could destroy the antiferromagnetic state and induce superconductivity by immersing the samples (some of which were as big as one centimeter long by one millimeter wide by one millimeter thick) in high magnetic fields. Specifically, at temperatures below about half a degree K. magnetic fields on the order of 15 tesla make cerium-lead-3 superconducting. (The magnetic field of the earth is roughly a ten-thousandth of a tesla.)

Although practical applications of the heavy-fermion superconductor are not in the immediate offing, Mihalisin notes that future work should improve understanding of the interplay between magnetism and superconductivity. It could even lead to the design of novel magnetic materials.

TAKE THE 3-VOLUME HANDBOOK OF ARTIFICIAL INTELLIGENCE (A $141.95 VALUE) FOR ONLY $4.95

when you join the Library of Computer and Information Sciences.

You simply agree to buy 3 more books—at handsome discounts—within the next 12 months.

Just completed, the massive, 3-volume HANDBOOK OF ARTIFICIAL INTELLIGENCE promises to become the standard reference work in the growing AI field.

Conceived and produced by leading scientists and researchers at Stanford University, with contributions from universities and laboratories across the nation, the *Handbook* makes available to scientists, engineers, students, and hobbyists who are encountering AI for the first time the techniques and concepts in this rapidly expanding computer universe.

The 200 articles cover the emerging issues, technical problems, and design strategies which have been developed during the past 25 years of research. The *Handbook* has been written for people with no background in AI; jargon has been eliminated; and, the hierarchical organization of the book allows the reader to delve deeply into a particular subject or browse the articles which serve as overviews of the various subfields.

The 15 chapters (5 per volume) include: the history, goals, and current areas of research activity; the key concept of "search"; research on "natural languages"; the design of programs that understand spoken language; applications-oriented AI research in science, medicine, and education; automatic programming; models of cognition; automatic deduction; vision and learning research; and, planning and problem solving.

The comprehensive HANDBOOK OF ARTIFICIAL INTELLIGENCE answers questions like:

- What is a "heuristic problem-solving program?"
- How do computers understand English?
- Can computer programs outperform human experts?

AND INCLUDES:

- Over 1,450 pages.
- More than 200 articles in 15 chapters.
- With numerous charts, tables, and schematics.
- Edited by Avron Barr, Paul Cohen and Edward Feigenbaum.

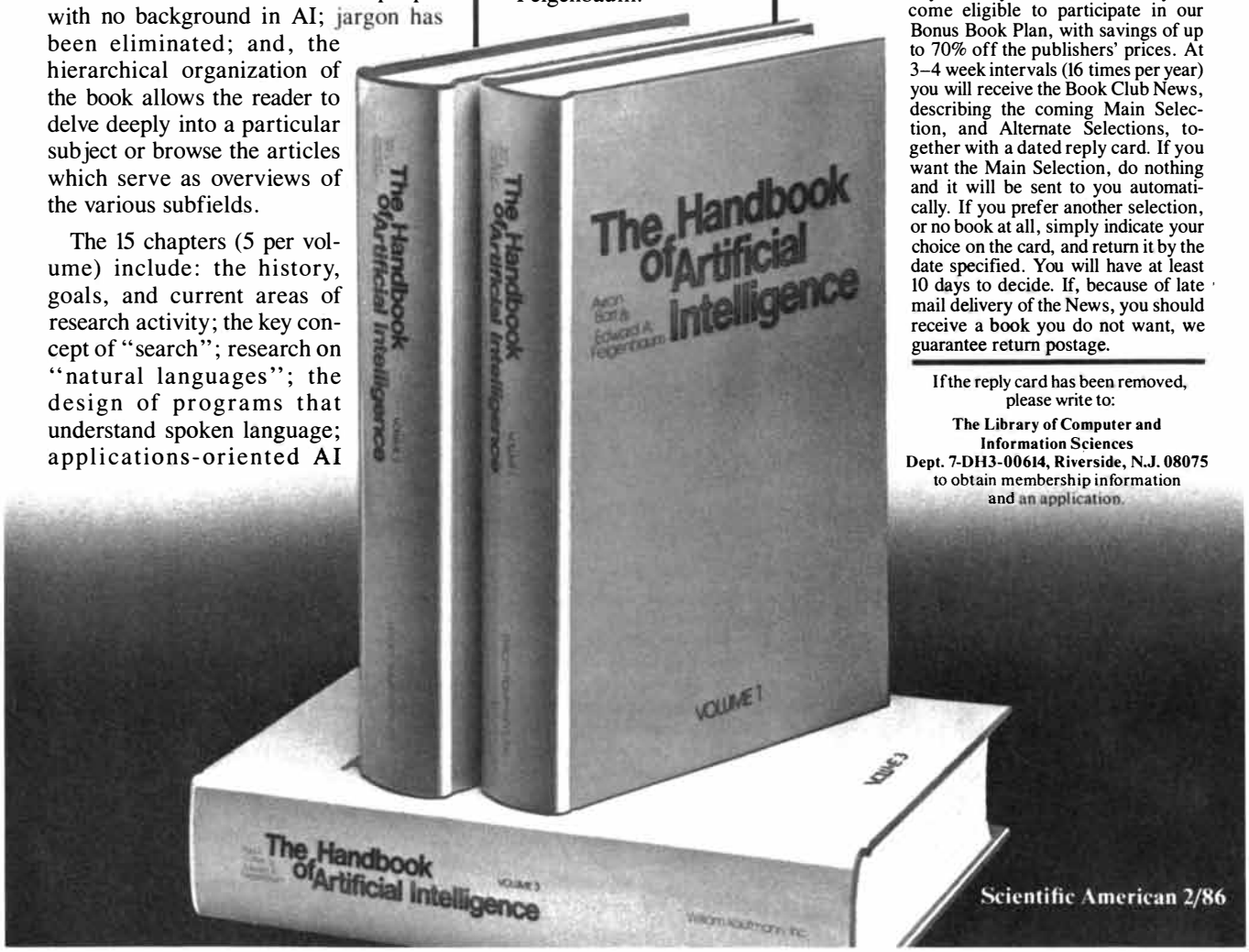
The Library of Computer and Information Sciences is the oldest and largest book club especially designed for the computer professional. In the incredibly fast-moving world of data processing, where up-to-date knowledge is essential, we make it easy for you to keep totally informed on all areas of the information sciences.

Begin enjoying the club's benefits today!

MEMBERSHIP BENEFITS: In addition to getting the 3-volume Handbook of Artificial Intelligence for only $4.95, when you join, you keep saving substantially on the books you buy. Also, you will immediately become eligible to participate in our Bonus Book Plan, with savings of up to 70% off the publishers' prices. At 3–4 week intervals (16 times per year) you will receive the Book Club News, describing the coming Main Selection, and Alternate Selections, together with a dated reply card. If you want the Main Selection, do nothing and it will be sent to you automatically. If you prefer another selection, or no book at all, simply indicate your choice on the card, and return it by the date specified. You will have at least 10 days to decide. If, because of late mail delivery of the News, you should receive a book you do not want, we guarantee return postage.

If the reply card has been removed, please write to:

The Library of Computer and Information Sciences
Dept. 7-DH3-00614, Riverside, N.J. 08075
to obtain membership information and an application.



Scientific American 2/86

Seismic Images of Plate Boundaries

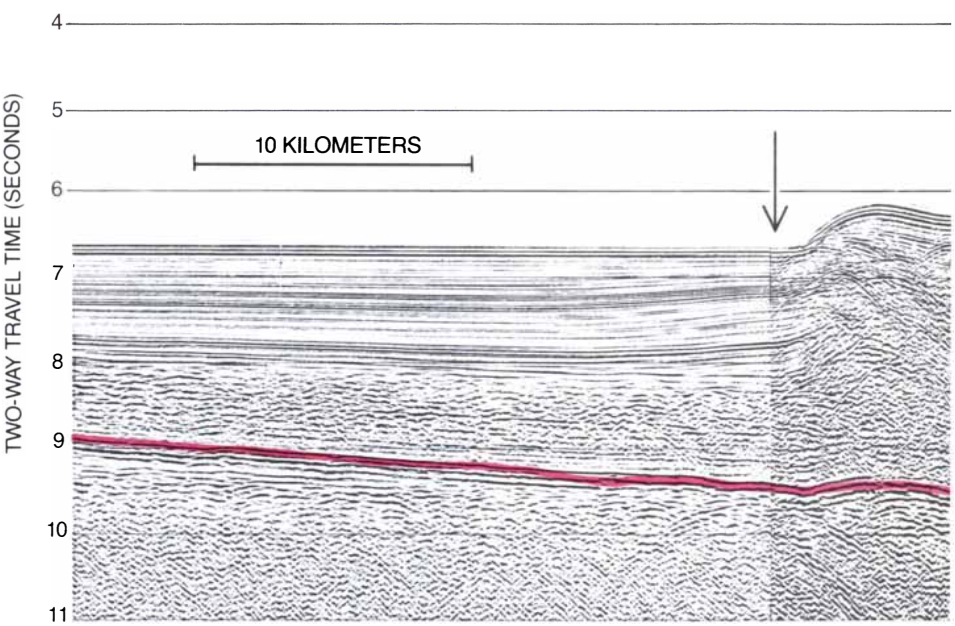
By bouncing sound off rock layers under the sea floor and recording the reflections with many detectors, structural images of the crust can be made at the boundaries where plates collide and rift apart

by John C. Mutter

How can geologists know what lies below the ocean floor? Over the past two decades the triumph of plate-tectonic theory has infused the question with special significance. The shifting plates that make up the earth's outer layer range in size from a few hundred to tens of millions of square kilometers, but they have an important feature in common: they are internally rigid. As a result they undergo major deformations only at their boundaries, and it is there that geology on a large scale really happens. Mountains and volcanoes are built where plates collide; ocean basins are opened and new sea floor is created where plates spread apart; earthquakes occur in both these regions but also at transform faults, where plates jerk past each other. To understand the mechanics of such interactions, investigators must study the deformation of the crust at plate boundaries. And since 70 percent of the earth's surface is covered by oceans, most of the boundaries are deeply submerged.

Unlike geologists on land, marine geologists and geophysicists cannot readily look at their subject matter. To be sure, spectacular images of the sea floor have been made from submersible research vehicles and by surface ships towing cameras on cables several kilometers long. Yet only an insignificant fraction of the total ocean floor has been observed visually, and that situation is not likely to change in the near future. Deep-sea drilling has provided essential information on rock formations under the sea floor, but again only at isolated points. Furthermore, the deepest holes drilled so far have penetrated less than two kilometers into the upper, sedimentary layers of the oceanic crust, and only one kilometer into the igneous basement layers that are formed by upwelling magma at midocean spreading centers.

Nevertheless, it is safe to say that the structure of the oceanic crust is known, at least in outline, in most parts of the world; for instance, the igneous layers are known to be on the average between five and seven kilometers thick. The knowledge is derived from the crust's response to sound. Sound waves set off by explosives or air guns deployed from a surface ship can penetrate to the base of the crust and beyond. At the same time some of the waves are reflected off the boundaries between successive layers of rock, because layers of different composition and density transmit acoustic (seismic) energy at different velocities. The reflected waves travel back to the surface, where they can be detected by hydrophones towed behind the ship.



SUBDUCTION ZONES, where two lithospheric plates converge and one plunges under the other into the asthenosphere, are readily recognized on multichannel seismic reflection images. Most of the crust (the top layer of the lithosphere) on the descending plate is carried into the asthenosphere. At many subduction zones, however, the upper, unconsolidated layers of sediment are scraped off the descending crust by the overriding plate. The sediments pile up, forming what is called an accretionary prism. An example is the Curacao Ridge off Venezuela (*above*), which includes the islands of the Netherlands Antilles. There the Caribbean plate is subducted by the South American plate. On the multichannel image the boundary between the unconsolidated sediments of the Venezuela Basin and the hardened sediments below is seen as a strong acoustic reflector (*color*). Known as A'', it continues largely intact for about 50 kilometers past the base of the Curacao Ridge (*arrow*), indicating that the sedimentary rock under it is being subducted. In contrast, the layers of soft sediment above the reflector are deformed into a chaotic mass; they are scraped off the hard underlying layer and pile up to form the ridge. On seismic images the position of a reflector is represented by the time a sound signal takes to travel from sea level to the reflector and back. The reflector's depth depends on the speed of sound in the rock above it.

By analyzing the reflections one can construct a profile—in effect an acoustic image—of the rock strata.

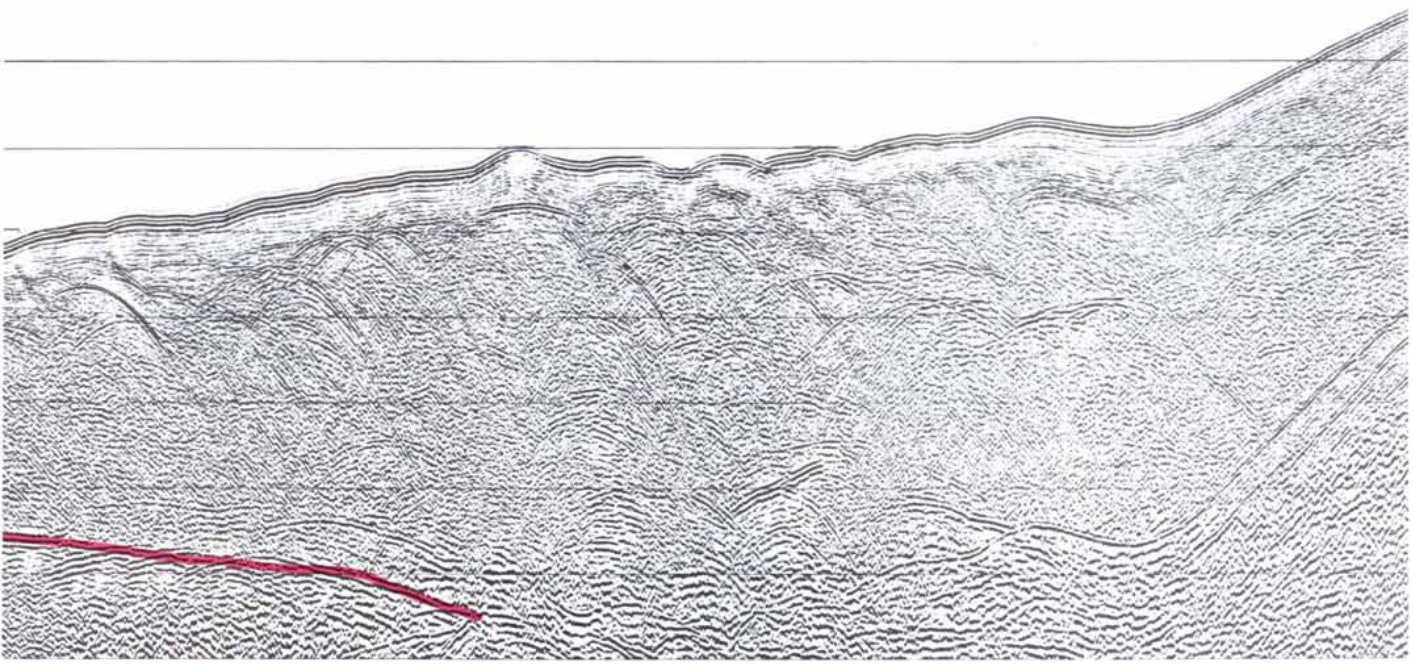
Reflection profiles have been made in some form since the 1950's. Since the late 1970's, however, they have grown considerably more sophisticated. The principle underlying the recent advances is simple: if a number of reflections are recorded from a particular point on the sea floor, the individual profiles can be combined to produce a single, enhanced profile of the crust at that point. The larger the number of reflections, the greater the resolution of the composite image. Modern seismic research vessels, including the one my colleagues and I at Columbia University's Lamont-Doherty Geological Observatory use for most of our investigations, deploy thousands of hydrophones. The equipment enables us to record many reflections from many points on the sea floor in a short time.

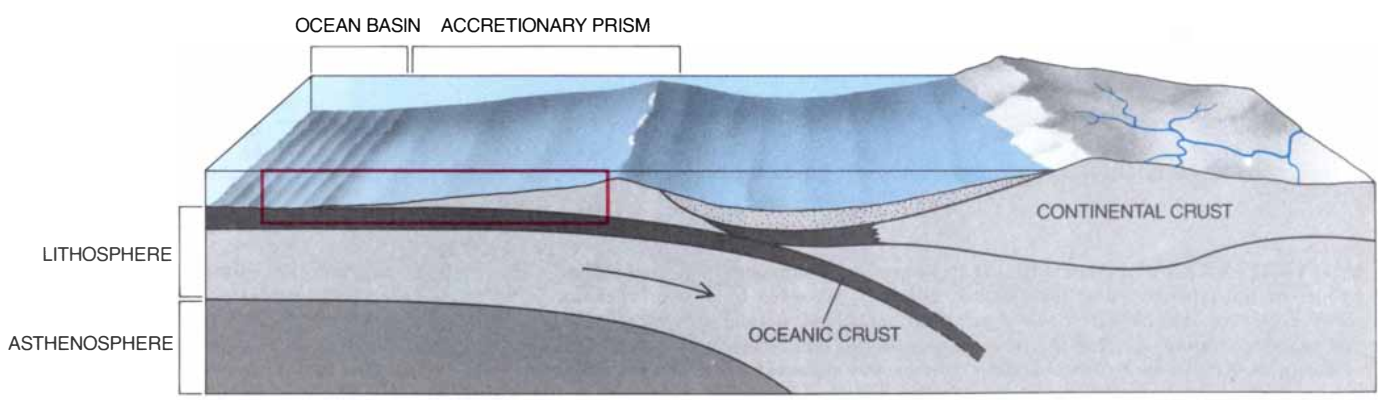
This technique, called multichannel seismic reflection profiling, has proved its worth in a variety of applications. It was initially developed by the oil-exploration industry; since the velocity of seismic waves is much slower in oil or gas than it is in rock, sub-sea-floor reservoirs often show up as strong reflections on crustal profiles. My own interest is in the application of multichannel seismic reflection profiling to the study of plate tectonics. Structural images of plate boundaries have confirmed some of the fundamental predictions of the theory and have led to a greatly refined understanding of plate interactions. In doing so they have also raised some provocative questions.

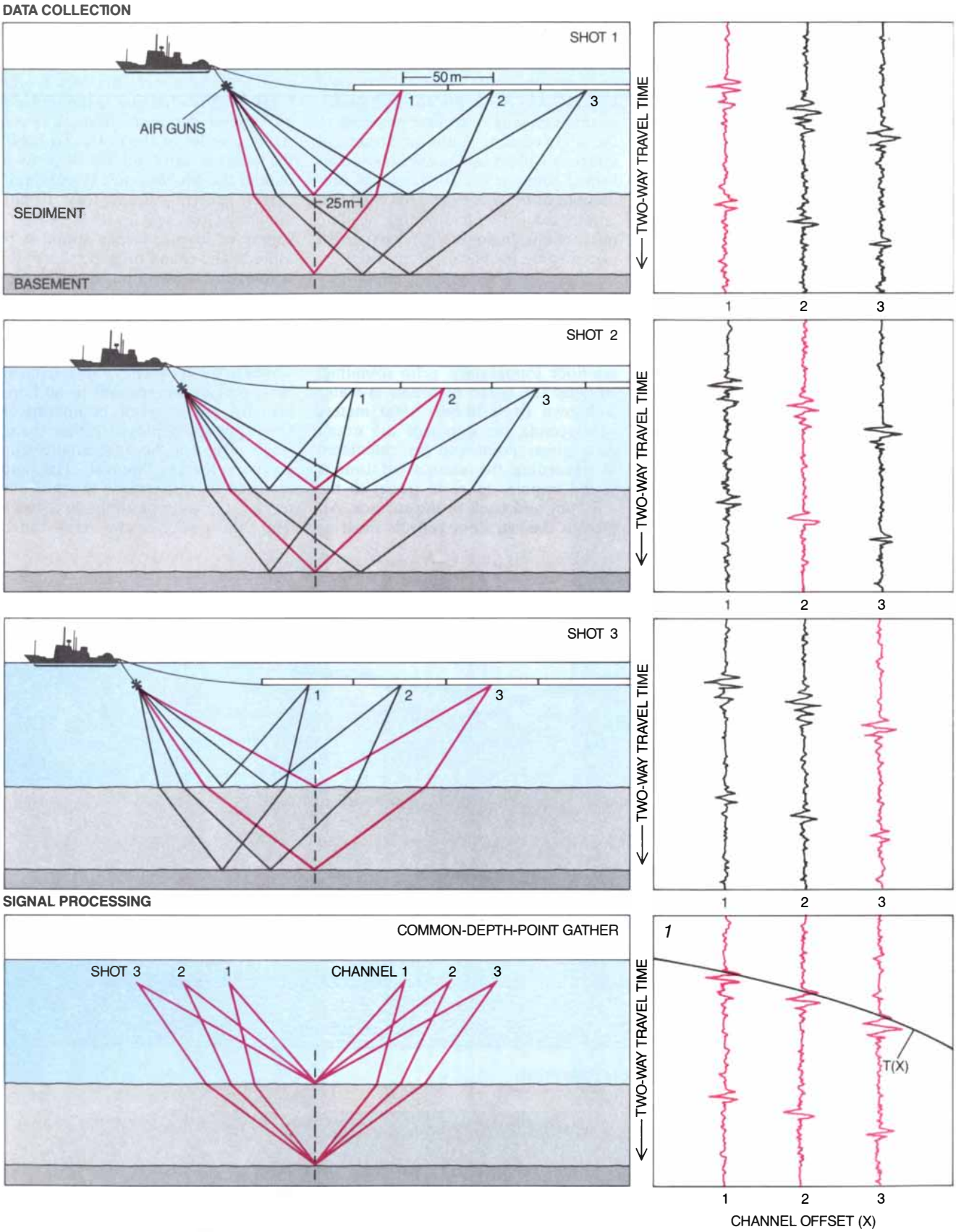
Collecting Seismic Data

Reflection profiling is derived from a well-known technique of mapping sea-floor topography: echo sounding. Because the speed of sound in water is known (it is about 1,500 meters per second), the depth of the ocean at a given point can be calculated by recording the amount of time a sound impulse takes to travel to the sea floor and back to the surface. Although the sea floor reflects most of the acoustic energy impinging on it, at low frequencies a significant fraction of the energy is refracted rather than reflected. The refracted waves penetrate into the underlying rock formations. That is what makes it possible to go beyond depth sounding and construct profiles of the crust. To profile the oceanic crust all the way to its base at the Mohorovičić discontinuity (which generally lies at least 10 kilometers below sea level) a powerful source of low-frequency sound is required. The sound must be sharp; if it is long and drawn out, the reflections from closely spaced rock layers will overlap and the profile will be blurred.

The most commonly used sound sources are air guns that emit sound waves in the frequency range from six hertz (cycles per second) to 60 hertz. Usually an array of comparatively small guns is deployed rather than a single large one, because smaller guns produce sharper sounds. The guns are fired simultaneously at regular intervals, typically of about 20 seconds. The total power of the array can be







MULTICHANNEL SEISMIC PROFILING produces enhanced images of the crust by combining sound reflections recorded by many receivers. The research vessel tows air guns and a long "streamer" containing a string of hydrophones grouped into receiver channels. After a shot each channel records sound reflected by the sea floor as well as sound refracted by the sea floor and then reflected by underlying rock layers. (Both the sediment and the basement rock may have internal reflecting boundaries.) Different channels record reflections from different points. The firing interval and the speed of the ship are coordinated so that a point sampled by the first channel on one shot is sampled by the second channel on the next shot; between shots the ship advances by a distance equal to

quite large. Commercial oil-exploration vessels often tow 30 or more guns, each discharging up to 4,000 pounds per square inch of air pressure. The sound from such a battery can sometimes penetrate as much as 15 kilometers into the earth.

Reflections from the crust are detected by a "streamer": a long plastic tube filled with a low-density fluid and a single line of hydrophones. The fluid gives the streamer buoyancy; together with the plastic, it also reduces the contrast in sound velocity between the hydrophones and the seawater, thereby reducing the extent to which the reflections from the crust are themselves reflected off the hydrophones without being registered. The streamer is towed at a fixed distance behind the air guns. It is ballasted so that it floats at a depth of 10 meters or so, which helps to isolate the sound detectors from the din of surface waves.

Streamers come in many sizes. The one belonging to Lamont-Doherty is 2.4 kilometers long and has a total of 2,400 hydrophones at one-meter intervals. Groups of 50 hydrophones are electrically linked to form 48 separate receiving channels spaced 50 meters apart. Each channel converts the acoustic signal it receives into an electric signal that is amplified, converted into digital form and stored on magnetic tape. Hence for every shot fired by the air guns 48 separate return signals are recorded. The signals have traveled along different paths from source to receiver and have bounced off different points on the sea floor, the longest path being the one to the last channel on the streamer.

The recording, or trace, made by a single channel after one shot is in itself a depth profile of the crust at the bounce point. Peaks on the trace mark the arrival time and intensity of reflections off the boundaries between rock strata. A single trace, however, reveals information only on the shallowest layers of the crust. The reflections from deep layers are very weak when they reach the surface, because the signal has been reflected off overlying boundaries and attenuated by its passage through many kilometers of rock. Furthermore, the signal is masked by background noise: the roar of waves, the drone of the ship's engines and the sloshing of water past the moving hydrophones. On an individual trace the noise often drowns out the signal.

Fortunately the noise is random, whereas the signal is not. If many profiles from a single bounce point are combined, the noise peaks will more often than not be out of phase and will therefore cancel. The signals, on the other hand, can be made to add in phase. This fact is exploited in multichannel reflection profiling to derive a composite image in which the signal is enhanced, the noise is suppressed and the reflecting layers in the crust are resolved more clearly.

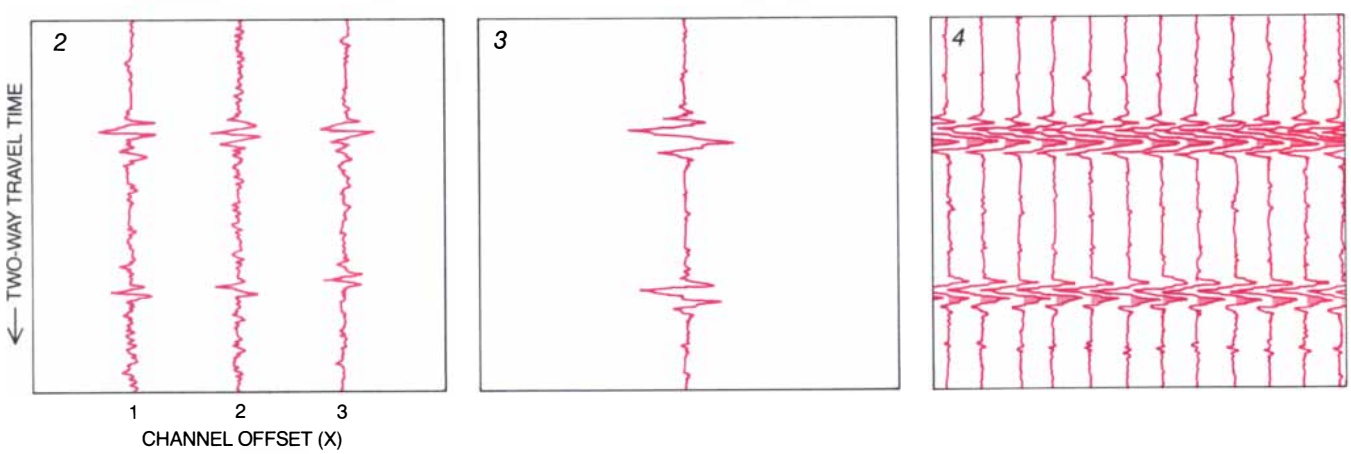
With a multichannel streamer it is not hard to record many traces from roughly the same bounce point: the firing interval of the air guns and the speed of the vessel are adjusted so that reflections off the same point are successively recorded by all the channels. Because the angle of reflection of a sound wave is equal to the angle of incidence at the reflecting boundary, the ideal bounce-point spacing is half the channel spacing. If the ship advances by this distance over the sea floor between shots, a bounce point sampled by one channel on the first shot will be resampled by the next channel on the next shot.

Signal Processing

When the ship returns from sea, the data it has collected are analyzed by computer. The first step is to extract from the individual shot records the traces corresponding to a particular point on the sea floor. The resulting group of traces is called a common-depth-point gather.

The traces cannot be summed immediately. Reflections from the same rock layer take different amounts of time to reach different channels on the streamer; the signal recorded by the last channel on the streamer, for example, has traveled much farther than the one picked up by the channel nearest the air guns. As a result the reflection peaks on different traces are initially not in phase. They must be brought into phase by shifting each peak's arrival time to a standard reference time: the arrival time it would have had if the sound waves had traveled vertically downward from the ship to the reflecting layer and then vertically back to the hydrophones. The reference time is called normal time.

This mathematical shifting of the traces is possible because there is a regular relation among the arrival times of corresponding peaks at different channels. The arrival time is a function of the distance of the channel from the air guns (the offset) and the seismic velocities in all the strata, including the water, above the reflecting layer. To simplify the calculations the velocities are replaced by a single quantity called the root-mean-square (RMS) velocity:



half the channel spacing. The first step in signal processing is to gather the recordings of reflections from a common depth point (*1*). A computer then identifies peaks corresponding to reflections from the same boundary; such reflections have traveled through the same rock layers, and their arrival times at different channels are related by an approximate formula that incorporates the velocity of seismic waves in those layers. The traces are shifted so that corresponding peaks have the same arrival time (*2*). When the traces are added, the reflection peaks are amplified, but random noise tends to cancel (*3*). The result is an enhanced profile of the crust at one point. To make a continuous image of a long section of crust, thousands of point profiles are put side by side and the peaks are darkened (*4*).

the square of the average of the individual velocities weighted by the travel time in each layer.

When an analysis of a common-depth-point gather begins, neither the arrival times of the reflection peaks nor the corresponding RMS velocities are known. The reflection peaks generally cannot be distinguished from noise on the individual traces. The computer finds both the reflection peaks and the RMS velocities through a trial-and-error search. In effect it advances down the arrival-time axis, stopping at regular intervals to scan across all the traces along the offset axis (the *x* axis). At each time interval the computer looks for a set of "coherent" peaks: peaks whose arrival times are specified by the same RMS velocity. Peaks with the same RMS velocity must have been formed by signals that have traversed the same rock layers and been reflected from the same sub-sea-floor boundary.

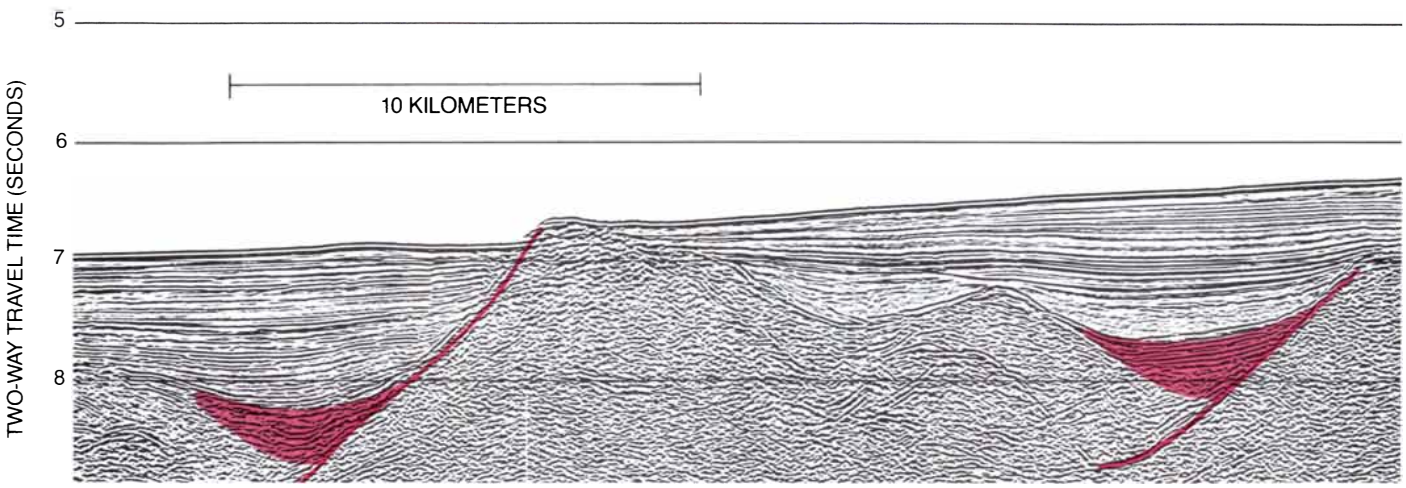
The computer repeats this procedure until it reaches the bottom of the traces. The RMS velocities are then used to shift all the arrival times to normal time. Finally the shifted traces are added together. Reflection peaks add in phase and are amplified; random noise tends to cancel. The result is an enhanced profile of the crust at one point. The images that accompany this article are made up of thousands of such profiles, placed side by side so that the darkened peaks run together and sometimes form continuous reflecting boundaries.

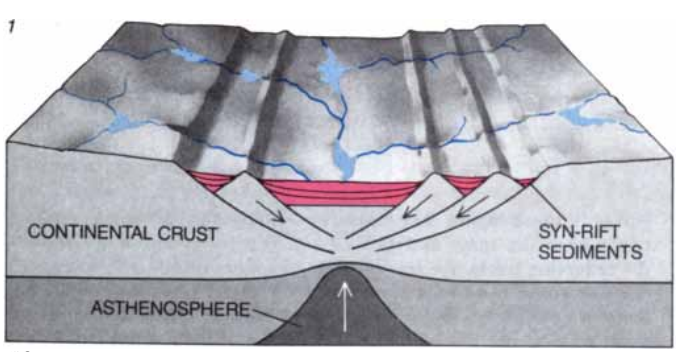
Commonly the position of a boundary is plotted not according to its depth but according to the two-way travel times of the reflections. To determine the depth of a reflector one must also know the seismic velocities in the individual rock layers above it. These can be estimated from the RMS velocities. In searching for a reflecting boundary the computer determines the RMS velocity of all the layers above the boundary. Because the speed of sound in seawater, the first layer, is known, the velocities of successively deeper layers can be calculated one by one from the RMS velocities. The analysis yields more than just the depth of the reflecting boundaries: the seismic velocity in a rock stratum is an important clue to its composition.

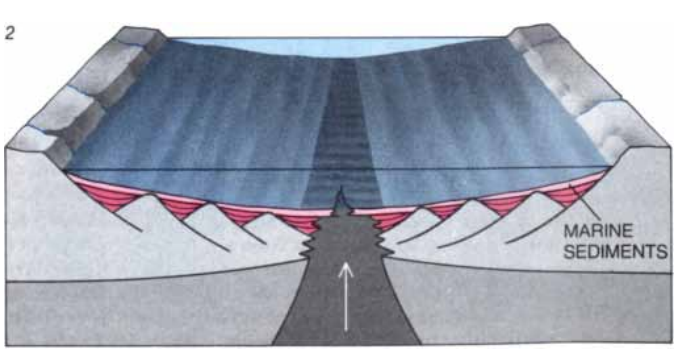
Refinements

A number of other steps can be taken to further enhance a multichannel image and filter out noise. Here I should like to mention only one refinement. Clearly the quality of a common-depth-point profile increases with the number of traces that go into it and therefore with the number of channels. This is particularly true of profiles of the deepest layers of the crust, from which the reflected signals are weak. With our 48-channel streamer we can usually detect the Mohorovičić discontinuity at the base of the crust; it is a relatively strong reflector, because the physical contrast between the crust and the upper mantle is great. Within the deep crust itself, however, there are few reflecting boundaries. The basement below the sedimentary layers consists of igneous rocks whose physical properties are nearly uniform or change only gradually with depth. Forty-eight channels are not enough to resolve the internal structure of the igneous basement, or even enough to track the Moho continuously at its deepest points.

One way of adding channels is simply to make the streamer longer. With a longer streamer, however, one is more likely to encounter technical difficulties that can delay the collection of data. Peter Buhl and his colleagues at Lamont-Doherty have developed a way of simulating a long streamer by teaming up two ships, each equipped with air guns and a streamer. The two ships steam along the same line, one behind the other, alternately firing their guns. Each ship records reflections not only from its own shots but







also from those of the other ship. If both ships tow 2.4-kilometer, 48-channel streamers, and if the trailing ship keeps 2.4 kilometers behind the end of the leading ship's streamer, the two streamers together simulate a single 7.2-kilometer streamer with 144 channels. The technique is called wide-aperture common-depth-point profiling.

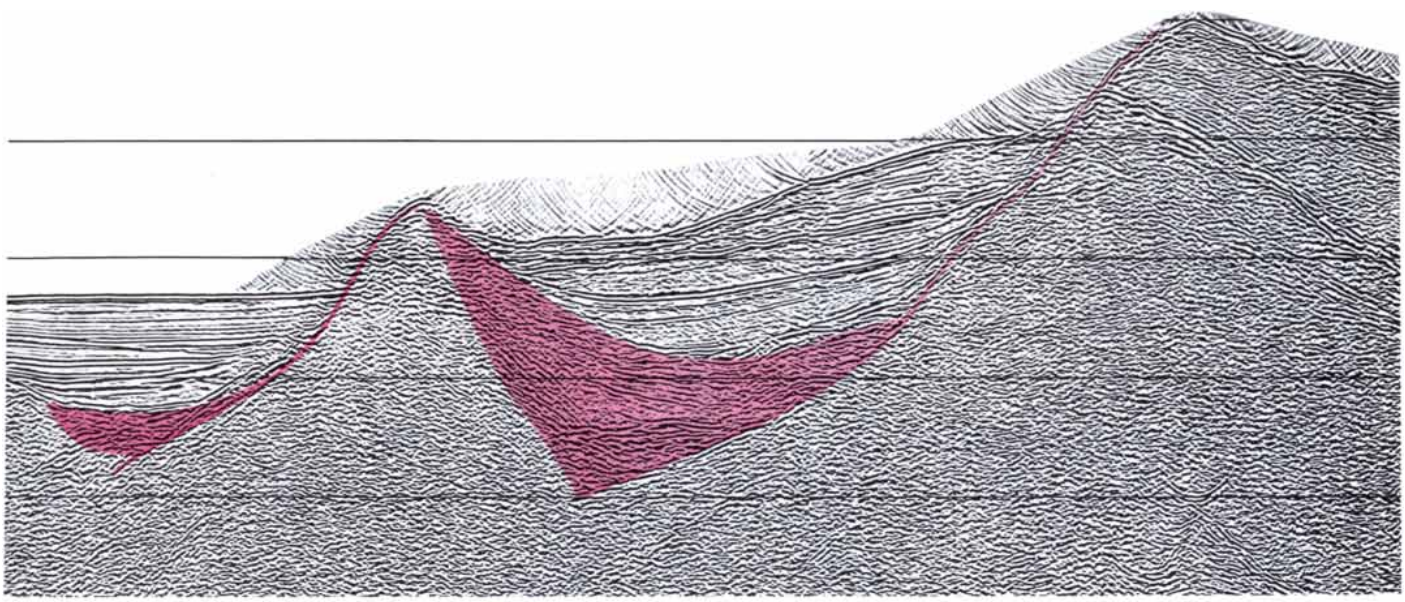
With one ship or with two, the cycle of firing the air guns and recording reflections can proceed continuously as long as stormy weather does not force the crew to haul in the streamers. In a matter of days, with the ships steaming ahead at five knots, enough data can be collected to make an image that tracks crustal structure for hundreds of kilometers under the ships' path. Multichannel seismic reflection profiling is therefore well suited to the task of searching the crust for the large-scale effects of plate motion.

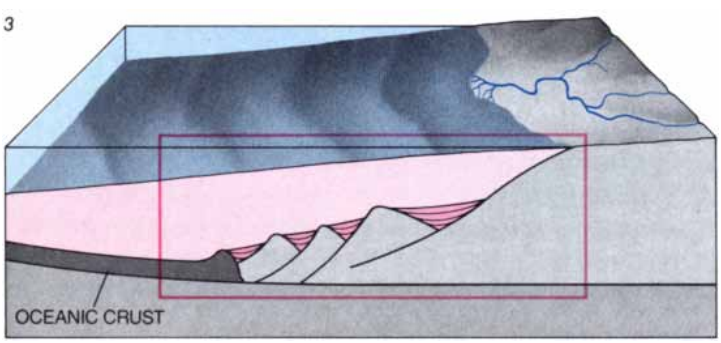
Subduction Zones

The plates, which together make up the rigid lithosphere, move over a partially molten layer of the mantle called the asthenosphere. The crust is only the top part of the lithosphere; the plates are from 75 to 150 kilometers thick and include a significant part of the upper mantle. Where two of them converge one plunges under the other and is subducted into the asthenosphere. What happens to the crust? Usually most of the crust plunges into the asthenosphere with the subducted plate. Some of the upper sedimentary layers may, however, be scraped off the descending plate by the overriding plate. In particular, poorly consolidated, waterlogged material on the ocean floor may be skimmed off the stronger underlying rock much as snow is skimmed off a hard road by a plow. The sediments pile up in front of the overriding plate, forming what is called an accretionary prism.

Reflection profiles of the Caribbean made by John W. Ladd of Lamont-Doherty vividly illustrate the process [*see illustration on pages 66 and 67*]. At the continental margin off Venezuela the Caribbean plate dives under the South American plate. The crust entering the subduction zone consists of undeformed, layered sediments, the deepest of which are about 85 million years old, deposited on a basement of igneous basalt. On the profiles there is a prominent reflector called A″ at a depth of about seven kilometers below sea level, equivalent to an age of 37 million years. From drilling samples it is known that the sedimentary layers below the A″ horizon are highly lithified formations of chalk and limestone. Above the reflector, on the other hand, there is a poorly consolidated sequence of clays and turbidites (sediments derived from material eroded off the land and the continental shelf).

The beginning of the subduction zone is marked by the gently sloping Curaçao Ridge. The A″ horizon continues landward under the ridge for at least 50 kilometers. It appears to be slightly distorted, but its overall structure is easily recognized. Sediments under the horizon seem to continue intact into the subduction zone, presumably to be destroyed in the mantle along with the crustal basement and the rest of the subducted lithosphere. In contrast, sediments above the A″ horizon undergo an abrupt change immediately under the base of the ridge: previously undisturbed layers become a chaotic, featureless mass. The Curaçao Ridge displays little evidence of internal structure. It appears to be just





CONTINENTAL RIFT is thought to form through stretching and thinning of the lithosphere over millions of years. When the brittle upper layers of the continental crust are stretched, blocks of crust slide down curved fault planes, producing a series of ridges and troughs (*1*). As the crust subsides during the rifting process, the troughs are filled by "syn-rift" sediments eroded off the fault blocks and the continents and deposited in lakes. If stretching continues, the thinned plate separates into two spreading plates (*2*) and a new ocean basin forms (*3*). The syn-rift sediments are overlain by marine sediments. On a multichannel seismic image of the continental margin off the coast of Spain, curved faults caused by stretching are clearly visible (*dark color*). The syn-rift layers (*medium color*) in the troughs converge toward the top of the blocks, indicating they were deposited while the blocks were sliding. The overlying marine sediments were deposited after rifting and are undisturbed.

a pile of sediments swept up by the South American plate.

Other accretionary prisms are built up more systematically. An example is the one at the Nanki Trough, a part of the Japan Trench, where the westward-moving Pacific plate plunges under the Eurasian plates. (The deep oceanic trenches of the western Pacific are all subduction zones.) Most of the sediments at the Nanki Trough seem to be carried some distance into the subduction zone. Water is squeezed out of them, and they are compressed and strengthened. Eventually a wedge of compressed sediment fails along a fault that dips slightly from the horizontal. It thereby becomes separated from the unconsolidated sediment that lies behind it on the descending plate. This newer sediment is then thrust along the fault plane under the compressed wedge. It too is compressed until a second wedge breaks off it. In this manner wedge after wedge of sediment is driven into the bottom of the accretionary prism—the new material below the old—and the prism is gradually lifted and tilted. On the reflection profiles of the Nanki Trough the thrust faults between the sedimentary wedges form strong reflecting boundaries.

Both at the Nanki Trough and in the Caribbean some of the sediments on the descending plate are subducted. At other subduction zones all the sediments seem to be accreted by the overriding plate. At still others, none of the sediments escape subduction; indeed, there is evidence in some cases that the descending plate may actually erode material off the front of the upper plate. Multichannel reflection profiles have shown that no two subduction zones are wholly alike, which naturally raises a question: What accounts for their differences? The fate of the crust may depend on the speed at which the two plates converge, but that factor alone cannot explain the observed variety of crustal structures. So far no entirely satisfactory answer to the question has been found.

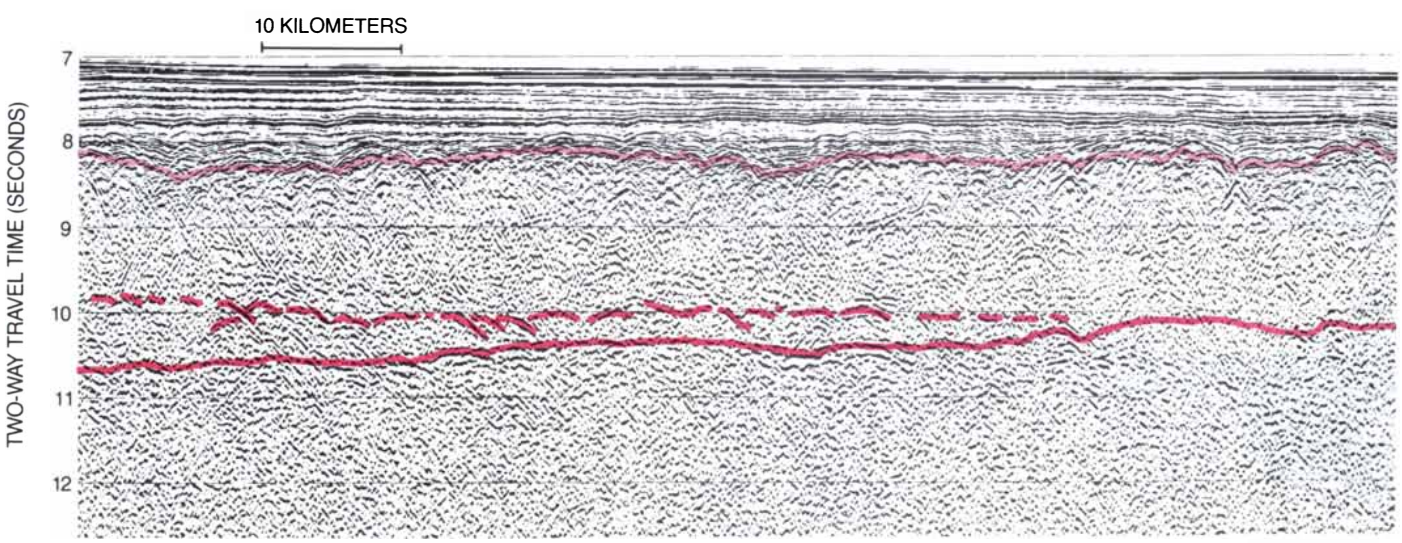
Rifts

An even greater puzzle about plate convergence is the question of how it begins. There is little evidence on which to base an answer: in every case most of the structures created during the period when a pair of plates began to converge have been subducted. In contrast, the formation of a divergent plate boundary is relatively well documented. To understand how North America and Europe rifted apart, for example, one can look at the margins of those continents, where evidence of the process is preserved.

Rifting involves the entire lithosphere, not just the crust. It is generally believed that a plate does not rupture instantaneously; rather, it is stretched and thinned over a period of millions of years. The evidence for stretching is often indirect: it comes from the observation of sedimentary basins such as the North Sea. According to theoretical calculations by D. P. McKenzie of the University of Cambridge, sedimentary basins can form when the bottom of a thinned plate is replaced by hot material from the asthenosphere. As the hot material cools it contracts, causing the lithosphere to subside. In this view many sedimentary basins are either failed continental rifts or rifts in the making.

If the lithosphere is slowly stretched before rifting, more direct evidence of the process should be detectable in the structure of the crust at continental margins. Lucien Montadert of the Institut Français du Pétrole, David Roberts of British Petroleum in London and their colleagues have found such evidence on reflection profiles of the margins off France, Spain and Portugal [*see illustration on preceding two pages*]. The basement formations there are broken into a series of jagged blocks. On their landward side the blocks are bounded by fault planes that dip steeply near the surface but then curve and become nearly horizontal with depth—precisely the type of faulting expected in brittle material subjected to severe extensional stress. As Europe and North America were pulled apart, Montadert and Roberts argue, the fault blocks slid down the curved planes and formed a series of jagged ridges and troughs.

Sediments deposited in the troughs confirm that the lithosphere was slowly stretched before it rifted. The layers converge toward the top of the blocks, indicating that they were deposited while the blocks were still moving. Drilling samples show that these "syn-rift" sediments include debris eroded off the basement blocks. More important, the samples make it possible to date the sediments: they were deposited over a period of about 10 million years just before the oldest oceanic crust in the Atlantic was formed. (The formation of oceanic crust at a new midocean spreading center indicates that rifting is complete and that the rifted plates have begun to diverge.) Judging from the syn-rift sediments, then, the divergence of North America and Europe, which off France and



OCEANIC FRACTURE ZONES appear on multichannel seismic images as regions of thin crust. Where a fracture zone dissects a midocean ridge the two spreading plates slide past each other along transform faults. Although such a fault may be short, the anomalous crust of the fracture zone may extend along the fault line for thousands of kilometers. The profile shows the Blake Spur Fracture

Spain began some 110 million years ago, was preceded by at least 10 million years of stretching. After rifting had ceased and the continental margin had subsided below sea level, the syn-rift sediments were themselves overlain by flat layers of sediment attributable to the steady snow of pelagic plant and animal life on the sea floor.

The connection between lithospheric stretching and plate divergence is not always so straightforward. Off Norway, for instance, the crust on the continental margin shows clear indications of having been stretched, but the stretching seems to have taken place nearly 80 million years before sea-floor spreading began. The block-faulted basement and the syn-rift sediments are about 135 million years old, whereas the oldest oceanic crust in the Norwegian-Greenland Sea dates from only 59 million years ago. In the intervening period flat beds of sediment accumulated over the syn-rift sequence. These beds were barely deformed by the completion of the rift, which must not have required further stretching. Apparently the lithosphere stretched near its breaking point, stayed that way for 80 million years and then ruptured catastrophically.

Sea-Floor Spreading

Once a plate has rifted apart, the two new plates diverge. Molten rock rises up from the asthenosphere to fill the gap and form new lithosphere. A small fraction of the material erupts at the surface as lava; the rest cools and attaches to the edges of the spreading plates. Multichannel profiles of continental margins, including the Norwegian one, suggest this volcanic activity is most intense in the first few million years of sea-floor spreading.

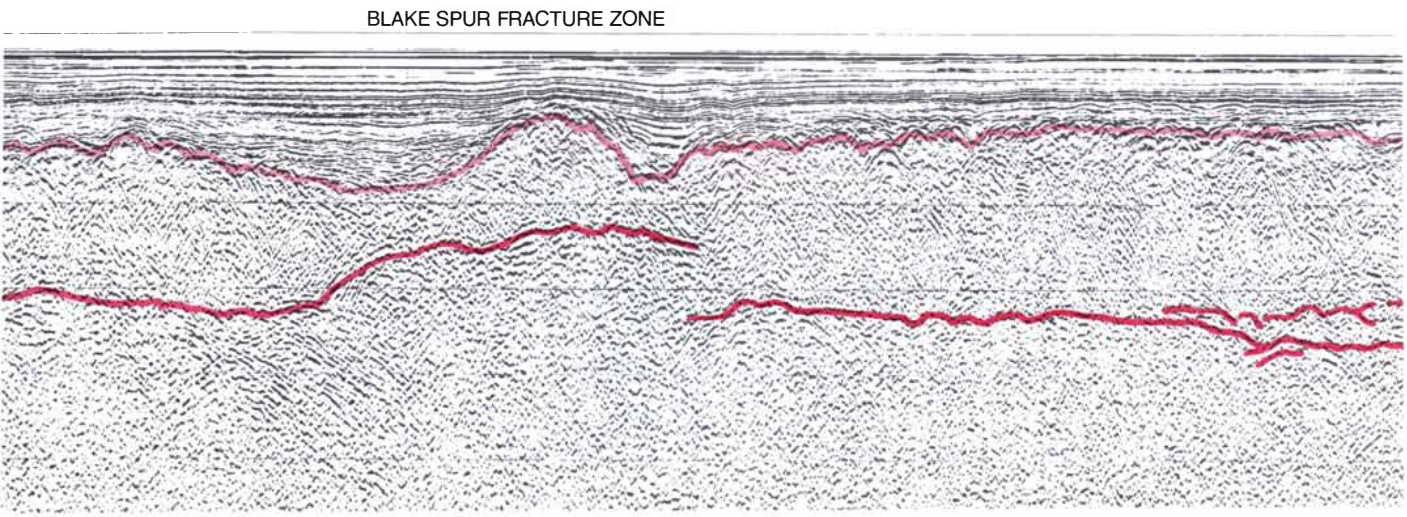
As the lithosphere spreads away from a midocean ridge it cools further, contracts and subsides. The deepest abysses in the Atlantic are found adjacent to the continental margins, where the oceanic lithosphere is oldest and has therefore cooled the longest. Off the Norwegian margin, however, the sea floor is anomalously shallow. Drilling samples show that the crust in the shallow areas is indeed oceanic and not continental; seismic profiles indicate it is about twice as thick as expected for crust that (along with the rest of the lithosphere) has been cooling and contracting for nearly 60 million years. When the thick crust was formed, during the first three to five million years of spreading, magma must have been welling up at the ridge at a rate much higher than normal.

Indeed, multichannel reflection profiles of the Norwegian margin suggest the upwelling was intense enough to lift the ridge above sea level. The profiles reveal a distinct internal structure in the upper part of the basement: it consists of a stack of curved layers that dip slowly toward the sea. Manik Talwani of Rice University and I have proposed that the layers were formed by successive volcanic flows from a spreading center that must have been exposed. (Lava erupting from a submerged spreading center cools and solidifies into "pillows" before it can flow far.) The hypothesis explains the seaward dip of the layers: most of the lava in each flow remained near the site of the eruption, and so the load placed on the underlying layers by each flow was progressively greater toward the spreading center. Much the same thing is happening today in Iceland, where the Mid-Atlantic Ridge lies above sea level. Lava beds in eastern Iceland have a structure similar to that of the crust at the Norwegian margin.

By analyzing multichannel seismic profiles, Karl Hinz of the Federal Institute for Geosciences and Natural Resources in Hanover has found evidence for dipping lava sequences at continental margins all over the world, in regions as diverse as Antarctica and India. It is becoming clear that the early evolution of many spreading centers was marked by intense volcanism. In most cases the volcanic activity probably diminished after a few million years. The spreading center, no longer buoyed by upwelling magma, sank below sea level. The lava still erupting from it thereupon began to pile up in the pillows typical of midocean ridges.

Magma Chambers

The detailed structure of spreading centers is a matter of considerable debate among geologists. It is widely assumed that a midocean ridge is fed by a persistent chamber of magma in the crust. Studies of ophiolites (fragments of oceanic crust thrust onto land by plate-tectonic movements) offer the primary support for this assumption. The bottom sections of ophiolites often consist of layered gabbros: coarse-grained plutonic rocks that are thought to have cooled and crystallized slowly at a depth of several kilometers. The formation of layered gabbros is hard to explain without positing a magma chamber several kilometers below the ridge. In such a chamber minerals could crystallize out of the melt in layers, which would then be carried away



Zone about 1,000 kilometers east of Florida. The light colored line is the boundary between sediments and the igneous basement, which was formed at the Mid-Atlantic Ridge 130 million years ago. Near the fracture zone the crust thins from below. The Moho (*dark colored line*), which marks the base of the crust, rises sharply, and the internal reflector called Horizon *R* (*broken colored line*) disappears.

on the diverging plates. Periodically the chamber would be replenished with magma from the asthenosphere.

Norman H. Sleep of Stanford University and Bruce R. Rosendahl of Duke University have derived a thermal model for spreading centers that also implies the presence of magma chambers. Not surprisingly, the model predicts a large magma chamber under a ridge where lithosphere is being created at a fast rate, say about five centimeters per year, and a narrow chamber under a slow spreading center. The magma chamber under a fast spreading center is expected to be a wedge-shaped structure, perhaps as much as 20 kilometers wide at its base and narrowing to a few kilometers at the top.

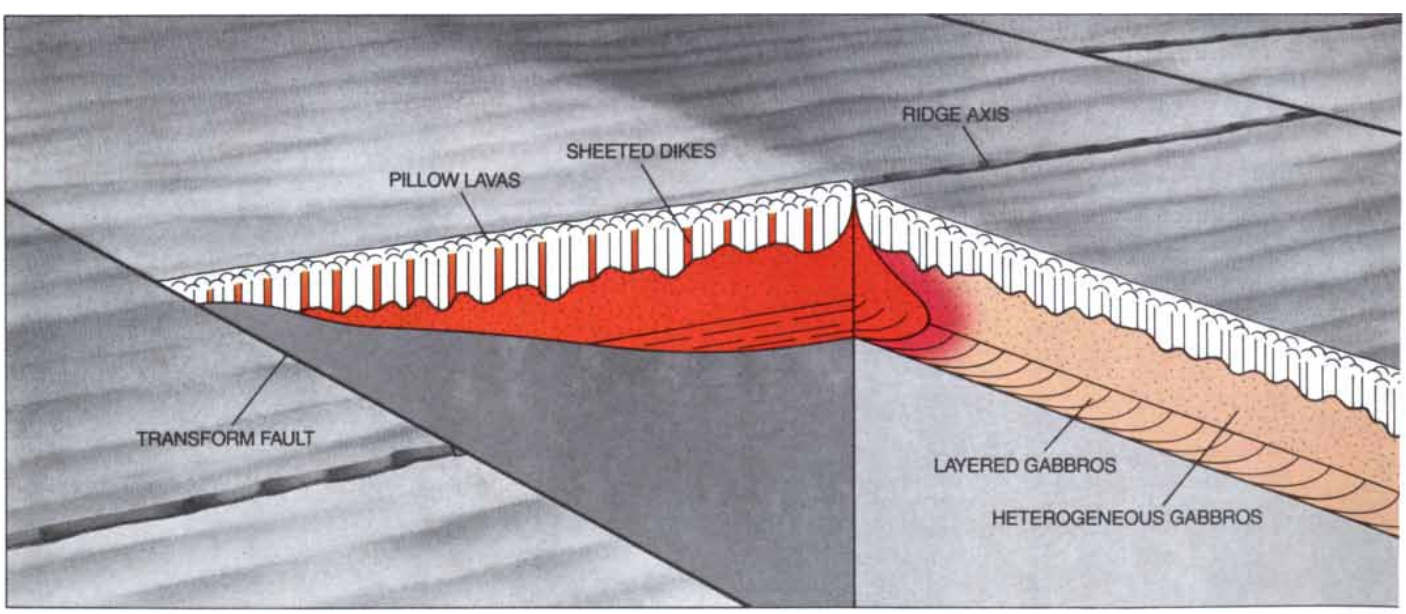
A large magma chamber should in principle be detectable on seismic profiles, because the magma would have a much lower seismic velocity than the solid rock surrounding it. Hence on a reflection profile one should see reflections off the top of the chamber. A second type of evidence may be found with a different seismic technique, one that depends on detecting refracted sound waves. In most parts of the crust seismic velocity increases with depth, and a sound wave that penetrates the crust without being reflected is refracted toward the horizontal at each successive boundary, until eventually it propagates horizontally along a boundary. Finally it is refracted back toward the surface, where it can be detected; the deeper the layer in which the signal traveled horizontally, the greater the distance at which it is detected. A sound wave entering low-velocity magma, on the other hand, is not bent toward the horizontal. Instead it is refracted downward, away from the detectors. A magma chamber may therefore reveal itself as a "shadow zone" on a profile assembled from refracted signals: at the detector distance corresponding to the depth of the chamber, the strength of the signal is dramatically reduced.

Using instruments on the sea floor and explosives, John A. Orcutt and his co-workers at the Scripps Institution of Oceanography detected a shadow zone below the East Pacific Rise, one of the fastest spreading centers in the world. Later a Lamont-Doherty group led by Thomas J. Herron recorded a distinct reflector on a multichannel profile of the ridge crest near where Orcutt did his experiment. The depth of the reflector corresponds almost exactly to the top of Orcutt's shadow zone. Taken together, these two results are compelling evidence for the presence of some type of magma chamber under the East Pacific Rise.

The magma chamber, however, does not appear to be nearly as large as expected for a fast spreading center. Last summer Orcutt, Robert S. Detrick of the University of Rhode Island, Thomas Brocher of the Woods Hole Oceanographic Institution, Peter Buhl and I made several crossings of the rise, collecting both reflection and refraction data with two-ship, high-resolution methods. A preliminary analysis of the data has found evidence for a shadow zone, but only directly under the ridge crest. The analysis suggests the magma chamber under that section of the East Pacific Rise is no more than a few kilometers wide. This sharply contradicts both the ophiolite studies and Sleep's theoretical model, which imply that the chamber should be on the order of 20 kilometers in diameter. It is not yet clear how the conflict between geological inference and seismological observation is to be resolved; Sleep has argued that seawater circulating through the crust may be cooling and shrinking the magma chamber.

Ancient Crust

A constant problem in trying to generate clear images of the crust under a midocean ridge is the extreme ruggedness of the terrain: much of the sound energy that would penetrate a smooth surface is scattered in random directions by a rugged one. A way to avoid this problem and yet focus on the same basic question—how oceanic crust is formed—is to study the crust away from a ridge, where it has acquired a smooth blanket of sediment. The structure of the crust provides information on conditions at the ridge, albeit the conditions that prevailed many millions of years ago. Applying the wide-aperture technique, my colleagues and I have made multichannel reflection profiles all the way from

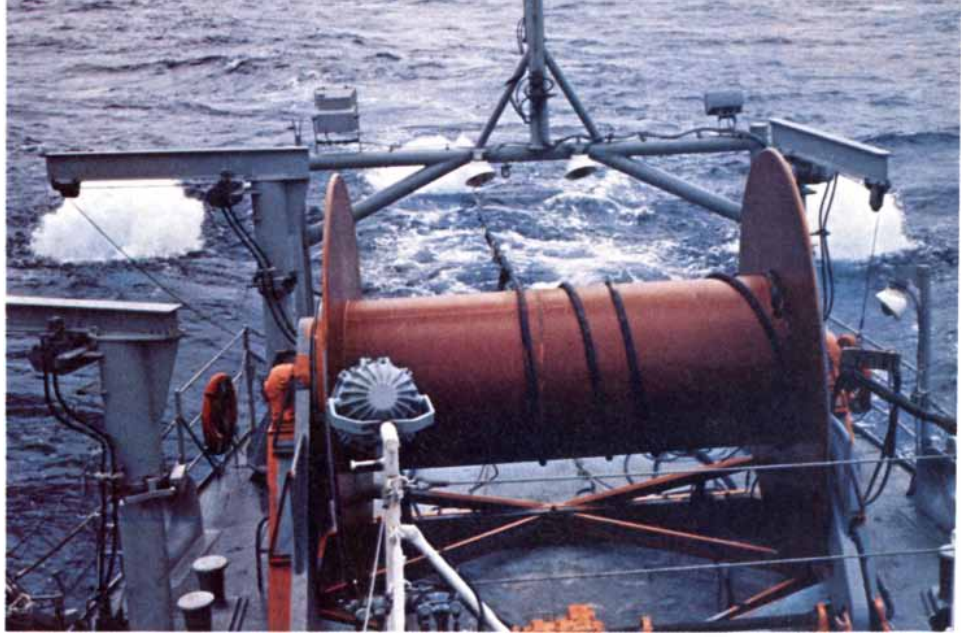


MAGMA CHAMBER under a midocean ridge may explain the thin crust at fracture zones. Most oceanic crust is thought to form through cooling and crystallization of magma within such a chamber; only a small part of the magma is extruded at the surface as pillow lava. Horizon *R* may be the boundary between the heterogeneous gabbros, which form as magma cools against the side walls and roof of the chamber, and the layered gabbros, which form from mineral crystals that have settled to the bottom. The Moho represents the base of the magma chamber. Each ridge segment bounded by fracture zones may be fed by a separate magma chamber cen-

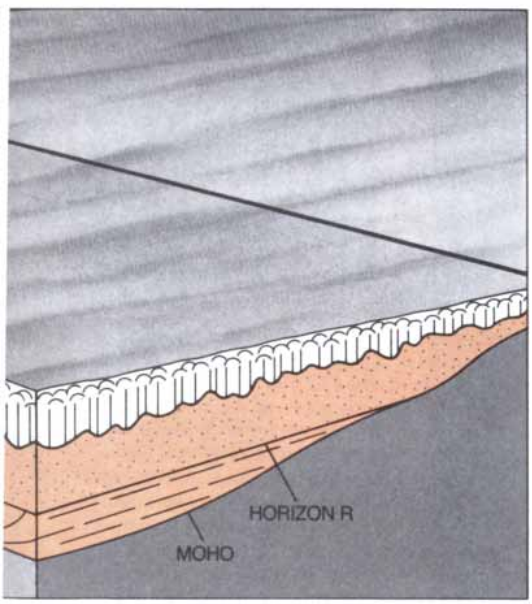
Florida to the Mid-Atlantic Ridge, from the oldest to the youngest Atlantic crust. The profiles yield indirect evidence for the presence of a magma chamber under the ridge.

On our profiles the Moho can be tracked continuously over great distances at a nearly constant depth of from 12 to 15 kilometers below sea level. Its reflectivity is also fairly uniform. Whatever the process that forms the lower crust, in the North Atlantic it appears to have been relatively invariant for the past 100 million years. This supports the notion that the spreading center is fed by a steady-state magma chamber; if the crust were produced by episodic injections of magma from the mantle, as some workers have suggested, the Moho would probably not be so uniform and continuous.

Another feature of the profiles provides further evidence for a magma chamber. Roughly three kilometers above the Moho there is an internal reflecting boundary I call Horizon *R*. Unlike the Moho, it is not smooth and continuous, indicating that the boundary in physical properties it represents is neither constant in depth nor laterally uniform. I believe Horizon *R* may represent a boundary within a magma chamber: the boundary between the layered gabbros, which form when minerals crystallize out of the magma and settle to the bottom of the chamber, and the more heterogeneous gabbros, which form by crystallizing against the side and top walls of the chamber. The physical contrast between these rock types is slight, but it may be enough to produce a reflector on wide-aperture images.



AIR GUNS towed by the Lamont-Doherty research vessel *Robert D. Conrad* are fired every 20 seconds or so. Reflections from the sea floor and underlying boundaries are detected with a 2.4-kilometer-long streamer deployed from a drum and towed below the surface.



tered under the segment. Toward the fracture zones the chamber becomes smaller and is perhaps only intermittently present. Hence the resulting crust is thinner; Horizon *R* and the layered gabbros disappear.

Fracture Zones

The structure of the deep crust, as revealed by Horizon *R* and the Moho, undergoes a striking change in the vicinity of an oceanic fracture zone. A fracture zone consists in part of an offset in a midocean ridge. The offset is a transform fault where the two diverging plates slide past each other. On each side of the ridge the crustal features formed at the transform fault may be recognizable for thousands of kilometers along the line of the offset. Our profiles of the North Atlantic cross several fracture zones at some distance from the ridge [*see illustration on pages 72 and 73*].

We have found that the crust becomes progressively thinner as a fracture zone is approached. It appears to thin from below: the layer between Horizon *R* and the Moho tapers as the Moho rises. Under the fracture zone itself Horizon *R* disappears. Under the Blake Spur Fracture Zone, for instance, the Moho rises abruptly until it is well above the normal depth of Horizon *R;* the thickness of the igneous part of the crust is reduced by from two to three kilometers, to just over half its normal value. The thickest crust is found midway between fracture zones.

The pattern has a simple explanation that is based on an idea initially put forward by Hans Schouten of Woods Hole. Each segment of the ridge may be fed by a separate magma chamber situated in a roughly central position between two fracture zones. Toward the fracture zones the magma chamber becomes progressively smaller. In a small chamber there is no room for a large quantity of mineral crystals to settle out of the melt, and so layered gabbros—which make up the crust between Horizon *R* and the Moho—are not formed. Under transform faults there may be no magma chamber at all; the normal process of crustal accretion may stop entirely. The hypothesis is supported by geochemical evidence: rocks erupted along different ridge segments are chemically distinct. Nevertheless, the hypothesis must be considered tentative.

Knowledge of the internal structure of fracture zones and midocean ridges is still meager, simply because few seismic images of their deep levels have been made. The same is true to a lesser extent of subduction zones, the third type of plate boundary. What I have presented here are only a few early insights from multichannel seismic reflection profiling; in terms of its application to basic geologic questions, the technique is in its infancy. Certainly the quality of the images will improve steadily, as investigators deploy more air guns, more sound detectors and better signal-processing methods. Equally important to an understanding of the earth's vast dynamic processes, however, is the quantity of images available. The small cohort of workers engaged in profiling the crust under the ocean floor have many crossings ahead of them.

The Heart as an Endocrine Gland

It is more than a pump. The atria secrete a recently discovered hormone, atrial natriuretic factor, that interacts with other hormones to fine-tune control of blood pressure and volume

by Marc Cantin and Jacques Genest

The heart is a pump: a muscular organ that contracts in rhythm, impelling the blood first to the lungs for oxygenation and then out into the vascular system to supply oxygen and nutrients to every cell in the body. That has been known since the publication in 1628 of William Harvey's *Essay on the Motion of the Heart and the Blood in Animals.*

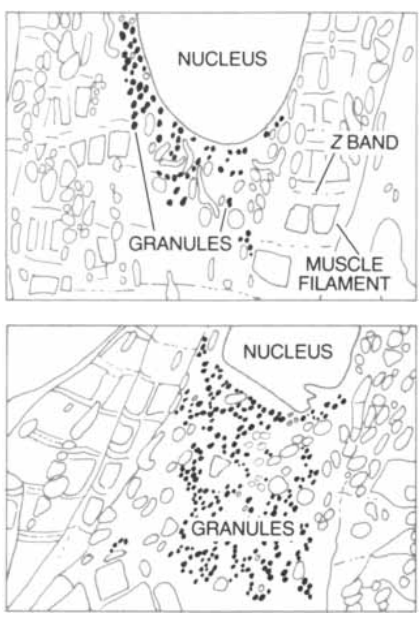
Within the past few years it has been discovered that the heart is something more than a pump. It is also an endocrine gland. It secretes a powerful peptide hormone called atrial natriuretic factor (ANF). The hormone has an important role in the regulation of blood pressure and blood volume and in the excretion of water, sodium and potassium. It exerts its effects widely: on the blood vessels themselves, on the kidneys and the adrenal glands and on a large number of regulatory regions in the brain.

The recent discovery of ANF solved a long-standing mystery. As early as 1935 the late John Peters of the Yale University School of Medicine speculated that there must be a mechanism in or near the heart to "sense the fullness of the bloodstream" and fine-tune the regulation of blood volume. During the 1950's and 1960's numerous investigators searched in vain for a hypothesized "natriuretic hormone." Such a hormone would explain the occurrence of natriuresis (excretion of sodium) and concomitant diuresis (excretion of water) in the absence of changes in known regulatory processes. Such unexplained natriuresis and diuresis were observed to follow distention of the atria, the two upper chambers of the heart, which receive blood from the pulmonary veins or the vena cava and deliver it to the adjoining ventricles. The putative hormone was referred to as the "third factor," since it would complement the activity of two known regulators of blood pressure and blood volume: the hormone aldosterone and the filtration of blood by the kidney.

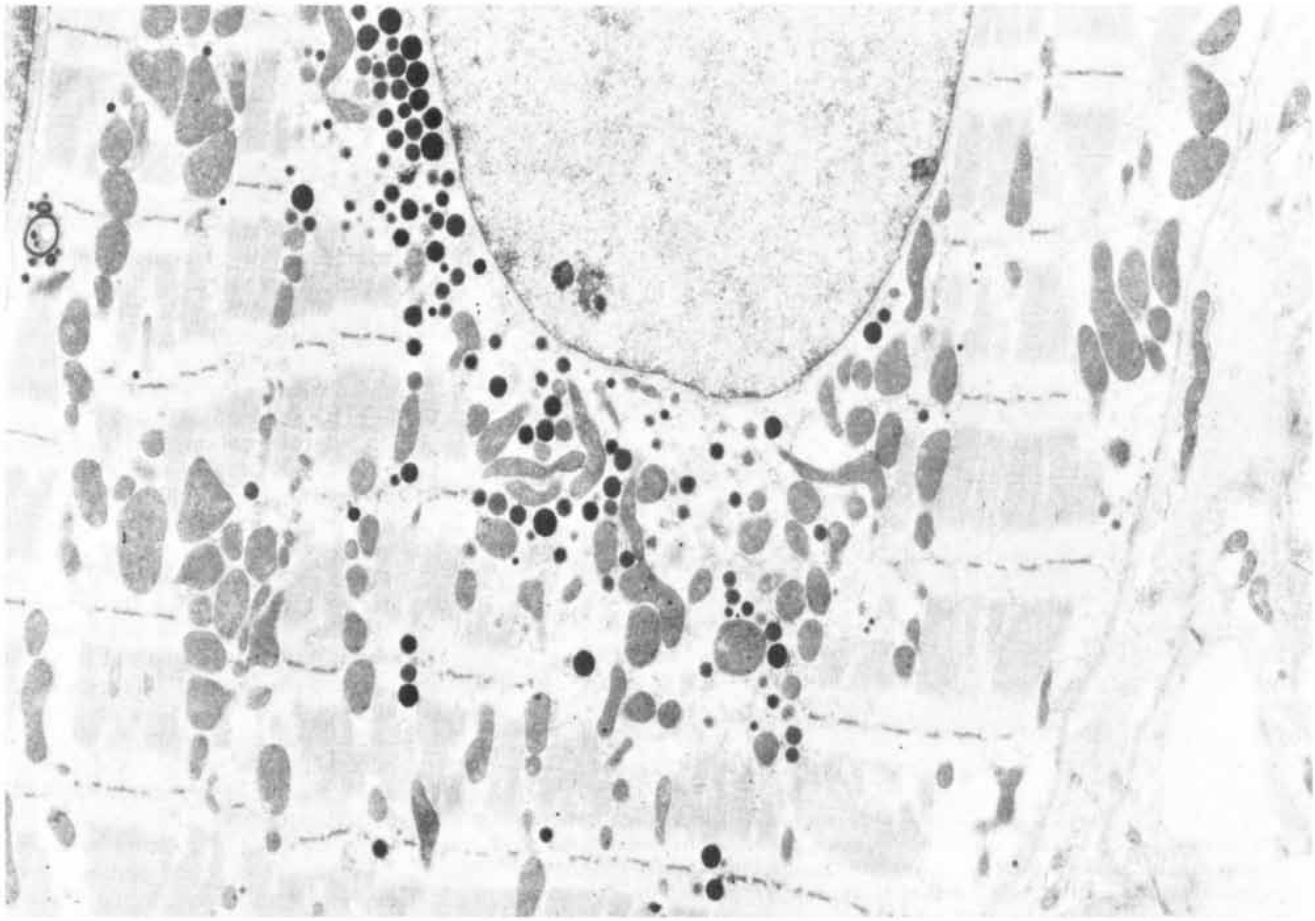
The first step toward the discovery of the third factor came in 1956, when Bruno Kisch of the American College of Cardiology noted the presence of what he called dense bodies in the cardiocytes, or heart-muscle cells, of guinea pig atria. In 1964 James D. Jamieson and George E. Palade of the Yale School of Medicine reported that such bodies, whose function was still not known, seemed to be present in the atria of all mammals they examined, including human beings. Our group at the University of Montreal noted in 1974 that the granules were very similar to storage granules seen in the endocrine (hormone-secreting) cells of, for example, the pancreas or the anterior pituitary gland. We found that when radioactively labeled amino acids were introduced into animals, they rapidly appeared in the atrial granules, incorporated into newly synthesized polypeptides (protein chains)—just as they would in the storage granules of endocrine cells.

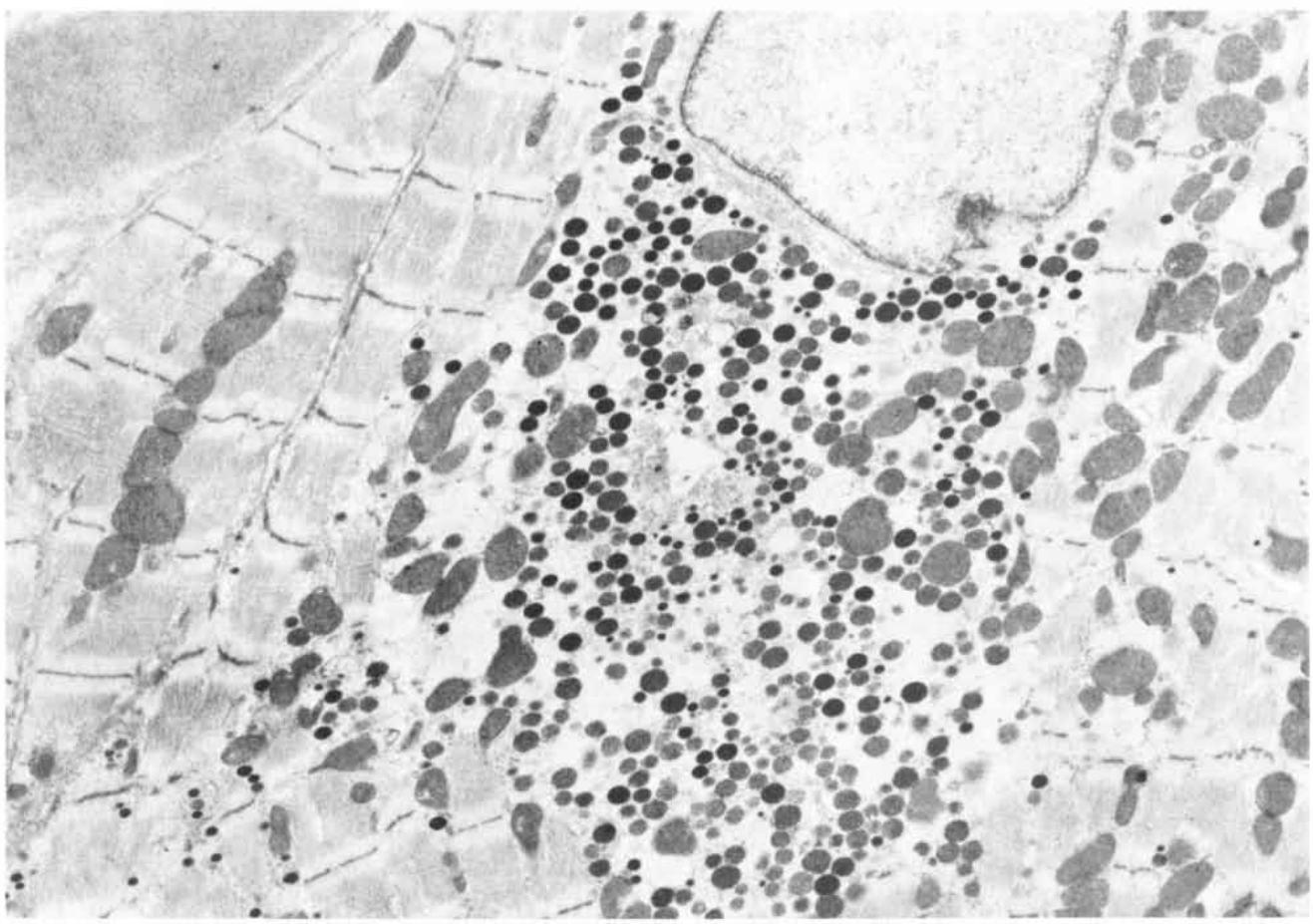
In 1976 Pierre-Yves Hatt and his colleagues at the University of Paris correlated current knowledge about the granules with earlier findings covering the regulation of sodium and water levels. They showed that the number of granules in the atrial cardiocytes increases when the amount of sodium in an animal diet is reduced. This implied that the granules must store some substance that has to do with sodium balance. A breakthrough was made in 1981, when Adolfo J. de Bold, Harald Sonnenberg and their colleagues at Queen's University at Kingston in Ontario injected homogenized rat atria into rats and observed a rapid, massive and short-lasting diuresis and natriuresis. They concluded that the atria indeed contained a "factor" that promotes these effects, and they named it atrial natriuretic factor.

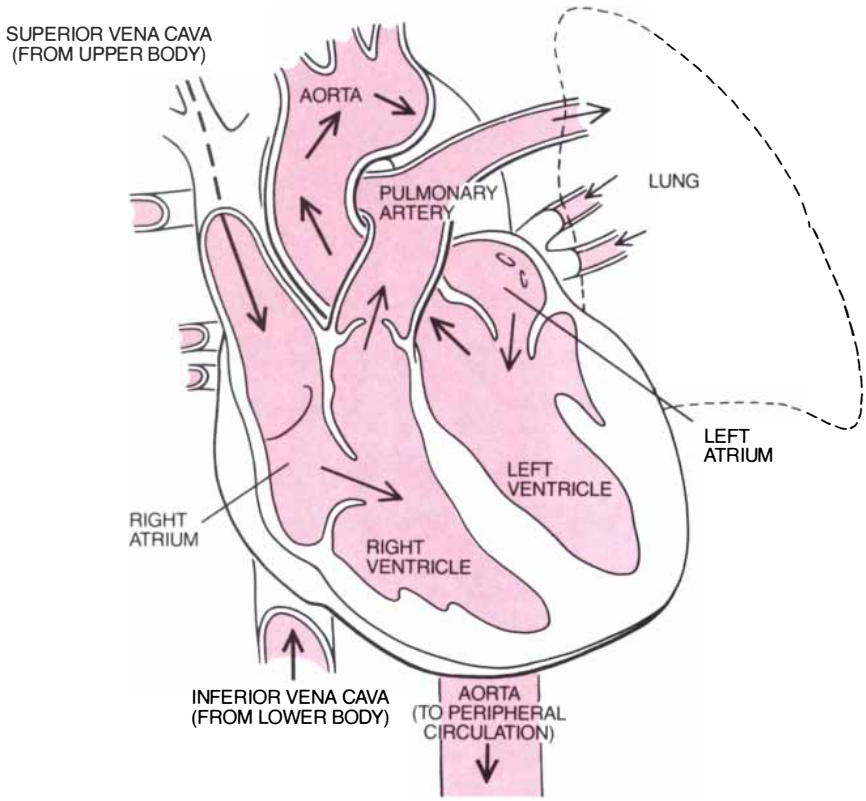
In the next three years the first direct evidence was reported for the location and biochemical identity of ANF. Workers found there are from two to two and a half times as many of the granules in the right atrium as there are in the left atrium in rats. The granules are highly concentrated near the surface of the heart and in the exteri-



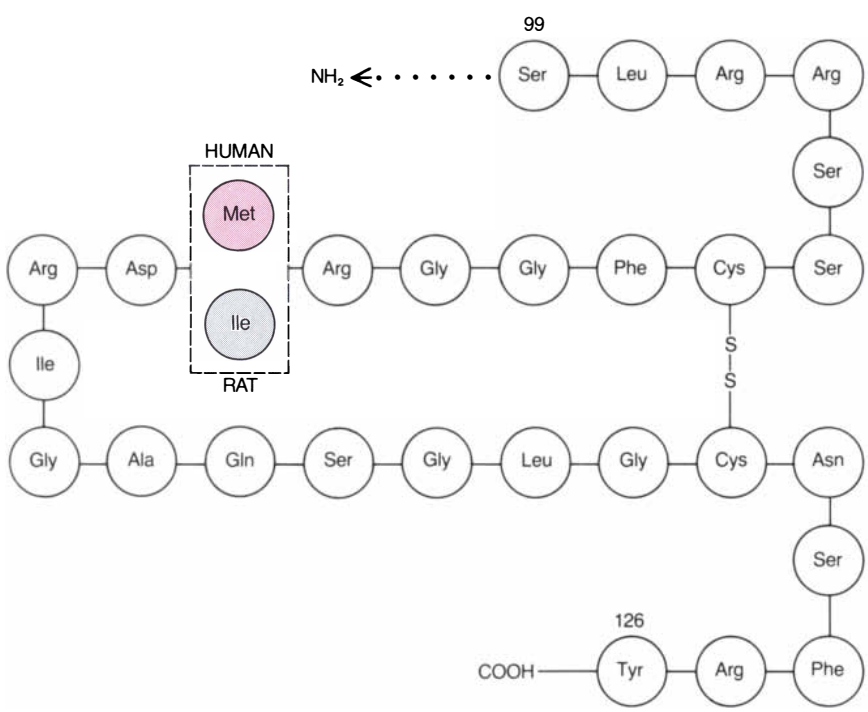
HORMONE-STORAGE GRANULES in cardiocytes, or heart-muscle cells, of rats are enlarged some 12,000 diameters in electron micrographs made in the authors' laboratory. Discovery of such granules first suggested that the heart is an endocrine organ. Stretching of the cardiocyte's contractile apparatus (*filaments with* Z *bands*) stimulates the release of the hormone called atrial natriuretic factor (ANF). In a cell of a normal rat (*top*) granules are seen clustered near the nucleus. In a cell from a rat fed a sodium-deficient diet for 30 days, the number of granules is increased (*bottom*), possibly because the blood volume is lowered. Reduced blood volume decreases the circulating level of ANF, leading eventually to an accumulation of granules in the cell.







ANF IS SECRETED in mammals by cardiocytes in the right and left atria of the heart. Oxygen-depleted blood from the periphery enters the right atrium from the great veins, empties into the right ventricle and is pumped through the pulmonary artery to the lungs. Oxygen-replenished blood returns to the left atrium and is pumped from the left ventricle to the aorta, to be distributed to the periphery. In some nonmammals the ventricles as well as the atria seem to exhibit diuretic and natriuretic activity, and so they too may secrete ANF.



AMINO ACID SEQUENCE of the circulating ANF molecule appears to be identical in the human and the rat except at position 110, where the human ANF has methionine and the rat ANF has isoleucine. Both the human and the rat active ANF molecules consist of 28 amino acids, and they both include a disulfide bond between two cysteines that is essential to ANF's activity. The active circulating hormone is cleaved from a much larger precursor polypeptide molecule that is 152 amino acids long in the rat, 151 amino acids in the human.

or regions of the atria. They have not been found in the ventricles of rats or any other mammals, and the injection of mammalian ventricular extracts does not affect blood vessels, diuresis or natriuresis. In contrast, granules have been discovered in the ventricles as well as in the atria of nonmammalian species and can be shown to be related to diuretic and natriuretic effects. The presence of the granules in the ventricles of nonmammalian species but not in the ventricles of mammals is consistent with the fact that heart cells tend to be more specialized in higher species.

Once the location of ANF was determined, the peptide was isolated and purified in June, 1983, by our group; it was synthesized two months later by Ruth F. Nutt of the Merck Sharp & Dohme Research Laboratories and her colleagues. ANF is the active part of a larger precursor molecule. When various groups determined the amino acid sequence of the polypeptide, they all found ANF has the same core of 21 amino acids. The active, circulating hormone in the rat has 28 amino acids and a molecular weight of 3,060. The active hormone is attached to an inactive peptide of 100 amino acids and a 24-amino-acid signal peptide that is cleaved when the molecule is synthesized. The circulating form in the human has not yet been determined, although we strongly suspect that it too is made up of 28 amino acids.

Recently the human ANF gene has been cloned and sequenced, making it possible to synthesize the hormone chemically or by inserting the gene into yeast or bacteria. By either method the hormone can be produced in quantity for studies of its activity throughout the body. In addition antibodies to the hormone have been developed, so that sensitive immunological tests can be carried out to trace the release of ANF and to find the sites where it is active.

When rats are subjected to the stress of immobilization, there is a five- to twentyfold increase in the blood level of ANF. The release of ANF from the heart was also measured in human patients with valvular disease and expansion of blood volume who were undergoing cardiac catheterization, a procedure that allows sampling of blood in the arteries and heart chambers. The plasma level of ANF was from two to eight times as high in venous blood from the coronary sinus, which drains the heart's atria, as it was in blood circulating in the arteries or peripheral veins. This confirms that ANF is released primarily, if not exclusively, by cells in the atria. In both the rats subjected to stress and the catheter-

ized patients the atrial cardiocytes are stretched, and the stretching is the signal for the release of ANF. An increase in blood volume can also cause the atrial cardiocytes to stretch and release ANF, as was shown when the blood volume of rats was experimentally increased by the infusion of a salt solution.

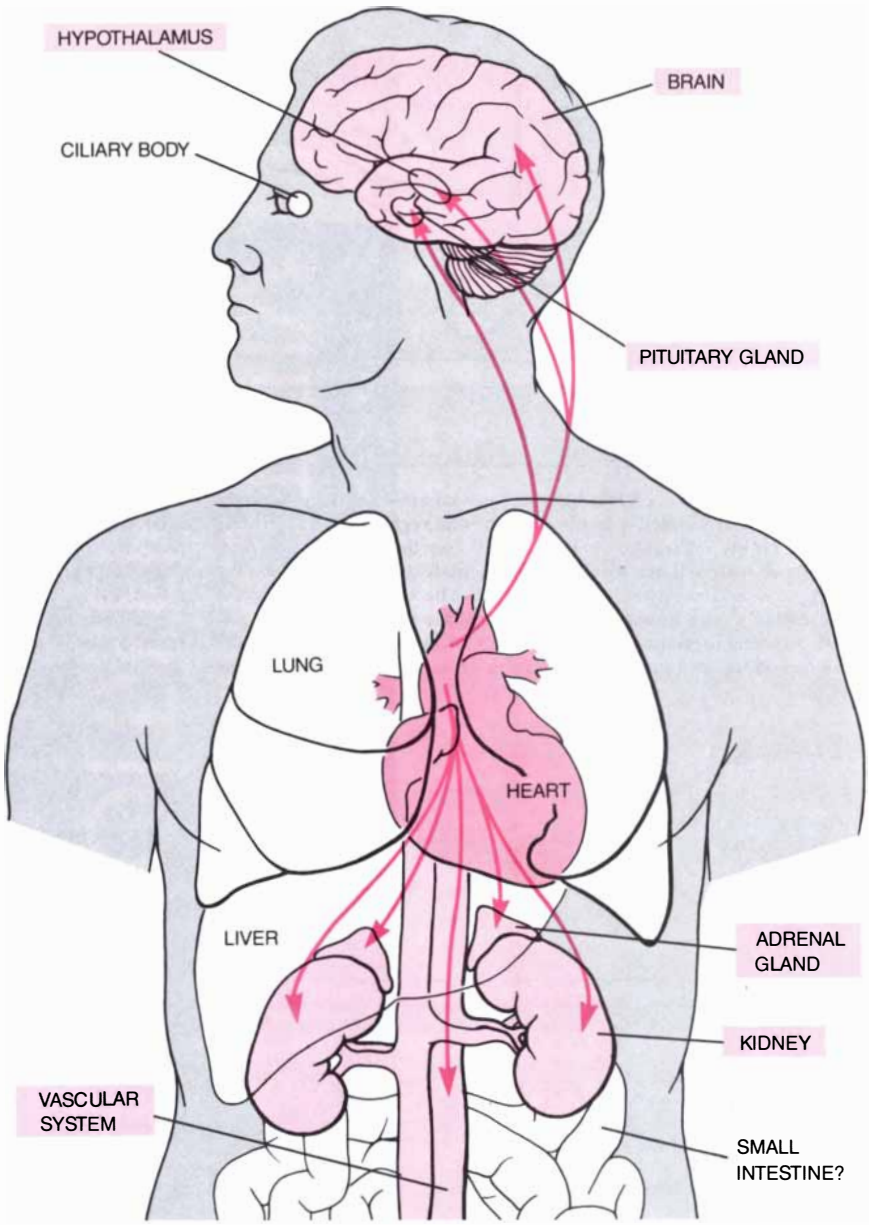
Once the cardiocytes respond to stretching by releasing ANF, the peptide travels through the arteries to targets in the kidneys, the adrenal glands, the brain and various other tissues. In general what ANF does is to modify the activity of a complicated homeostatic feedback loop regulating blood pressure, blood volume and sodium retention: the renin-angiotensin system, which links certain functions of the brain, heart, arteries, adrenals, kidneys and other organs. One of the key substances in the system is the enzyme renin. It is secreted by cells in the arteries that lead to the glomeruli, which are saclike structures at the entrance to the kidney. These juxtaglomerular cells secrete renin into the bloodstream whenever the level of sodium is low in the distal tubules of the kidney or when the local pressure in the kidney is low.

Circulating renin cleaves a polypeptide called angiotensinogen to produce angiotensin I, which in turn is converted into angiotensin II. This small peptide is a powerful constrictor of vascular smooth muscle. Angiotensin II also has a feedback effect, partially suppressing renin secretion from the juxtaglomerular cells. Finally, it stimulates the adrenal gland to secrete the hormone aldosterone, which travels to the kidney and the posterior pituitary to inhibit the excretion of sodium and water.

ANF affects the renin-angiotensin system by somehow inhibiting the secretion of renin and also by directly inhibiting the adrenal secretion of aldosterone. The relation between ANF and aldosterone was elucidated in our laboratory by a series of experiments with cultured bovine and rat adrenal cells. ANF inhibited the normal production of adrenal aldosterone by 20 percent; it reduced by from 40 to 70 percent the stepped-up production that ordinarily follows stimulation of the adrenal cells with angiotensin II or the pituitary hormone ACTH. Similar significant decreases in plasma aldosterone levels have been noted in rats and dogs after the injection of ANF.

In order to determine whether the decreases in aldosterone levels were caused by the presence of specific binding sites for ANF on the surface of adrenal cells, we introduced radioactively labeled ANF into adrenal-cell cultures and then added "cold," or unlabeled, ANF. We observed that the concentration of the labeled ANF decreased significantly after introduction of the cold ANF, a sign that the cold ANF was displacing the labeled peptide at many sites on the surface of the cells. We determined that ANF binds at very specific sites on the cell by introducing ACTH, angiotensin II and other active peptides. The fact that these peptides did not exert the same displacement effect as cold ANF testified to the presence of binding sites that are specific for ANF.

In addition to ANF's effects on the renin-angiotensin system, the peptide acts directly at various sites in the kidney to regulate water and sodium excretion. Radioactively labeled sam-

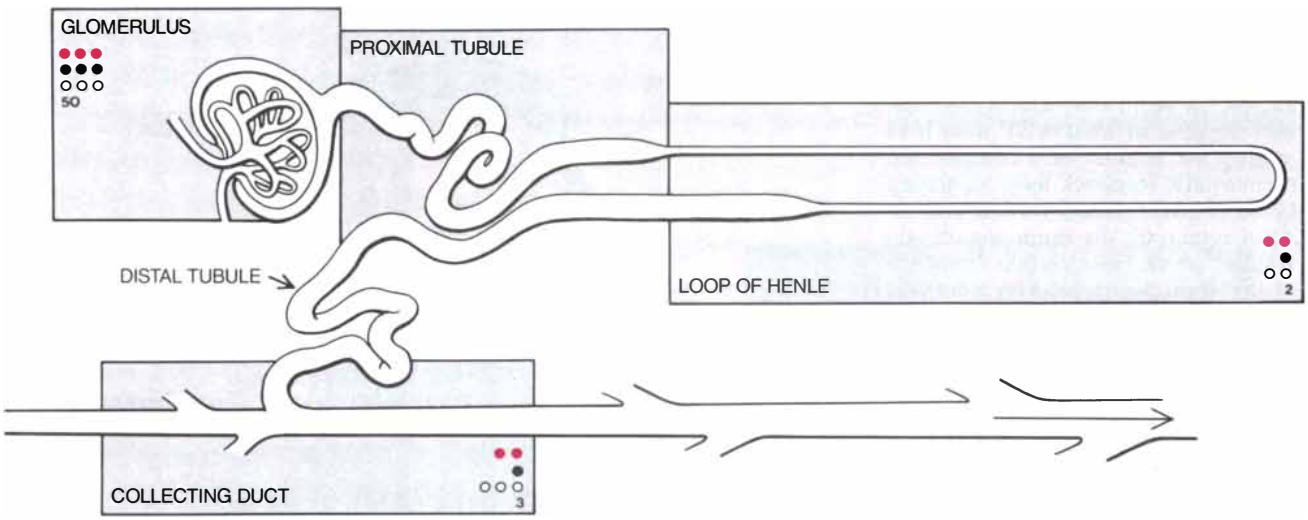


ANF IS KNOWN TO AFFECT various regions of the brain (including the hypothalamus), the posterior pituitary gland, the adrenal gland, the kidney and the vascular system. ANF also binds to targets in the lung, the liver, the ciliary body (which secretes the lymphlike aqueous humor of the eye) and probably the small intestine, but its effects in these tissues are not clear. In the brain ANF binds to various sites that are involved in blood-pressure control and the regulation of sodium and water; in the hypothalamus ANF inhibits the release of vasopressin, a hormone stored in the posterior pituitary that is antidiuretic and can cause arterioles and capillaries to constrict. ANF relaxes the smooth-muscle cells of blood vessels, inhibits secretion of aldosterone (a hormone that tends to raise blood pressure) from the adrenal gland and stimulates the kidney to increase excretion of sodium and water from the blood-filtering glomeruli. In addition ANF appears to decrease reabsorption of sodium into the blood from the distal tubules and collecting ducts of the kidney.

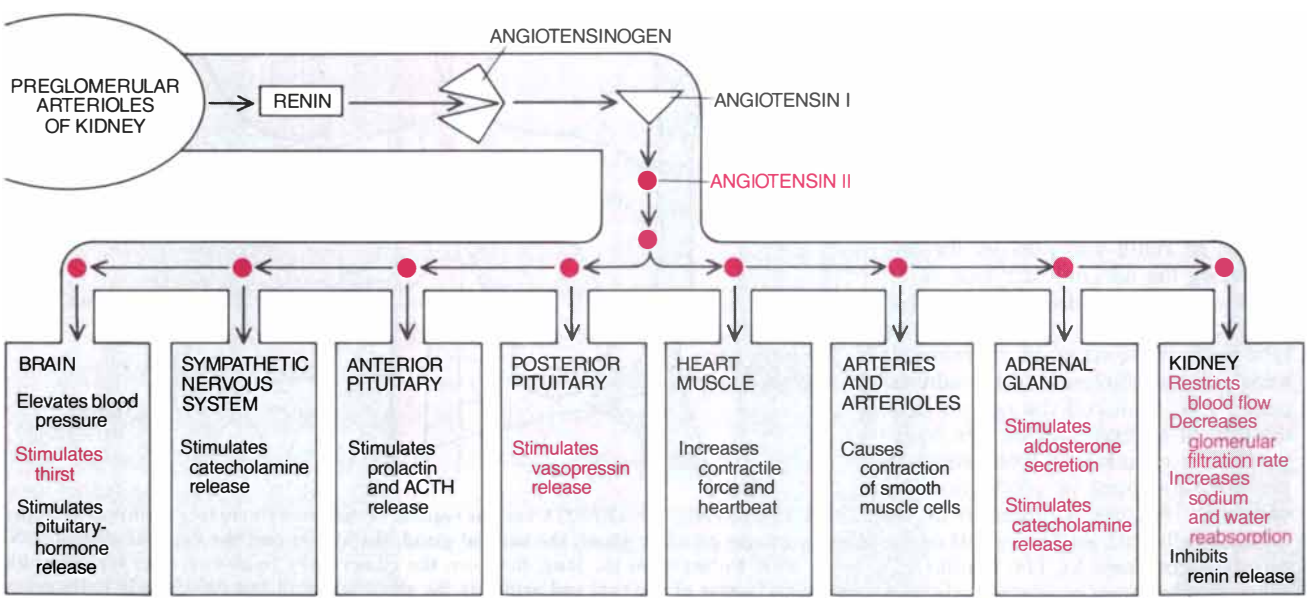
ples injected into the rat aorta reveal that ANF binds to epithelial cells in the glomeruli and to numerous receptor sites throughout the blood vessels in the vicinity of the glomeruli and tubules. The hormone somehow exhibits a short-term effect on the mechanism by which the glomeruli filter the blood. ANF probably causes the lining of the glomeruli to become more permeable, allowing larger quantities of water and sodium to be filtered from the blood.

ANF also acts directly in the kidney's tubules, where urine is formed from the filtered blood plasma transported from the glomeruli. The distal tubules "reabsorb" sodium from the filtrate and pass it back into the bloodstream, and ANF probably decreases this activity. The mechanism for ANF's effect in the tubules remains a mystery, since there is no evidence that it requires energy: ANF activity does not require the consumption of oxygen or the breakdown of glucose, unlike other mechanisms of reabsorption.

The role of ANF as a relaxer of



IN THE KIDNEY, ANF is apparently most active in the glomeruli and somewhat less active in other functional regions, as is indicated by this highly schematic diagram. ANF functions by binding to a target cell, where it often activates particulate guanylate cyclase, an enzyme localized on the cell membrane. The enzyme activates the nucleotide cyclic guanosine monophosphate (cyclic GMP), a so-called second messenger that carries ANF's message to a site within the cell. ANF also inhibits stimulation of adenylate cyclase, which is involved in a different second-messenger system. Studies with radioactively labeled ANF showed a very high density of binding (*colored circles*) in rat glomeruli. On the same 0-to-3 scale other tests detected marked guanylate cyclase activity (*black circles*) and adenylate cyclase inhibition (*open circles*) in the glomeruli; the cyclic-GMP level (*numbers*) increased more than 50 times over the basal level. The collecting ducts and the thick part of the loops of Henle showed some ANF activity, but the proximal tubules showed none.



ANF INHIBITS certain actions of angiotensin II, a peptide hormone that raises blood pressure and increases blood volume, among other effects. When the renal blood pressure is low, the kidney secretes the enzyme renin into the bloodstream. Renin cleaves a polypeptide precursor called angiotensinogen to produce angiotensin I. This is converted in turn into the peptide angiotensin II, which the bloodstream delivers to the organs shown here. Angiotensin II has its hypertensive effects by directly making blood vessels contract, by stimulating the secretion of other contractile hormones (such as vasopressin or adrenaline and other catecholamines) and by stimulating the adrenal gland to release aldosterone, which acts in the kidney to promote the retention of salt and water. Angiotensin II also has a feedback effect, inhibiting the release of renin from the kidney. Actions known to be inhibited by ANF are shown in color.

muscle cells throughout the vascular system is as important as its functions in the glomeruli and tubules of the kidney. Although injections of ANF can be shown to relax and expand the large vertebral, femoral, common carotid and coronary arteries, the peptide's effect is profoundest in the small arteries of the kidney. When synthetic ANF was added to baths containing strips of rat and rabbit renal (kidney) arteries, the usual constricting activity of angiotensin II and the hormone norepinephrine on the vascular tissue was profoundly inhibited for from 30 to 80 minutes. The infusion of ANF into an isolated rat kidney resulted in a rapid drop in perfusion pressure that lasted for 18 minutes. These findings meant the ANF must be acting either in the smooth-muscle cells of the blood vessels or in the endothelium, the cellular lining of the vessels. When workers added the peptide to blood vessels after destroying the endothelium, the vasorelaxation effect persisted, indicating that ANF acts on the smooth-muscle cells.

The precise effect of ANF on the smooth-muscle cells is not known. We are doing experiments to test our hypothesis that it affects either the entry of calcium into the cells or its relocation within the cells. ANF could do this indirectly by activating cyclic guanosine monophosphate (cyclic GMP). This is a nucleotide that acts as a second messenger: a substance that transmits into the interior of the cell the message that has been delivered by a hormone to the cell's surface. Evidence for a relation between ANF and cyclic GMP was demonstrated by the finding that in rats the injection of the peptide leads to significant increases in cyclic GMP in the plasma and urine. ANF may have this effect by activating the enzyme guanylate cyclase, which is attached to the cell membrane and has a role in the activation of cyclic GMP. ANF also inhibits the stimulation of adenylate cyclase, an enzyme that is implicated in second-messenger activity in some cells.

Experiments with radioactively labeled synthetic ANF have shown that the peptide acts at multiple sites in the brain of rats and guinea pigs, binding to areas implicated in the regulation of blood pressure and the control of sodium, potassium and water levels. Investigators have also learned that ANF inhibits the production of vasopressin, a hormone that is synthesized in the hypothalamus at the base of the brain and moves to the posterior pituitary for storage. When vasopressin is released from the pituitary, it constricts blood vessels, raises blood pressure and influences the reabsorption of water by kidney tubules. Finally, experiments have revealed that ANF binds at various sites in the ciliary body of the eye, possibly contributing to the control of ocular pressure there.

While research on the physiology of ANF's effects continues, several groups are investigating how ANF might serve as a drug to control hypertension and congestive heart failure. In our laboratory we have observed a significant short-term (less than one hour) reduction in blood pressure when hypertensive rats are given a single injection of one microgram of synthetic ANF. We found the greatest reductions in animals whose hypertension was dependent on the activity of renin. When ANF was infused at the rate of one microgram per hour for seven days, blood pressure dropped significantly: to normal levels after the second day. A similar treatment with smaller doses administered for 12 days led to a drop in blood pressure over the final 10 days and to lowering of plasma and urinary levels of aldosterone.

In subsequent studies we measured ANF levels in a line of rats genetically predisposed to high blood pressure. We found a high level of ANF in the bloodstream and a lower than normal level in the left (but not in the right) atrium. It would appear that the high circulating level is a sign of the body's attempt to lower the blood pressure; the lower level found in the left atrium suggests that ANF there becomes depleted in hypertension.

ANF appears also to have a major role in congestive heart failure. This is a condition in which the heart fails to pump the blood properly, with the result that the patient becomes short of breath and develops marked edema of the legs. Although it is not known how ANF contributes to the disease, our studies of a line of hamsters afflicted with spontaneous congestive heart failure have correlated changes in ANF levels with the progress of the disease.

In these hamsters the arterial blood pressure is always abnormally low; the venous pressure increases with the severity of the disease. At all stages the amount of ANF in the atria is lower than it is in control animals. The most striking finding is a significant increase in circulating ANF, which is noted as soon as the venous pressure within the heart begins to increase. The level of circulating ANF reaches a peak when the disease is moderately advanced and then decreases in the final stages. Postmortem studies of the diseased atria reveal what we call exhaustion hyperplasia: an increase in the rough endoplasmic reticulum, the site of peptide synthesis; an increase in the size of the Golgi apparatus, where peptides are processed, and a decrease in the number of secretory granules, each of which also harbors less ANF.

These results suggest that a small increase in atrial pressure is enough to trigger hypersecretion of ANF. The decreased ANF level within the atria suggests depletion. We believe sodium retention and activation of the renin-angiotensin-aldosterone system, two indications of congestive heart failure, are postponed until the later stages of the disease by the release of large quantities of ANF. The eventual appearance of these effects may be the result of "down regulation": in the presence of excessive amounts of ANF the target cells may decrease the number of ANF-binding sites on their outer membrane, slowing the cellular reactions that are ordinarily triggered by the binding of ANF.

Further study of ANF should lead to new treatments for hypertension and other blood-pressure diseases, for blood-volume disorders and for kidney diseases affecting the excretion of salts and water. In spite of progress in studying diseases associated with ANF, however, much work lies ahead before synthetic forms of the peptide can be administered to treat patients. The physiology of ANF's effects on the renal tubules remains to be elucidated, as does the relation between the relaxation of blood vessels, calcium movements in the smooth-muscle cells and the effects on adenylate and guanylate cyclases. Workers must also investigate the factors that activate the release of ANF from cardiocytes and determine in detail how the peptide acts in various regions of the brain. Fortunately recent clinical trials have confirmed that all the effects of ANF demonstrated in animals also occur in man; these findings should shorten the route to the development of treatments for disease.

In addition to a more detailed understanding of ANF's physiological activity, progress toward ANF therapy will depend on the development of techniques for tailoring particular analogues of ANF as drugs with which to treat each disorder at a specific binding site. Investigators will have to develop ways of modifying the drugs so that they will be protected from stomach enzymes and acids and can be absorbed easily when they are administered by mouth. Biotechnology and chemical synthesis will provide the needed techniques for these tasks, but several years will probably pass before the first ANF analogues are ready for controlled testing in human patients.

Reconstructing Bird Phylogeny by Comparing DNA's

Differences between DNA's reveal evolutionary distances between species, making it possible to reconstruct and date the branchings of avian lineages and providing a basis for classifying living groups

by Charles G. Sibley and Jon E. Ahlquist

All organisms have ancestors; therefore all organisms have an evolutionary history. Because all the plants and animals presumably evolved from a single origin, they share a single phylogeny, or history. The reconstruction of this phylogeny is a primary goal of evolutionary biology. Living species are the topmost twigs of a vast phylogenetic tree whose larger branches and trunk are no longer directly visible. To reconstruct the tree of life it is necessary to determine the branching pattern and, if possible, date the branching events of the past.

Over the past 10 years we have used a technique that extracts evidence of phylogeny from the genetic material, DNA. The method, DNA-DNA hybridization, has enabled us to reconstruct the branching pattern of the major lineages of birds. The approximately 9,000 living species of birds are the descendants of lineages that began to diverge from one another about 150 million years ago in the late Jurassic and early Cretaceous periods, after the origin of birds from a reptilian ancestor. Pierce Brodkorb of the University of Florida has estimated that about 150,000 species of birds have existed. The living species are only 6 percent of the total; the rest are extinct.

Our approach has been to measure the average difference between the DNA's of species representing the major groups of living birds and to use the results to reconstruct the branching sequence of the avian tree. The reconstructed phylogeny provides the basis for a classification of birds in which living species are assigned to taxonomic categories on the basis of their genealogical relationships. In some cases our results have indicated changes from traditional avian classifications.

The elements of a phylogeny are the branching pattern and the date of each branching event. A branching occurs when a barrier, usually a geographic one, divides a single species into two populations, which then diverge genetically and become the ancestors of two lineages. Each lineage in turn may split, and the process may repeat itself to produce a radiation of morphologically and ecologically varied species.

Until recently comparisons of the anatomical characters of living species were the only source of information about the pattern of branching. Such comparisons have answered many questions and have established the outlines of the history of life. Anatomical characters, however, are shaped by functional requirements; thus structure may provide false clues about phylogeny because the process of convergent evolution can produce similar structures in unrelated organisms. Swifts and swallows, for example, are superficially alike because both groups are specialized to feed on flying insects. In early classifications of birds the two groups were placed together. Later studies showed fundamental differences in anatomy, and it was eventually realized that the groups are not closely associated: swifts are distant relatives of hummingbirds, and swallows are related to other songbirds. Many cases of convergence are so subtle that they defy solution by anatomical comparisons.

To determine the dating of the branching events that were inferred from anatomical comparisons it has been necessary to rely on the fossil record. Partial phylogenies for certain groups have been reconstructed from fossil data, but the fossil record for some groups, including birds, is fragmentary. Moreover, although a dated fossil indicates approximately when the individual organism perished, the time its lineage diverged from that of its relatives usually remains uncertain.

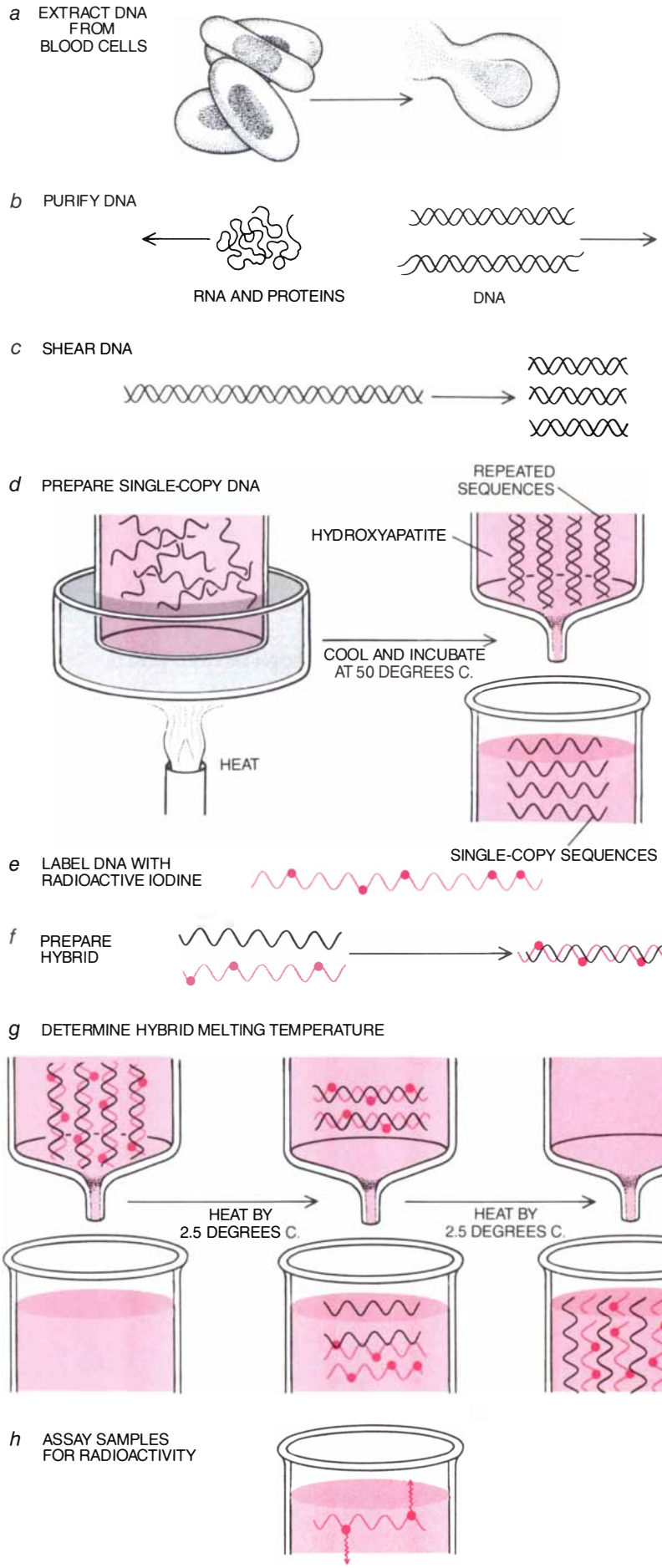
Clearly a direct method for measuring genealogical distances among extant lineages and dating the divergences between lineages should improve the reconstruction of phylogenies. The genetic relationships among living species reflect their evolutionary history; because genetic change is mainly divergent, the genetic difference between any two lineages is related to the length of time since the lineages last shared a common ancestor. To reconstruct the phylogeny of birds we therefore studied their genetic material.

In all organisms except for certain viruses DNA is the genetic material. It is a double-stranded molecule, in which each strand is a sequence of four kinds of chemical units called nucleotides. Each nucleotide is composed of a five-carbon sugar, a phosphate group and a base. The nucleotides differ from one another only in their bases: adenine (*A*), thymine (*T*), cytosine (*C*) and guanine (*G*). The bases along the two strands form complementary pairs, held together by hydrogen bonds: an *A* pairs with a *T* and a *C* with a *G*. Genetic information is encoded in the sequence of bases. Specific sequences of bases form genes, which code for the many kinds of proteins that make up most of the structures of plants and animals and control their functions.

Within the genome, or complete set of genes, of a cell most genes are present as a single copy. Between about 3 and 5 percent of the different sequences in higher organisms occur as more than one copy in each genome. These repeated sequences may make up about 40 percent of the total volume of DNA in a cell.

There are about two billion pairs of nucleotides in the genome of a bird. The technique of DNA-DNA hybrid-

BARBET AND TOUCAN SPECIES perch on a branch symbolizing the genealogical relationships of the groups to which they belong. The barbets of the New World and those of Africa and Asia have traditionally been placed in one family and the toucans in another. Comparisons of DNA's showed that in spite of the birds' appearance the New World barbets are related more closely to the toucans than to the Old World barbets. The degrees of difference between DNA's indicated the African barbets last shared an ancestor with the New World barbets and the toucans about 55 million years ago; the latter groups diverged about 30 million years ago.



ization enables us to compare these huge numbers of genetic units and measure the genetic differences between living species. From these measurements we are able to reconstruct the branching pattern of the phylogeny and, by calibrating the measured genetic differences against time, calculate the approximate dates of the divergences between living lineages.

The technique of hybridization depends on the properties of DNA molecules. When a solution containing double-stranded DNA is heated to boiling, the hydrogen bonds between complementary bases "melt," or rupture, and the DNA separates into single strands. The hydrogen bonds are the weakest bonds in DNA, and the rest of the molecule is not damaged by boiling. As the melted sample cools, the single strands collide by chance. If colliding strands have complementary sequences of bases, they will reassociate into the double-stranded structure as complementary bases "recognize" each other and the hydrogen bonds between them are reestablished. If the reassociation takes place at a low temperature, the restored duplex, or double-stranded DNA, can contain numerous mismatches, but at a temperature of 60 degrees C. about 80 percent of the bases must be properly matched for the duplex to be stable. Under such conditions single strands of DNA will

DNA-DNA HYBRIDIZATION requires a cellular source of DNA; for studies of birds the DNA is extracted from red blood cells (which in birds have nuclei) by rupturing them (*a*). The DNA is separated from the RNA and proteins (*b*), and the long strands are sheared into shorter fragments (*c*). The DNA is then boiled briefly, causing the double-stranded molecules to separate into single strands. As the DNA cools, sequences that are present as numerous copies reassociate more rapidly than those present as single copies. The reassociated DNA is passed through a column of hydroxyapatite, which binds double-stranded but not single-stranded DNA (*d*). Most of the repeated sequences, which have two strands, are bound to the hydroxyapatite; the fragments of DNA containing single copies of each gene pass through. The single-copy DNA is labeled with radioactive iodine (*e*) and mixed with unlabeled DNA from the same species or a different one. The combined DNA's are incubated at 60 degrees Celsius for 120 hours, forming a DNA hybrid containing one labeled and one unlabeled strand (*f*). Each DNA hybrid is placed in a column of hydroxyapatite, which is heated in a water bath from 55 degrees C. to 95 degrees in increments of 2.5 degrees. At each increment the single strands released by the melting of the hybrids are washed from the column into a vial (*g*). Measurement of radioactivity in the vials indicates how much of the hybrid melted at each temperature (*h*).

reassociate only with their complementary partners, and the original double-stranded DNA will be restored.

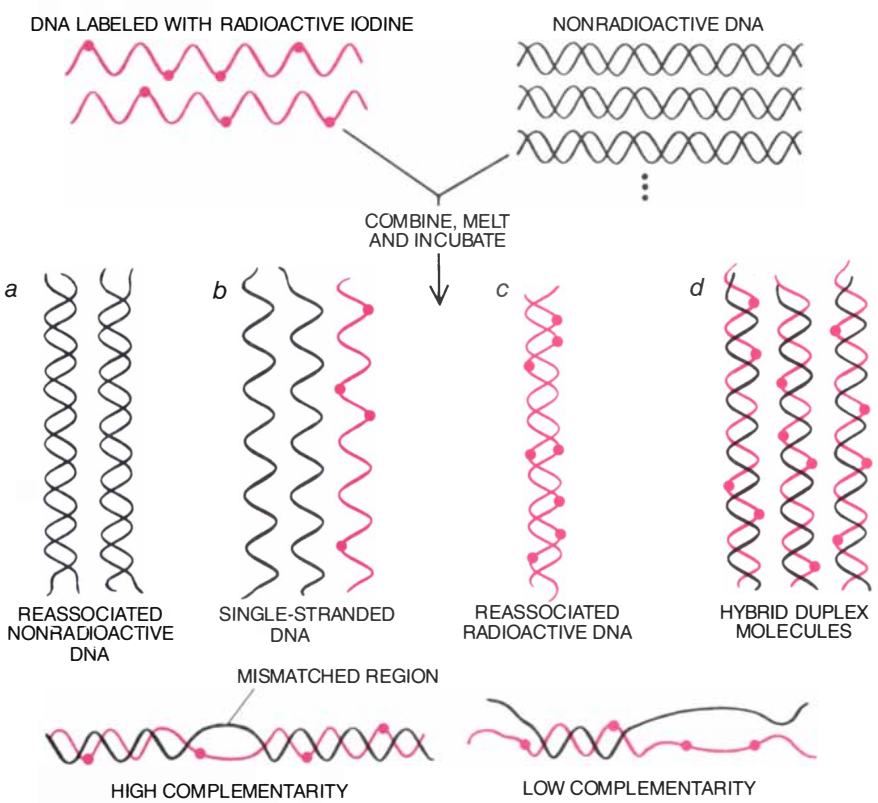
When the single-stranded DNA's from two different species are combined and incubated at 60 degrees C., hybrid double-stranded DNA will form only between homologous base sequences: sequences inherited from a common ancestor of the two species. Only homologous sequences contain enough complementary pairs to form thermally stable duplexes at 60 degrees. A hybrid duplex of DNA from different species will contain mismatched bases because the two lineages have incorporated different sets of mutations since they last shared an ancestor. Thus an *A* may be opposite a *C* or a *G* may be opposite a *T*, and no bonds will form between the bases. Since the melting temperature of the duplex is proportional to the number of hydrogen bonds between the two strands, such mismatches will cause the hybrid DNA to melt at a temperature lower than that required to melt perfectly base-paired double strands.

In the DNA-DNA hybridization procedure DNA is extracted from the nuclei of cells and separated from the proteins and other cell constituents. The long strands are sheared into fragments averaging 500 nucleotides in length. Most of the repeated gene sequences are removed from the DNA of the species that is to be compared with other species, and the "single copy" DNA that remains is labeled with radioactive iodine. A small quantity of the radioactive DNA (known as the tracer) is combined with a much larger amount of unlabeled DNA (the driver) from the same species. The same tracer is also combined with the driver DNA's of other species. Each mixture will yield a different hybrid: either a homoduplex, in which the tracer and driver species are the same, or a heteroduplex, in which the two strands represent different species. The homoduplex form provides a standard against which the melting properties of the heteroduplexes can be compared.

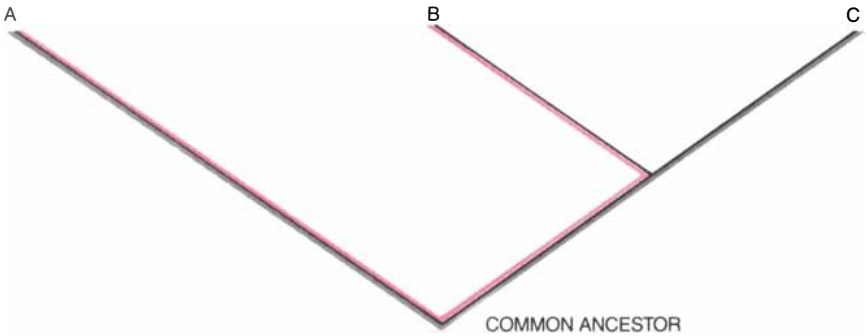
The tracer-driver mixtures are then boiled for five minutes to dissociate the double-stranded molecules into single strands. To allow the single strands to reassociate into duplexes they are incubated for 120 hours at 60 degrees C. in a sodium phosphate buffer solution. Each of the different double-stranded hybrids that results is placed on a column of hydroxyapatite, a form of calcium phosphate that binds double-stranded DNA but not the single-stranded form. The columns are immersed in a water bath at 55 degrees C., and the temperature of the bath is raised in increments of 2.5 degrees to 95 degrees. At each of the 17 temperatures the single-stranded DNA fragments produced by the melting of the duplexes are washed from each column into a vial. The radioactivity in each vial is counted, indicating how much of the corresponding hybrid had melted. The results are plotted as a melting curve: a graph showing how much of the hybrid had melted at each temperature. The median difference in degrees Celsius between the homoduplex curve and each heteroduplex curve is a measure of the median genetic difference between the tracer species and each driver species it was compared with.

The difference between the DNA's of two species can serve as an indicator of the genealogical distance between them only if DNA can be assumed to change at an average rate that is the same in all lineages. Emile Zuckerkandl and Linus Pauling, then at the California Institute of Technology, proposed in 1962 that proteins evolve at constant rates, and "molecular clocks" have been discussed and debated ever since.

We have found that the DNA clock seems to tick at the same average rate in all lineages of birds. The evidence comes from a procedure known as the relative-rate test, which was suggested in 1967 by Vincent M. Sarich and Allan C. Wilson of the University of California at Berkeley. A relative-rate test compares any three species of which two are known to be more closely related to each other than either one is to the third. If we choose such a trio of bird species and compare tracer DNA from the outlying species with driver DNA from each of the other two, we find that the genetic distances between the outlier and each of the other species (indicated by the melting temperatures of the DNA hybrids) are always equal, within the limits of experimental error. Since the same length of evolutionary history separates both of the driver species from the last ances-



HYBRIDS ARE FORMED from a small amount of radioactively labeled DNA and 1,000 times as much unlabeled DNA from the same species or a different one (*top*). The DNA is melted to separate it into single strands, which are incubated to allow the single strands to reassociate. Much of the unlabeled DNA reassociates with complementary unlabeled strands (*a*); because these products are not radioactive, they do not affect later measurements. Some of the single strands fail to reassociate (*b*), and only about 1 percent of the radioactively labeled strands hybridize with other labeled strands (*c*). Some of the DNA forms hybrid duplex molecules, consisting of one labeled and one unlabeled strand (*d*). In hybrids of DNA from two species, the proportion of the nucleotide bases along a strand that are paired with complementary partners on the adjacent strand depends on the genetic similarity of the species. Because a greater number of bonds link the strands in well-matched hybrids (*bottom left*), they melt at a higher temperature than poorly matched ones (*bottom right*).



RELATIVE-RATE TEST confirms that DNA evolves at the same average rate in different species. The test can be done with any trio of species in which two of the species (*B, C*) are more closely related to each other than either one is to a third (*A*). If DNA from each of the two allied species is hybridized with DNA from the outlying species, the melting temperatures of the two hybrids will be identical, within the limits of experimental error, indicating that the hybrid DNA's contain equal numbers of mismatched bases. Thus the genetic distances between species *A* and *B* (*color*) and between species *A* and *C* (*gray*) are the same. Since the same length of time separates species *B* and *C* from the last common ancestor they shared with species *A*, the average rates at which their DNA has changed must be identical.

tor they shared with the tracer species, the DNA of both driver species must have changed at the same average rate. Our reconstructed phylogeny of living birds includes thousands of such trios of species that yield the same result and attest to the uniform average rate of the DNA clock in birds.

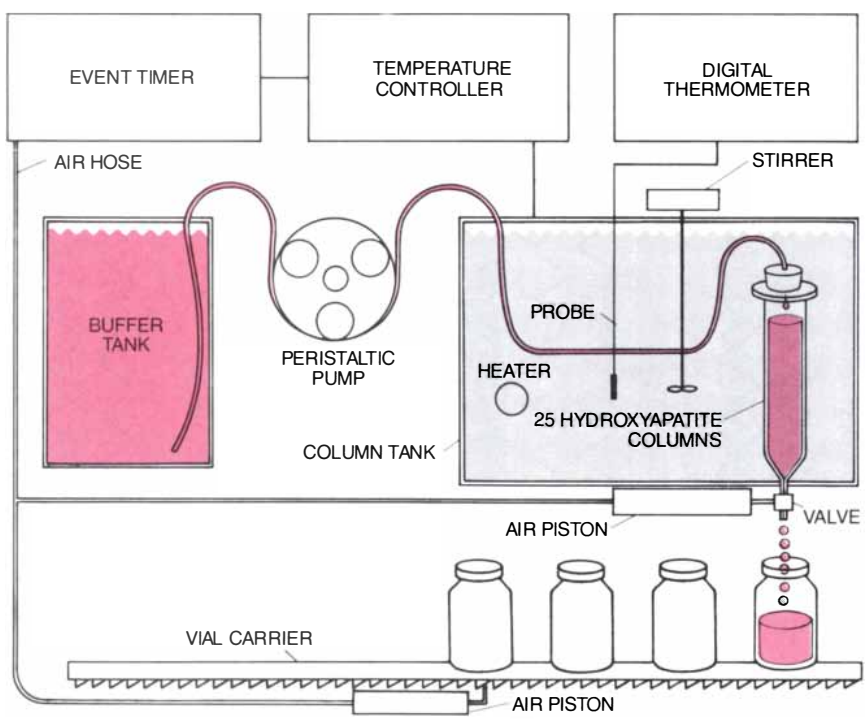
This apparent constancy may seem to be magic (or nonsense) at first glance, but it may be simply the result of measuring differences between sequences of billions of base pairs after millions of years of evolution. Natural selection dictates that different genes evolve at many different rates, and any individual gene may evolve at different rates at different times, but the range in the rates of all genes is narrow and the number of genes in the genome of a bird is huge. As the rate of evolution of one gene speeds up, another gene may be statistically likely to slow down by the same amount. The balancing of the rates of change in different genes need not occur simultaneously; the apparent constancy arises over millions of years. Whatever the correct explanation may be, avian DNA appears to evolve at a uniform average rate.

Because mismatched bases in DNA hybrids are the result of genetic changes that have been fixed in the two lineages since they last shared a common ancestor, the number of mismatches is proportional to the time the lineages have been diverging. The median melting temperature of a DNA-DNA hybrid is therefore an indirect measure of the length of time since the branching occurred. By correlating the median melting temperature with a dated geologic event that caused an ancestral species to be divided into separate lineages, we can calibrate the DNA clock in absolute time.

We have assumed, for example, that the common ancestor of the Ostrich of Africa and the rheas of South America ranged across the protocontinent of Gondwanaland before continental drift split the protocontinent into the southern continents of today and opened the Atlantic Ocean during the Cretaceous period. The geologic evidence indicates that the Atlantic became a barrier for flightless animals about 80 million years ago. Thus the Ostrich and rhea lineages must have diverged about then. Dividing 80 million years by the difference between the median melting temperature of Ostrich/rhea DNA heteroduplexes and that of Ostrich/Ostrich or rhea/rhea homoduplexes yields a calibration constant, in millions of years of divergence per degree of reduction in median melting temperature.

We have seven similar dated divergences between bird lineages caused by three geologic events, two of them about 80 million years ago and one some 40 million. Each instance yields a calibration constant of between 4.3 and 4.7, with the average at 4.5. Hence a median reduction in melting temperature of one degree C. is equivalent to about 4.5 million years since the two lineages shared their most recent common ancestor. This constant is tentative and subject to correction, but we use it to calculate approximate divergence dates.

During the past 10 years we have made more than 25,000 DNA-DNA hybrid comparisons, using genetic ma-

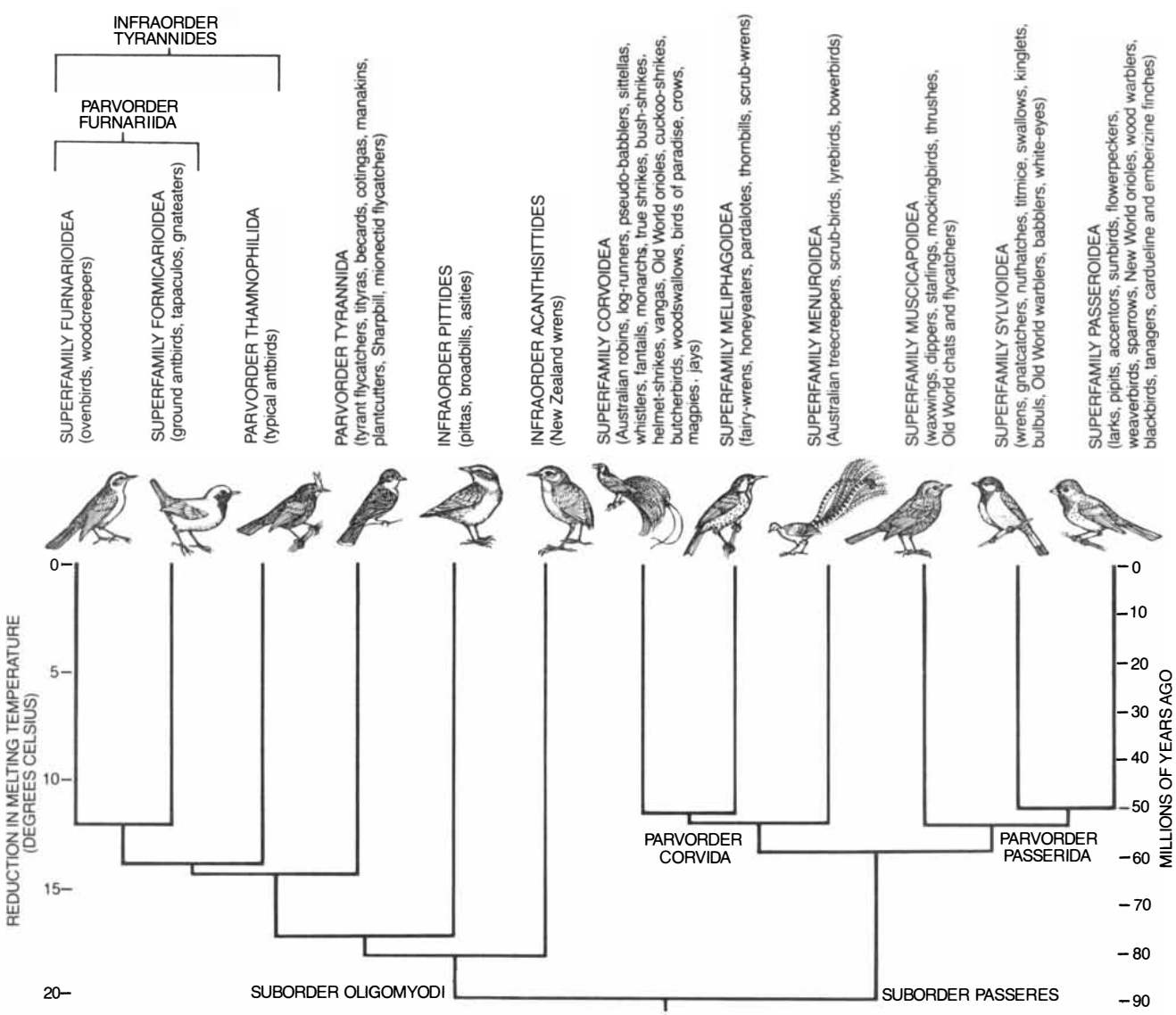


APPARATUS FOR DNA-HYBRID ANALYSIS, the DNAlyzer, processes 25 hybrids simultaneously. An event timer synchronizes the operation of the apparatus; a controller regulates the temperature of the water bath in which the hybrid DNA's, contained in a row of 25 hydroxyapatite columns, are immersed. When the timer releases a pulse of compressed air, the vial carrier moves vials into position under the columns and valves at the bottom of the columns open, allowing buffer solution supplied by the peristaltic pump to flow through the columns. The buffer washes the single-stranded DNA produced by the melting of the hybrids into the vials. The valves then close and the heater raises the water temperature in the column tank by 2.5 degrees C. before the timer releases another pulse and the next cycle begins. An assay of the amount of radioactively labeled DNA in each vial, done later, indicates how much of each hybrid DNA melted at each temperature.

terial from about 1,600 species, which together represent 168 of the 171 traditional families of living birds. From these data we have reconstructed the phylogeny of most of the groups of birds. The examples that follow illustrate some of the problems we have examined and the solutions indicated by the DNA hybridization data.

The barbets are small and usually brightly colored birds with tufts of bristles at the base of their relatively large bills. The Old World barbets live in Africa and southern Asia; the New World species live in the tropics of Central and South America. Traditionally the barbets have been placed in the family Capitonidae and were thought to be related to the woodpeckers (the Picidae) and to the large-billed, fruit-eating toucans (the Ramphastidae) of the New World tropics. Many taxonomists have remarked that the small species of toucans and the large species of New World barbets show similarities in morphology, and two recent studies have emphasized the close relationship between the barbets and the toucans. In 1984 Philip Burton of the British Museum of Natural History concluded from a study of the head region: "It seems reasonable simply to regard toucans as a specialized group of barbets which have arisen and radiated in South America." Similarly, in 1985 Lester L. Short, Jr., of the American Museum of Natural History stated: "Toucans effectively are large, specialized, toothbilled barbets."

The DNA comparisons agree with Burton and Short and enable us to add the dimension of time to their conclusions. The DNA indicates that the branching between the Old World and the New World barbets took place about 55 million years ago. The toucans branched from the New World barbet lineage more recently, about 30 million years ago. Hence the toucans are more closely related to the New World barbets than the two groups of barbets are to each other. To reflect these relationships our classification of these groups and the others belonging to the same order looks like this:



PHYLOGENY OF PASSERINE BIRDS, which include 5,300 of the 9,000 species of living birds, was reconstructed from DNA comparisons. For each bifurcation of the tree, the scale at the left shows the number of degrees by which the median melting point of hybrids formed of DNA's from species representing the two lineages is lowered with respect to perfectly matched DNA hybrids; the scale at the right dates the branchings. The reconstruction affects traditional classifications of living birds. In the suborder Oligomyodi, to which most of the passerine birds of South America belong, the DNA showed that ground-dwelling antbirds belong to a lineage different from that of the typical antbirds. In the suborder Passeres the results delineated two distinct groups: the parvorders Corvida and Passerida. The Corvida originated in Australia, although some groups in the Corvida are now distributed throughout the world.

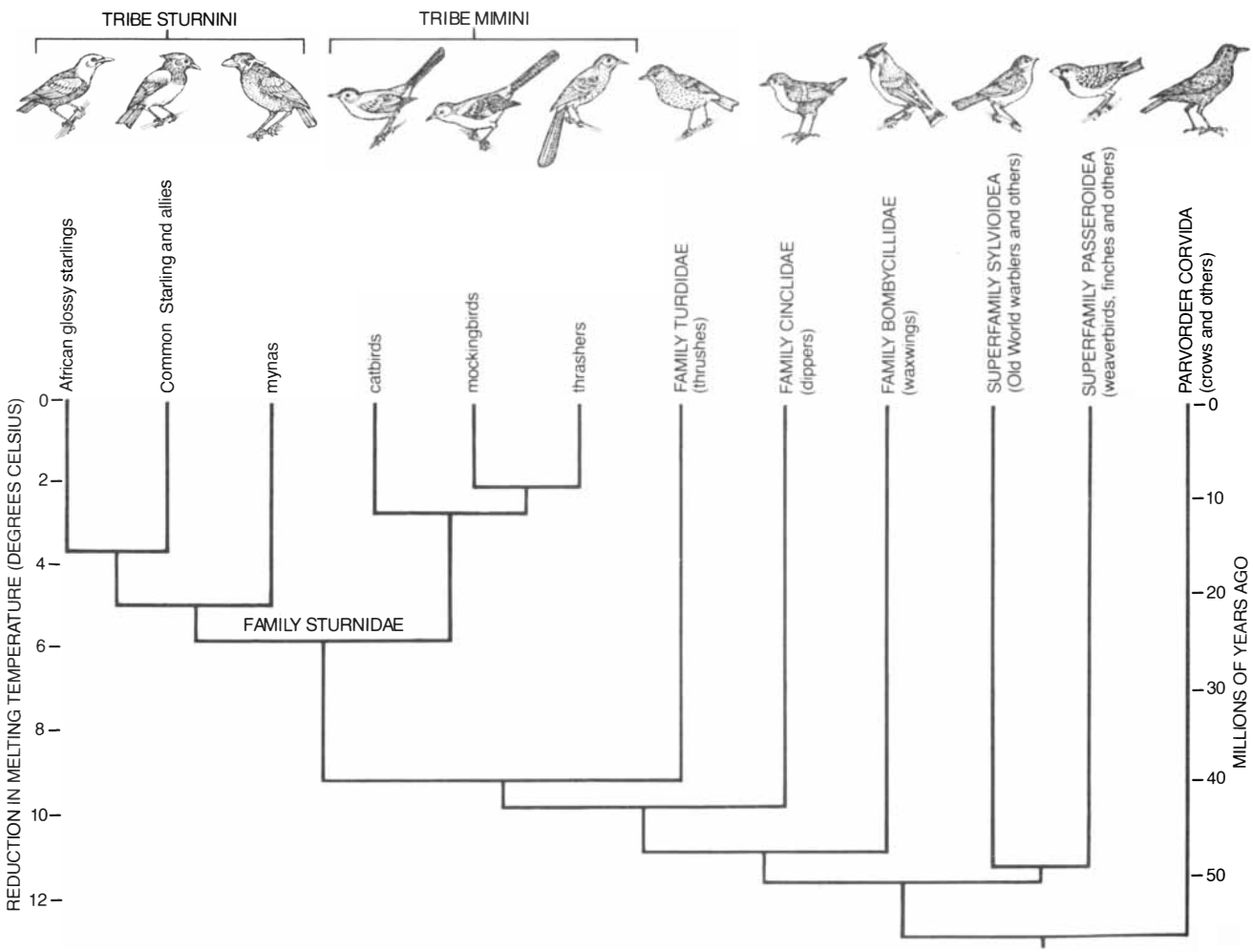
Order Piciformes
  Parvorder Picida
    Family Picidae (woodpeckers)
    Family Indicatoridae (honeyguides)
  Parvorder Ramphastida
    Superfamily Megalaimoidea
      Family Megalaimidae (Old World barbets)
    Superfamily Ramphastoidea
      Family Ramphastidae
        Subfamily Ramphastinae (toucans)
        Subfamily Capitoninae (New World barbets)

Our classification is based on the branching pattern of the phylogeny and the times of origin of the groups, as dated by the DNA comparisons. We divided the evolutionary time scale into segments of 10 million years and assigned a taxonomic category to each segment. Orders, in our scheme, are those lineages that branched from other lineages 90 to 100 million years ago and suborders are those that branched 80 to 90 million years ago. Infraorders originated 70 to 80 million years ago, parvorders 60 to 70, superfamilies 50 to 60, families 40 to 50, subfamilies 30 to 40 and tribes 20 to 30 million years ago. One result of the procedure is to make groups at the same categorical level approximately equal in their degree of evolutionary divergence. The boundaries we have constructed are not rigid, but they provide the basis for a classification that reflects the DNA-derived phylogeny and approaches the ideal of equivalent categories.

There is controversy, however, over how to classify organisms; some taxonomists prefer to base categorical rank on their evaluation of the degrees of morphological specialization of groups. On that basis the two groups of barbets might be placed in the same family and the toucans in an adjacent family or superfamily. Such a classification recognizes the birds' distinctive appearances but conceals their phylogenetic relationships. In one traditional arrangement, currently in wide use, the toucans, barbets and some other groups are classified as follows:

Order Piciformes
  Suborder Galbulae
    Superfamily Galbuloidea
      Family Galbulidae (jacamars)
      Family Bucconidae (puffbirds)
    Superfamily Capitonoidea
      Family Capitonidae (barbets)
      Family Indicatoridae (honeyguides)
    Superfamily Ramphastoidea
      Family Ramphastidae (toucans)
  Suborder Pici
      Family Picidae (woodpeckers)

Not only is this classification at odds with the DNA data on the toucans and barbets, but also the DNA indicates that the jacamars and puffbirds of the New World should be placed in a separate order, not in the Piciformes.



STARLINGS AND MOCKINGBIRDS occupy adjacent branches of the avian phylogenetic tree, according to data from DNA-DNA hybridization. The starlings (the tribe Sturnini) had been seen as related to the crows; the mockingbirds (the tribe Mimini) had been placed near the thrushes. If that classification were correct, the ancestors of the mockingbirds and the starlings would have diverged almost 60 million years ago. Hybrid melting-point data showed instead that the starlings and mockingbirds are each other's closest relatives, having diverged about 25 million years ago. Both groups, which make up the family Sturnidae, are related to the thrushes.

There is no sign that the debate about how to classify plants and animals will be resolved soon.

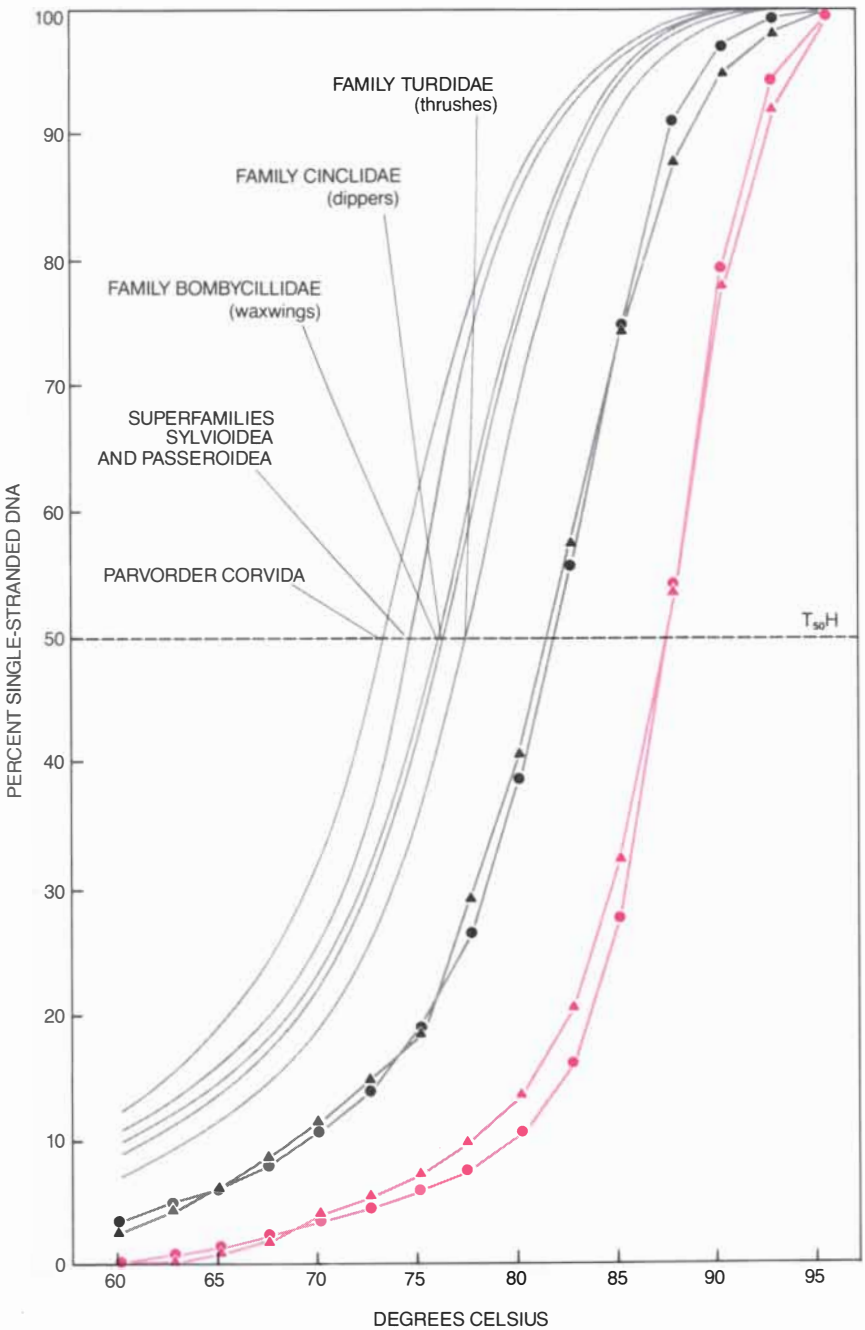
The vultures of the Old World are closely related to the hawks and eagles. The New World vultures, including the condors, the Turkey Vulture and the Black Vulture, superficially resemble the Old World vultures. Both groups are carrion eaters, and they have usually been placed together in the order Falconiformes: the diurnal birds of prey. The New World vultures share many morphological traits with the storks, however, and some taxonomists have argued that the New World vultures and the storks belong in the same order. This arrangement was proposed by Alfred B. Garrod in the 1870's and by David Ligon, then at the University of Michigan, in 1967, but recent classifications have ignored the evidence of a condor-stork alliance and continued to place the condors and their relatives in the Falconiformes.

The DNA comparisons support the morphological indications suggesting the New World vultures and the storks are each other's closest living relatives; the data show the groups diverged from a common ancestor about 35 to 40 million years ago. Similarities between the vultures of the New and Old worlds are the result of convergent evolution related to their carrion-eating habits.

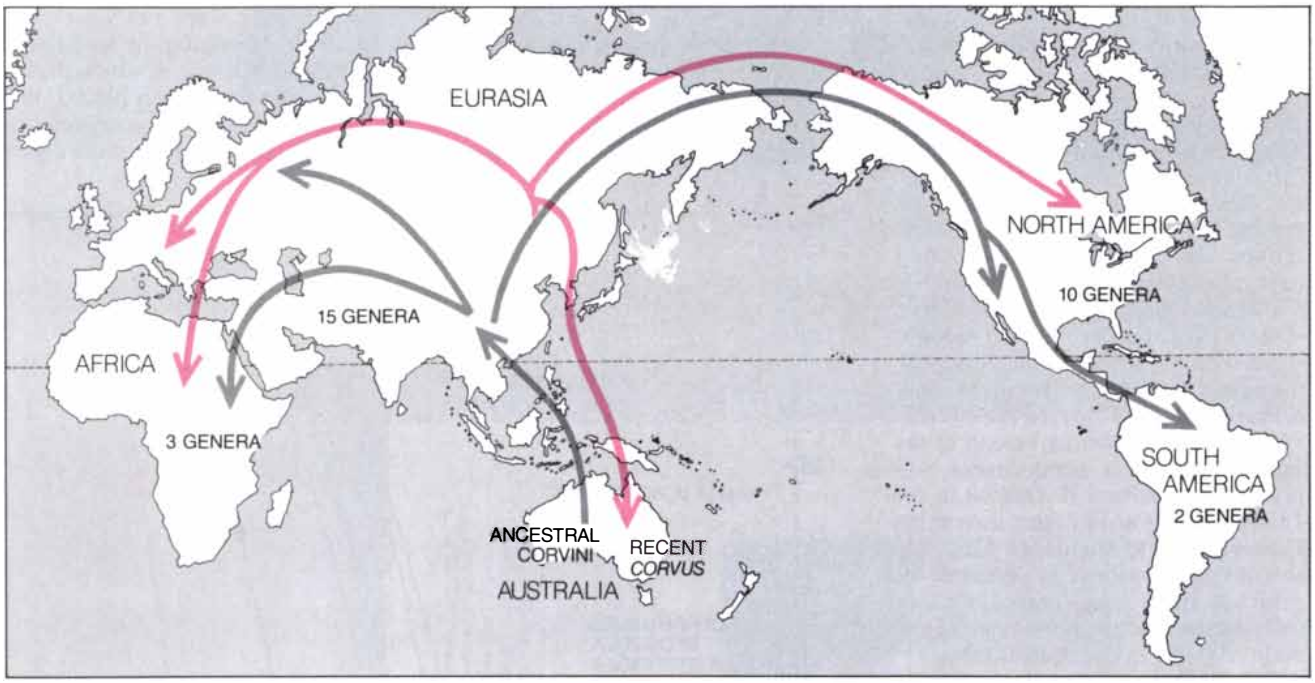
The totipalmate birds are those with webs between all four toes: pelicans, cormorants, anhingas, boobies, gannets, frigatebirds and tropicbirds. All except the tropicbirds also have an obvious gular pouch, or throat pouch, between the branches of the lower mandible; in tropicbirds the gular pouch is small and concealed. Mainly because of these two shared characters the totipalmate birds have usually been grouped together as members of the order Pelecaniformes, although it has sometimes been suggested that tropicbirds and frigatebirds are related to some other group.

Another bird, the Shoebill, has been proposed as a member of the same cluster. The Shoebill, a large storklike bird with a huge bill, lives in the swamps of eastern Africa and feeds on lungfishes and other aquatic prey. The Shoebill has usually been viewed as a relative of the storks or the herons, but a 1957 study of its skeleton led Patricia A. Cottam, then at the British Museum, to conclude that the Shoebill is most closely related to the pelicans. Cottam's evidence was dismissed by most taxonomists as the result of convergent evolution, but our DNA comparisons support her. The divergence between the Shoebill and the pelicans took place about 35 to 40 million years ago. The nearest cluster to the pelican-Shoebill group turns out to be the group that includes the storks and the New World vultures. The two groups diverged 40 to 45 million years ago.

The surprises were not over. The DNA also showed that the traditional order Pelecaniformes, in which the totipalmate birds had been placed, is a polyphyletic group: it is composed of several subgroups that are more close-



MELTING CURVES OF DNA HYBRIDS indicate the differences between the DNA's of starlings and mockingbirds and between starling or mockingbird DNA and DNA's of some other bird lineages. The horizontal scale shows temperatures to which DNA hybrids were heated; the vertical scale indicates the percentage of the DNA that melted at each temperature, forming single strands. The colored curves show the melting properties of DNA hybrids in which both strands were from the Long-billed Thrasher (*dots*), a close relative of the mockingbirds, or from a glossy starling (*triangles*). The unlabeled black curves represent hybrids between the thrasher and various starlings (*dots*) or between the starling and various mockingbirds (*triangles*). The remaining curves show the averaged melting properties of DNA hybrids consisting of one strand of mockingbird or starling DNA and one strand of DNA from the indicated bird group. The temperature at which half of the hybrid DNA melted ($T_{50}H$) is used for comparing hybrids. The lower the $T_{50}H$ of a DNA hybrid is, the poorer the match is between the DNA's and the more distantly the species are related.



DISPERSAL OF TRIBE CORVINI, which includes crows, ravens, jays, magpies and their relatives, is mapped (*gray arrows*). DNA-DNA hybridization identified the Corvini as part of a larger group of birds that evolved in Australia. The number of genera belonging to the Corvini on each continent reflects the order in which the continents were invaded after ancestors of the family crossed from Australia into Asia about 30 million years ago. The largest number of genera are found in Asia and the fewest in South America, which the Corvini reached only from three to five million years ago. The crows and ravens (genus *Corvus*) evolved in Asia, and they later spread to Africa and North America (*colored arrows*). Only recently, within the past 100,000 years or so, has the group colonized Australia.

ly related to other birds than they are to one another. The pelicans are most closely related to the Shoebill, storks and condors. The cormorants, anhingas and boobies are allied with one another and seem to be distantly related to the herons; the frigatebirds are relatives of the tube-nosed seabirds (the albatrosses and petrels), and the tropicbirds appear to represent a separate lineage with no close living relatives. Hence the totipalmate foot and the gular pouch either evolved by convergence in separate lineages or were inherited from a distant common ancestor of several groups of birds. In descendants that do not show these characters their genetic basis may have been repressed. Such dormant genes are known in some groups of birds and in other animals.

Polyphyletic clusters of bird lineages have been revealed in the past. In some early classifications the birds with palmate feet (feet with webs between the three front toes only) were grouped together. It was soon realized that not all palmate birds, which include the ducks, albatrosses, penguins, loons, gulls and auks, are closely related, and they were assigned to several different groups. In the case of the totipalmate birds, however, with the gular pouch as supporting evidence, it seemed impossible that they did not belong to a monophyletic cluster—one in which all the members share the same most recent common ancestor.

In a recent study Joel Cracraft of the University of Illinois Medical Center at Chicago compared 45 skeletal and seven behavioral characters among the totipalmate birds and the penguins, loons, grebes, albatrosses, petrels and the Shoebill. He found 12 traits supporting his hypothesis that the totipalmate species form a monophyletic group and six suggesting that the albatrosses and petrels are the sister group, or companion lineage, of the totipalmate birds. Among the totipalmate species, the tropicbirds emerged as a lineage distinct from the other totipalmate birds. Of the others, the frigatebirds appeared to be the descendants of the oldest branch, followed in order of branching by the pelicans, boobies, anhingas and cormorants.

Cracraft rejected the Shoebill as a relative of the totipalmate birds and ascribed its similarities to the pelicans to convergence. There are obvious disagreements between the DNA comparisons and the morphological evidence Cracraft used, but there are also several congruencies: the relationship between the totipalmate species (the frigatebirds in particular) and the albatrosses and petrels, the outlying position of the tropicbirds and the close relationships among the boobies, anhingas and cormorants. The major disagreements concern the relationship between the pelicans and the Shoebill and the position of the pelicans with respect to the other totipalmate birds.

It is unlikely that the DNA-based evidence of the polyphyly of the order Pelecaniformes will be accepted by most ornithologists soon. Nevertheless, we predict that appropriate comparisons will reveal morphological evidence that, like Cottam's 1957 study, is consistent with the DNA data.

The sandgrouse are pigeon- or ploverlike birds of the arid regions of Africa, Asia and southern Europe. Their relationships to other birds have been debated for more than a century. Are they related to the pigeons, to the plovers or to the galliform birds (chickens, pheasants and their relatives)? Each group has had proponents; most of the recent participants in the dispute support the pigeons or the plovers.

The DNA results are quite clear: sandgrouse are the sister group of a large part of the order Charadriiformes, including the sheathbills, thick-knees, plovers, oystercatchers, avocets, stilts, gulls, auks and coursers. The sandgrouse and their sister assemblage are in turn most closely related to the sandpipers and their allies. Hence the sandgrouse are neither pigeons nor plovers but are closer to the plovers than they are to the pigeons;

similarities between the sandgrouse and the pigeons must be due to convergent evolution.

About 5,300 of the 9,000 species of living birds belong to the order Passeriformes; it includes the flycatchers, warblers, thrushes, sparrows, starlings, wrens, swallows, larks, crows and other species, generally small. Most of the passerine birds of South America are members of the suborder Oligomyodi, also known as the suboscines. The structure of the syrinx (the vocal apparatus) and other anatomical characters distinguish the suboscines from the oscines, or songbirds, which make up the suborder Passeres (the other branch of the Passeriformes). The New World suboscines evolved in South America while it was isolated from the rest of the world as an island continent from about 80 million years ago in the late Cretaceous period until about five million years ago.

One evolutionary lesson taught by the New World suboscines concerns the tropical antbirds, a group of about 235 species that have traditionally been placed in the family Formicariidae. In the early 1960's Mary Heimerdinger Clench and Peter L. Ames, both at Yale University, found that some antbirds have two deep notches in the posterior edge of their sternum, or breastbone; the other species have four notches. Ames also found that the two groups differ in the musculature of the syrinx. The 185 two-notched species occupy a range of habitats, but the 50 four-notched species are long-legged, short-tailed ground dwellers. Clench and Ames also pointed out that four-notched sterna occur in two other groups of birds: the tapaculos (family Rhinocryptidae) and the gnateaters (family Conopophagidae). Clench and Ames suggested that the ground-dwelling antbirds are more closely related to those groups than they are to the other antbirds.

The DNA revealed the same alliances. The comparisons showed that the two-notched antbirds branched from the four-notched lineage before the four-notched antbirds diverged from the tapaculos and the gnateaters. Thus the morphological and molecular evidence agreed, and the DNA data provided the branching order and the approximate dates of the divergences.

The other branch of the passerines, the suborder Passeres, includes about 4,000 of the 5,300 passerine species. The DNA comparisons revealed that the Passeres consists of two major groups, which we call the parvorders Passerida and Corvida. These two lineages diverged from a common ancestor about 55 to 60 million years ago. The evidence shows the Passerida evolved in Africa, Eurasia and North America and the Corvida in Australia.

From about 60 to 30 million years ago, during the early and middle Tertiary period, Australia was isolated from other landmasses. The Corvida evolved many morphologically and ecologically specialized forms, including warblers, flycatchers, creepers, thrushes, babblers and nectar feeders—forms much like those to which the Passerida gave rise in other parts of the world. The birds of Australia were discovered and named, however, after European ornithologists had already classified those of most of the rest of the world. The Australian passerines seemed to fit into categories that had been founded on specimens from other areas. Thus the Australian warblerlike passerines were assigned to the Sylviidae (which includes the true warblers), the Australian flycatchers to the Muscicapidae (the Afro-Eurasian flycatchers), and the treecreepers to the Certhiidae (the Eurasian-American creepers). The sitellas, nuthatchlike birds of Australia, were placed in the Sittidae (the family of true nuthatches) and the Australian honeyeaters were grouped with the superficially similar nectar-eating Afro-Asian sunbirds.

When we compared DNA's among various Australian passerines and between Australian species and their supposed relatives from Africa, Eurasia and North America, we found that the Australian endemics are more closely related to one another than they are to their morphologically similar counterparts from other continents. Convergent evolution had produced similarities between unrelated species of the two parvorders, and museum taxonomists had assembled the species into polyphyletic clusters containing members of both the Corvida and the Passerida. The same mistake had been made for most of the 400 species of passerines in Australia and New Guinea today. Many of the convergences are so subtle that the true relationships of the Corvida and the Passerida probably could not have been resolved from anatomical comparisons alone.

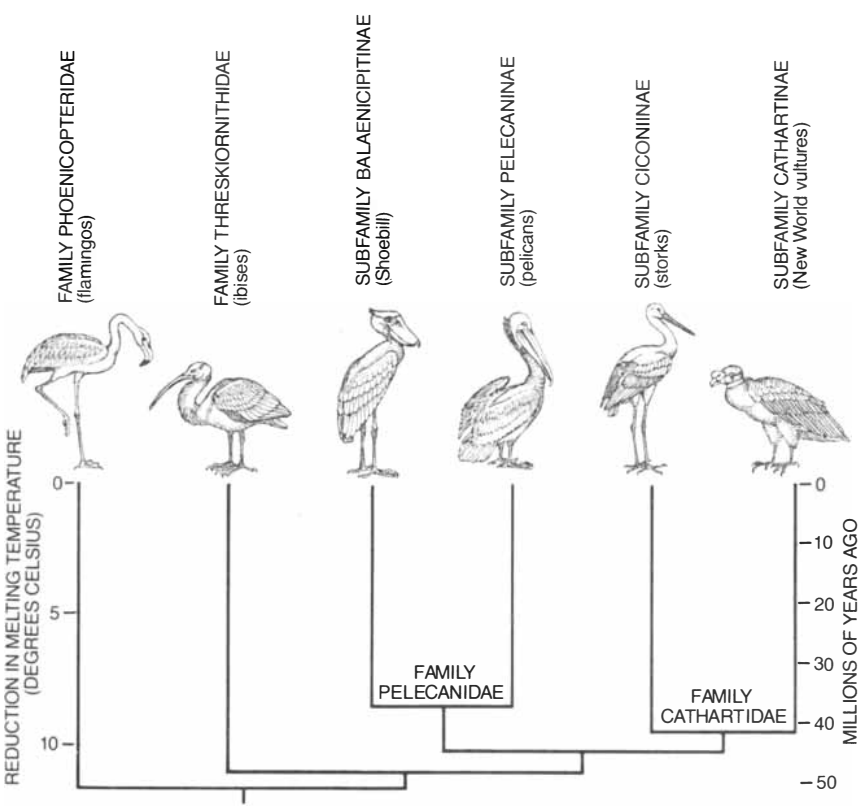
One result of the confusion was to conceal the fascinating story of the phylogeny of the Corvida, which parallels that of the marsupials. Both groups evolved while Australia was isolated; like the Australian passerines, the marsupials radiated into many of the same ecological niches occupied in Africa, Eurasia and North America by another group, in this case the placental mammals. In the process some of the marsupials assumed forms similar to those of other mammals. Unlike the

Australian passerines, however, the marsupials were not confused with their counterparts elsewhere because they have a marsupium, or pouch, and other distinctive traits.

The Corvida of Australia produced the ancestors of a few groups that were able to emigrate to Asia as Australia drifted northward during the Tertiary period. Among the groups whose ancestors evolved in Australia is the tribe Corvini, which includes the same species as the traditional family Corvidae: the crows, ravens, jays, magpies and their relatives.

Today there are 23 genera of the Corvini; of these 15 occur in Eurasia and 10 in North America. Only three genera are found in Africa, and in South America the tribe is represented only by two genera of jays. The numbers reveal the chronology and pattern of the dispersal from Australia. The earliest radiation occurred in southeastern Asia, and members of the lineages later extended their ranges to Europe, Africa and North America. South America was the last continent to be invaded. It was isolated from North America until between three and five million years ago; the two genera of jays that now occur in South America apparently expanded their ranges from the north after the emergence of Central America provided a land connection between the continents. The crows and ravens of the genus *Corvus* probably originated in Eurasia and spread nearly worldwide (except to South America), eventually even colonizing Australia.

Among the many surprising results from the comparisons of DNA's none was less expected than the discovery of the close relationship between two other groups of passerines: the starlings, which are native to the Old World, and the mockingbirds and thrashers of the New World. The starlings have usually been considered related to the crows, and the mockingbirds were correctly placed near the thrushes. If the starlings were close relatives of the crows, they would be members of the Corvida, but the DNA clearly identifies the mockingbirds as members of the Passerida. Thus, if the traditional classification were correct, the starling and mockingbird lineages would have diverged between 55 and 60 million years ago.



NEW WORLD VULTURES AND STORKS are each other's closest living relatives, according to the phylogeny reconstructed from DNA comparisons. Superficially the New World vultures (which include the condors) resemble the vultures of the Old World. The similarities must be due to convergent evolution; the Old World vultures belong to a different cluster of lineages. DNA studies also showed that the pelicans and the Shoebill, an African species, are the nearest neighbors of the storks and condors in the phylogenetic tree. In addition the DNA confirmed the close alliance between the Shoebill and the pelicans.

The DNA comparisons revealed, however, that the two lineages diverged about 25 million years ago. Other work supports this close starling-mockingbird alliance: immunological comparisons of muscle proteins done in 1961 by William B. Stallcup, Jr., of Southern Methodist University, studies of the anatomy of the head region done in 1953 by William J. Beecher, then at the University of Chicago, and comparisons of the syrinx made by Wesley E. Lanyon of the American Museum of Natural History. It may also be significant that some starlings—myna birds, for example—are, like mockingbirds, excellent mimics. Even the Common Starling mimics the songs of other birds.

The close relationship between the starlings and the mockingbirds may reflect the history of climatic change in the Northern Hemisphere. During the early and middle Tertiary period, 65 to 30 million years ago, the climate of the Arctic was temperate; broad-leaved trees grew in northern Canada and Greenland. It seems probable that the common ancestor of the starlings and the mockingbirds was widely distributed over those regions, which served as a bridge between the Old World and the New World. Evidence from plant fossils indicates that the climate grew colder beginning about 30 million years ago, and the ancestral populations presumably moved southward. About 25 million years ago the populations in America and Eurasia became separated and began to diverge.

These are just a few of the discoveries DNA-DNA hybridization has made possible. The DNA comparisons present us with new hypotheses of avian relationships. If the DNA-based phylogeny of birds is closer to the one true phylogeny, it will be congruent with evidence from other sources. The results we have so far obtained agree with geologic history better than many earlier proposals about avian phylogeny, and the relationships indicated by the DNA data are usually supported by at least some anatomical characters. We believe some aspects of morphology will prove to be congruent with the DNA evidence in all cases.

Yet DNA and the morphological characters traditionally used to reconstruct phylogeny serve to provide different kinds of information. Morphology shows how natural selection has modified structure to adapt organisms to the environment, whereas DNA comparisons give a direct indication of the branching pattern and the approximate branching dates among living lineages. Morphology is functional; the DNA clock keeps time.

MERKUR XR4Ti FROM GERMANY. PERFORMANCE THAT GOES BEYOND MERE STATISTICS.

Statistics can reveal a great deal about an automobile. But there is an equally important aspect of performance that numbers simply cannot express.

The way a car feels.

That's why it isn't enough to know that the Merkur XR4Ti has a turbocharged overhead cam engine that will accelerate it from rest to 60 mph in only 7.8 seconds. Or that its fully independent suspension mated with Pirelli P6 tires can achieve .800 g's on the skid pad.

Knowing XR4Ti has certain statistical advantages over its competitors from BMW, Saab, Volvo and Audi just isn't enough.

You have to take it out and feel it. Take it through a fast sweeper or a series of challenging turns and absorb the sensations it transmits through your body. It is only then that the true value of its statistics can be appreciated.

Anyone can build a car to generate impressive statistics in the sterile confines of a test track. Building one that surpasses your expectations in the real world is the true challenge.

For more information call toll-free: **1-800-445-4500.**

Ford

IMPORTED FROM GERMANY
EXCLUSIVELY FOR LINCOLN-MERCURY

MERKUR

MERKUR XR4Ti

Inca Stonemasonry

Enormous stone blocks were quarried, shaped and fitted so closely that a knife blade cannot be inserted into the joints. How was it done? To provide answers the author cut stones in an Inca quarry

by Jean-Pierre Protzen

It was only in the 100 years or so before the Spanish conquest of 1532 that Inca culture reached its height. During that century Inca society was transformed from a minor agricultural state in central Peru into a mighty empire stretching from Chile to Ecuador. One aspect of the cultural flowering was an ambitious program of new construction initiated by Pachakuti, the ninth Inca (or emperor), in 1438. Pachakuti ordered his stonemasons to rebuild Cuzco, the capital of the emerging empire. The rebuilding did not come to a stop at the death of the ninth Inca. His successors extended the new construction far beyond the boundaries of Cuzco. Throughout Peru temples, palaces, warehouses and waterworks were thrown up, breaking new ground or replacing older structures.

Pachakuti's construction agenda was not only ambitious but also technically innovative. Although most earlier Inca structures were probably built of adobe or mud-bonded stones, the new work was done entirely without mortar. Stone blocks weighing as much as 100,000 kilograms (about 220,000 pounds) were fitted so closely to their neighbors that even now a knife blade cannot be inserted into many of the joints.

For hundreds of years visitors to Peru have been intrigued by the size of the blocks in the Inca stonework and the precision with which each block is inserted among its neighbors. The fact that the Incas had no iron tools makes the stonework even more impressive. In 1979, on my way back to the U.S. from a temporary teaching appointment in Brazil, I visited some of the main Inca sites and marveled at the ingenuity of their construction. When I asked my guides how the Incas shaped the great stones and assembled them into buildings, however, I got answers that were less than satisfying. On returning to the University of California at Berkeley, where I teach architecture, I asked colleagues who are archaeologists about the scholarly literature on Inca stonemasonry. To my surprise I was told there was none.

Although I am not an archaeologist, I have a keen professional interest in construction techniques. After turning the matter over I decided to investigate the Inca walls on my own. A sabbatical leave in 1982 provided the opportunity to spend six months in Peru; since then I have returned for about a month each year. My investigation did not stop at the stage of hypotheses. When I had formulated a hypothesis, I put it directly to the test. Using materials available at the Inca sites, I cut, dressed and fitted stones to show that these tasks could have been carried out by the Incas as I propose. Some mysteries remain, particularly in the area of how the big stones were transported and handled at the building site, but by and large my investigation was successful. As a result speculation about how the Incas built their beautiful stone structures can now begin to be replaced by empirical findings.

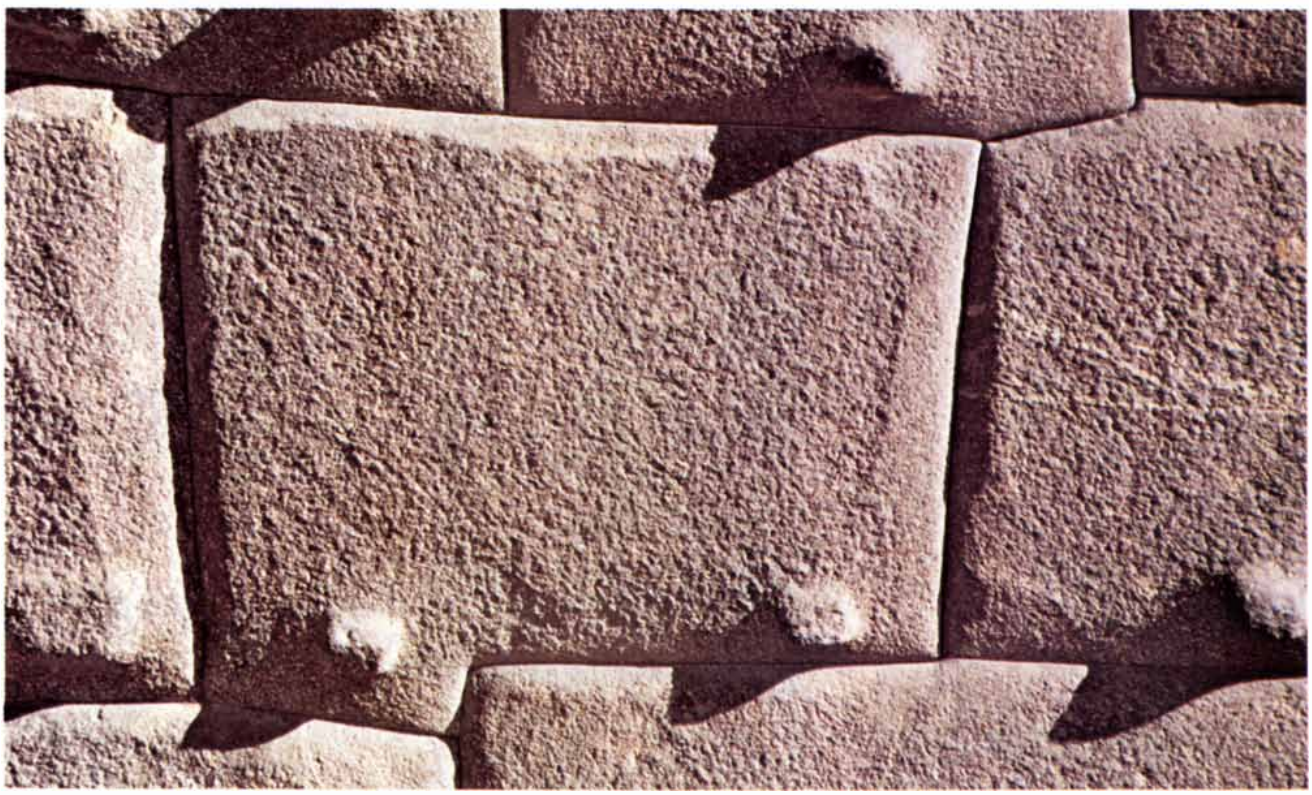
Much of my research involved analyzing specific Inca walls at Cuzco itself, and at the "fortresses" of Saqsaywaman and Ollantaytambo. Saqsaywaman is near Cuzco and Ollantaytambo is on the Urubamba River about 90 kilometers northeast of the Inca capital. Although many texts refer to Saqsaywaman and Ollantaytambo as forts, recent archaeological investigations suggest they had a religious function rather than a military one. Whatever role they had in Inca society, the two sites are impressive from the point of view of construction techniques. Saqsaywaman is a very large site that includes outworks made up of three separate stone walls, each one more than three meters tall. Ollantaytambo, built on the spur of a mountain, included a religious center, a royal estate and a town planned on a grid.

How were these great stone structures built? To make the question more manageable I divided it into four parts: the quarrying of the stone, the cutting and dressing of individual blocks, the fitting of the blocks and transportation. To investigate the quarrying I visited several Inca quarry sites, of which two—Kachiqhata and Rumiqolqa—were analyzed in detail. Kachiqhata lies about four kilometers across the Urubamba from Ollantaytambo. Its quarries supplied the porphyry (red granite) of which the Sun Temple, the most important structure at Ollantaytambo, is built. Rumiqolqa is 35 kilometers southwest of Cuzco; it supplied much of the andesite (an igneous rock) that Pachakuti's masons used as they rebuilt the imperial capital.

Several bits of circumstantial evidence indicate that quarrying was a matter of great significance to the Incas. Kachiqhata and Rumiqolqa are remote, difficult to reach and far from the construction sites where the stone blocks were assembled. The Incas' motive for exploiting such inconvenient quarries surely must have been that they put a high value on the type of stone found there.

Furthermore, the internal organization of the quarries shows that careful attention was paid to the process of obtaining building stone. Both Rumiqolqa and Kachiqhata have networks of access roads leading to the points where the building stones were retrieved. The quarries of Kachiqhata are reached by a road leading down from Ollantaytambo, across the Urubamba and up the far bank of the river to a series of large rockfalls, where rocks split off naturally from the face and accumulate in great piles. Where it reaches the rockfalls the access road divides into several branches extend-

SAQSAYWAMAN, near Cuzco, is the site of some of the most impressive Inca stonework. The photograph shows part of an outwork system that includes three walls, each one more than three meters high. The largest stones in these walls weigh about 100,000 kilograms. Although Saqsaywaman is often referred to as a fort, recent archaeological work suggests that it was probably a religious center.



REMARKABLE FIT of the Inca building stones is shown in a photograph of a wall from Ollantaytambo. The material of the blocks is a type of stone called meta-arkose. The protrusions were used in handling the stones at the construction site; they were often left on the blocks after the wall was finished. The blocks are covered with small scars made by the hammers that were employed to shape the stone. The scars are finer at the edges than in the center of the face, which suggests different hammers were applied to the two areas.

ing to the quarry sites. The road can readily be traced because it is reasonably well preserved and is lined with about 80 abandoned Inca blocks.

On uphill grades, on the flat and on mild downhill grades the access network consisted of ramps that were probably originally covered with gravel. On steep downhill grades the ramps are replaced by slides down which the blocks were allowed to plunge freely. The longest of the slides at Kachiqhata is an awesome 40-degree slope with a 250-meter vertical drop; at the bottom are four abandoned blocks. The quarries of Rumiqolqa have been worked extensively since the conquest and are not as well preserved as the ones at Kachiqhata. Even at Rumiqolqa, however, a network of roads leading to the quarry sites can be traced. At both quarries the Incas supplemented the access roads with other structures such as retaining walls, water canals and living quarters.

Although the two quarry sites are similar in plan, the Inca quarrymen used methods at Kachiqhata that were slightly different from those employed at Rumiqolqa. At Kachiqhata the Incas did not undertake quarrying in the technical sense, which implies that the stone is cut from a rock face or detached from the bedrock by undercutting. Instead the quarrymen simply combed the giant rockfalls and selected the blocks of coarse-grained red granite that met their specifications. My observations suggest that when a block had been selected at Kachiqhata, it was only minimally worked before being transported to Ollantaytambo. The later stages of dressing the stone and the adjustments for fitting appear to have been carried out at the construction site.

At Kachiqhata the rough work on a block was often begun before the ramp leading to the block had been completed. That this was so is particularly clear at the end of the highest ramp in the south quarry at Kachiqhata. Near the end of the ramp two huge blocks, one 4.5 by 2.5 by 1.7 meters, the other 6.5 by 2.7 by 2.1 meters, are raised on stone working platforms. Although the blocks are partially dressed, the access ramp does not extend to the platforms on which they stand.

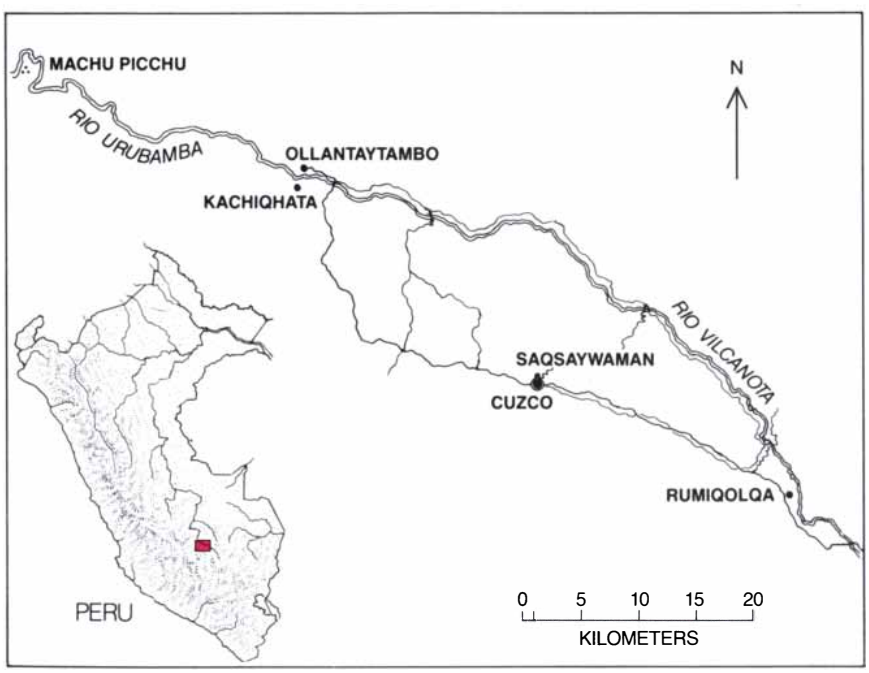
Intriguingly, the cutting marks on those blocks and on others in the Inca quarries are very similar to marks found on the pyramidion of the unfinished obelisk from Aswan in Egypt. (The pyramidion is the small triangular form at the top of an obelisk.) Both the pyramidion from Aswan and the stones at Kachiqhata have cuplike depressions on their surface. It is known that the Egyptians shaped their stones by pounding away at the workpiece with balls of dolerite (an igneous rock). It seems reasonable to think the Incas did the same.

After a careful search of the ground in the quarry at Kachiqhata I found some rounded stones of quartzite, a metamorphosed sandstone that does not occur naturally among the stones of the quarry but is present along the banks of the nearby Urubamba. An examination of the quartzite stones revealed pit marks on the smaller end of the stones, which indicates they were employed for pounding. I conclude that the Inca quarrymen at Kachiqhata picked up rounded river cobbles on the banks of the Urubamba and used them as hammers for imparting a rough shape to the blocks before the process of dressing was completed at Ollantaytambo.

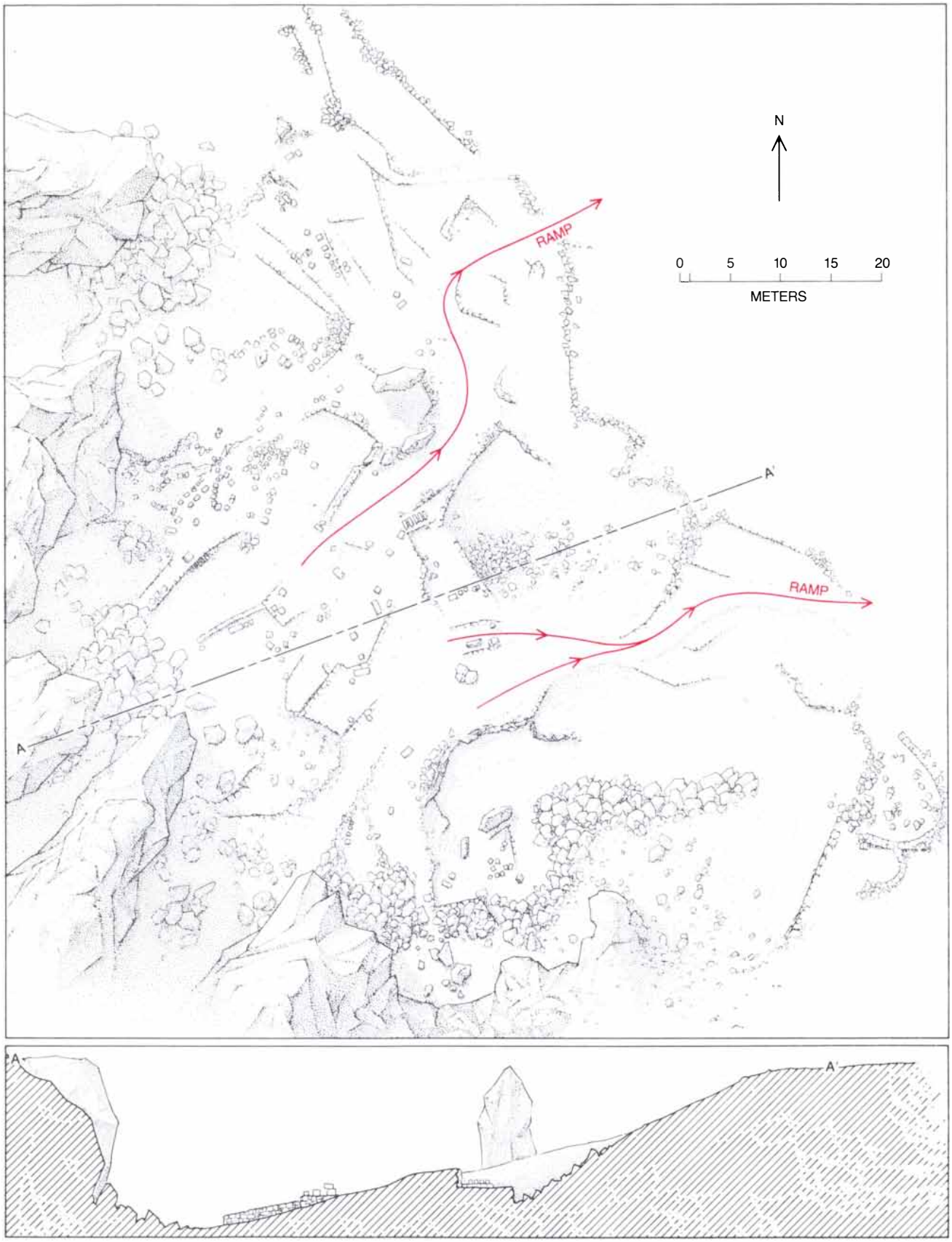
At Kachiqhata, then, stones were selected from the rockfalls rather than quarried in the technical sense and were only roughly dressed before being transported. At Rumiqolqa, on the other hand, there was true quarrying: the rock was broken off the face. Because Rumiqolqa has been worked since the conquest and is still being worked today, much of the evidence of the Inca exploitation of the rock has been obliterated. I did, however, succeed in finding one well-preserved Inca quarry pit in an area of Rumiqolqa that is hard to reach and therefore has not been worked in modern times. I named it the Llama Pit for the two petroglyphs of llamas carved on one of the rock faces.

The Llama Pit turned out to be a rich source of information about how the Incas quarried and dressed their building stones. Quarrying the andesite at Rumiqolqa does not pose major technical problems. Even the densest rock is fractured enough in its natural state to be broken out of the face quite easily. The quarrymen may have pried the stones they wanted from the rock face with pry bars like those that have been found at other Inca sites. The pry bars, which are made of bronze, are about a meter long; they have pointed ends and a rectangular cross section four or five centimeters on a side. The stone at Rumiqolqa is so fractured, however, that bronze pry bars would not have been necessary. I have seen quarrymen break the stones out of the face with sticks, and the Incas may have done the same.

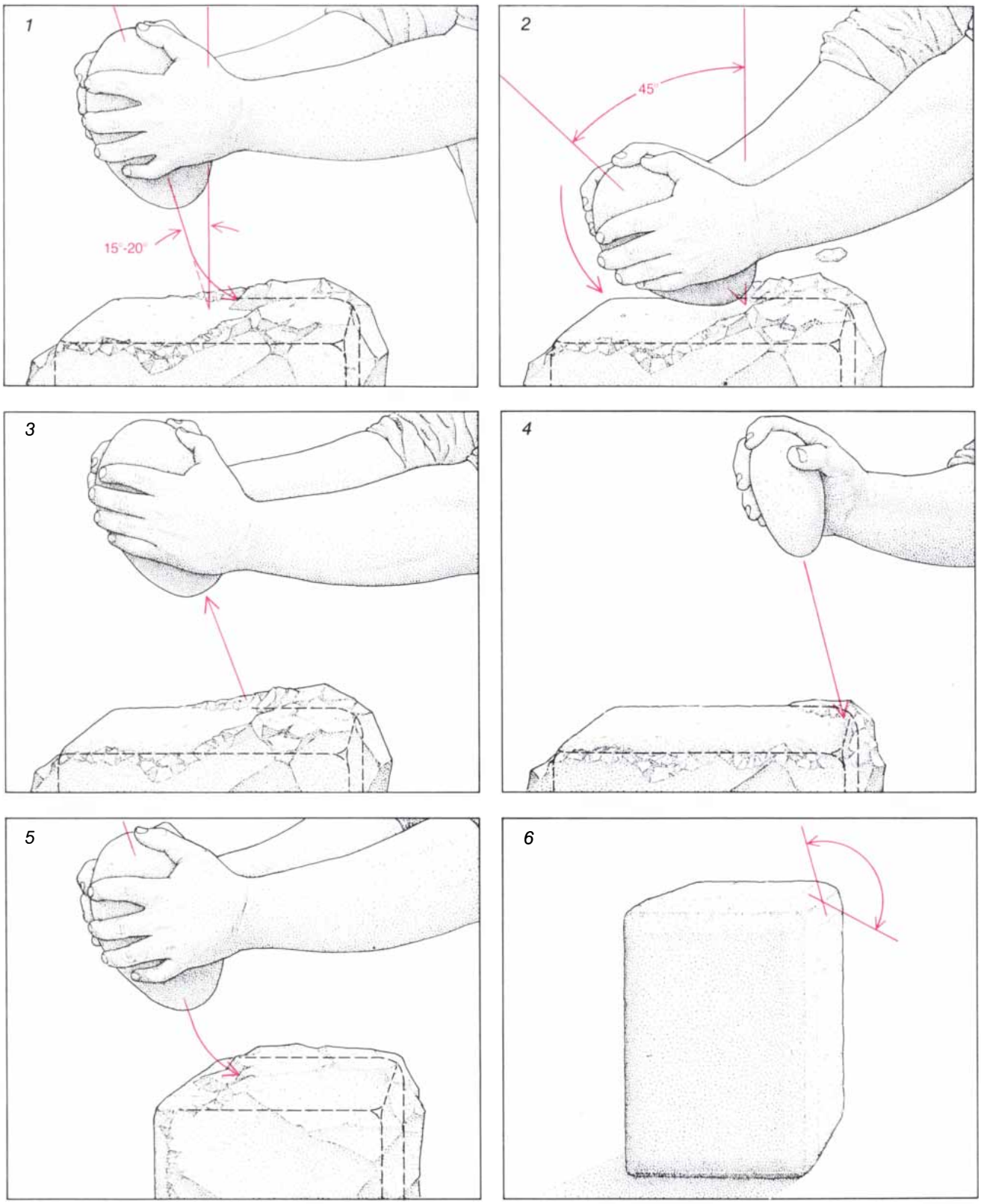
Understanding how the Incas quarried their stones is a fairly straight-



SOME INCA SITES are concentrated in the highlands of south central Peru near Cuzco. Cuzco was the capital of the Inca empire, and the technique of fitting stone blocks without mortar reached a new height there in the 15th century. The stone for many of the structures at Cuzco came from quarries at Rumiqolqa. At Ollantaytambo is an impressive Inca ruin that, like Saqsaywaman, is often referred to as a fort but was probably a religious center. Quarries at nearby Kachiqhata provided the stone for Ollantaytambo. Machu Picchu, one of the most famous and beautifully situated of all Inca ruins, sits among mountain peaks. Urubamba and Vilcanota are two names given to one river in different parts of its course.



LLAMA PIT is a well-preserved quarry site at Rumiqolqa. The Inca quarrymen probably pried stones from the rock face (*left*) with bronze pry bars or wood sticks. The building blocks were then shaped and dressed before being taken from the quarry. Scattered about the Llama Pit are 250 abandoned blocks in various stages of dressing. By examining these abandoned stones the author was able to reconstruct Inca stonecutting methods. The finished stones were carried away from the Llama Pit on ramps paved with gravel.



AUTHOR'S EXPERIMENT reveals how the Inca stonemasons might have dressed the building blocks. The author chose a block of andesite (an igneous rock) from the Llama Pit. After giving it a rough rectangular shape he took up a four-kilogram (8.8-pound) hammer and began pounding at one of the six faces (*1*). The hammer was held loosely between the hands and allowed to fall at an angle of between 15 and 20 degrees from the vertical. Just before striking the stone the hammer was given a twist with the wrists so that the angle of impact was about 45 degrees (*2*). After each strike the hammer rebounded about 25 centimeters (*3*). When the first face was finished, the block was left in the same position and a smaller hammer was employed to draft the edges of the next face to be dressed (*4*). The small hammer, which weighed 560 grams (1.2 pounds), was held tightly to deliver grazing blows directed away from the edge. Then the block was turned so that the large hammer could be used on the second face (*5*). The author's experimental technique resulted in a block with corners that are slightly convex, much like the corners observed on the Inca stone blocks (*6*).

forward matter. Understanding how the stones were cut and dressed poses greater technical difficulties. Here again the Llama Pit was helpful. The most striking feature of the Llama Pit is the 250 Inca blocks scattered about the site. In contrast to the stones from Kachiqhata, stones from Rumiqolqa were generally finished, or nearly finished, on five of their six surfaces while they were still in the quarry. One can find among the 250 blocks at the Llama Pit examples of all the stages of production, from raw stone to finely dressed blocks. By examining these stages I have been able to reconstruct the process by which the blocks were manufactured.

One of my first tasks was to identify the tools used in dressing the stones of the Llama Pit. Scattered among the chippings of andesite in the quarry pits I found stones foreign to the site in both shape and material. My search turned up enough of the foreign stones for me to be quite sure they had served as hammers for shaping the building blocks. As at Kachiqhata, most of these foreign stones are river cobbles. They appear to have come from the banks of the Vilcanota River, which flows close to the quarry. A few of the hammers are pure quartzite, others are granite and some are olivine basalt. (Basalt is an igneous rock and olivine is a mineral found in the basalt.)

The hammers and the andesite of the building stones have roughly the same hardness. One standard measure of hardness is called the Mohs scale. On the Mohs scale talc, the softest mineral, has a rating of 1 and diamond, the hardest, a rating of 10. The hammerstones I found at the Llama Pit have a rating of about 5.5, roughly the same as the hardness of the andesite in the building blocks. The hammerstones, however, are tougher than the blocks. Differential cooling during the formation of the andesite led to the accumulation of stresses in the rock. When the andesite is hit, the stresses are released and lead to a fragmentation of the rock. As a result the river cobbles make good hammers for shaping and dressing the building stones.

The Inca masons apparently employed hammers of different sizes for the various phases of the shaping process. In my search of the quarry sites I found three groups of hammers. The first group included hammers weighing from eight to 10 kilograms, the second those weighing from two to five kilograms and the third those weighing less than a kilogram. I believe each group had a specific function. The largest hammers could have

DISMANTLED INCA WALL at Ollantaytambo yields clues about how the Incas fitted stones. Each concave depression marks the place where a stone has been removed. The depressions were pounded out to precisely match the convex bottom surface of the upper stone.

DRAG MARKS on the bottom of a building block from Ollantaytambo suggest that some stones were pulled to the construction site over the gravel surface of the Inca roads. The marks can be analyzed to find the direction in which the block was dragged. For example, the circular depression (*left center*) is sharply defined on its left side and fuzzy on its right side. As the block was pulled along, gravel slipped into the depression under the front edge, which remained sharp. When the gravel reached the back of the depression, it was compressed between the roadbed and the rear edge of the recessed area, which became polished and fuzzy. Such reasoning suggests that the block shown in the photograph was dragged toward the left.

served for the rough work of breaking up and squaring off the blocks after they had been broken out of the quarry face. Most of the unfinished blocks show distinctive flaking scars similar to the scars on flaked stone tools, but much larger. The flaking scars are probably the result of pounding with the large squaring-off hammers. The medium hammers may have served for dressing the faces of the blocks and the small hammers for drafting the edges.

To find out whether the Inca masons could have employed the three groups of hammers in this way, I proceeded from observation to experiment. The raw material of my experiment was a rough block of andesite measuring about 25 by 25 by 30 centimeters. With a hammer that weighed about four kilograms I knocked off the largest protrusions to create a roughly rectangular block. Six blows were enough to do so. The next objective was to smooth one of the six faces of the rectangular block. For this purpose I chose a different four-kilogram hammer and began to pound. One might think wielding a four-kilogram hammer for an extended period would be very tiring. The work is made easier, however, by gravity. Holding the hammer lightly, one can allow it to fall onto the surface of the block while still guiding it in both hands. If the hammer is dropped onto andesite, it will rebound 15 to 25 centimeters; it can then be allowed to fall again. The process can be repeated for a long period, and the effort required is small.

Cutting stone in this fashion is essentially a matter of crushing the rock. If the hammer is directed at an angle of between 15 and 20 degrees from the normal (perpendicular) to the surface, however, tiny flakes chip off and the cutting is much accelerated. I found that the efficiency of the strike could be increased even more by giving the hammer a twist with the wrists just before it drops onto the surface of the block. Twisting the wrists increases the angle of impact to between 40 and 45 degrees from the normal [*see illustration on page 98*]. The mechanism by which the increase in angle augments the efficacy of cutting is readily explained. When the hammer is directed vertically, the entire force of the strike is converted into compression, which crushes the rock. On the other hand, if the strike deviates from the vertical, it gives rise to a shear in addition to the compression. The shear increases with the angle of the strike, and it is the shear that tears off the tiny flakes of stone and thereby accelerates the cutting.

After one of the block's six faces has been smoothed, the mason must change his technique. If the block were simply turned over and the same hammer used to cut the new face, large flakes would surely be torn from the edge of the new face by the blows of the large hammer. To avoid that result the mason must take up a smaller hammer and use it for drafting the edges of the new face before its inner part is smoothed. For this work I used a hammer weighing about 560 grams. The method is quite different from that of cutting the face. Rather than striking the surface of the block more or less vertically, the hammer grazes the edge. Gravity has little part in the work done on the edge of the block. The hammer of 560 grams is too small to be dropped and then allowed to rebound. It must be held tightly, with the force of the blow coming solely from the mason's arm.

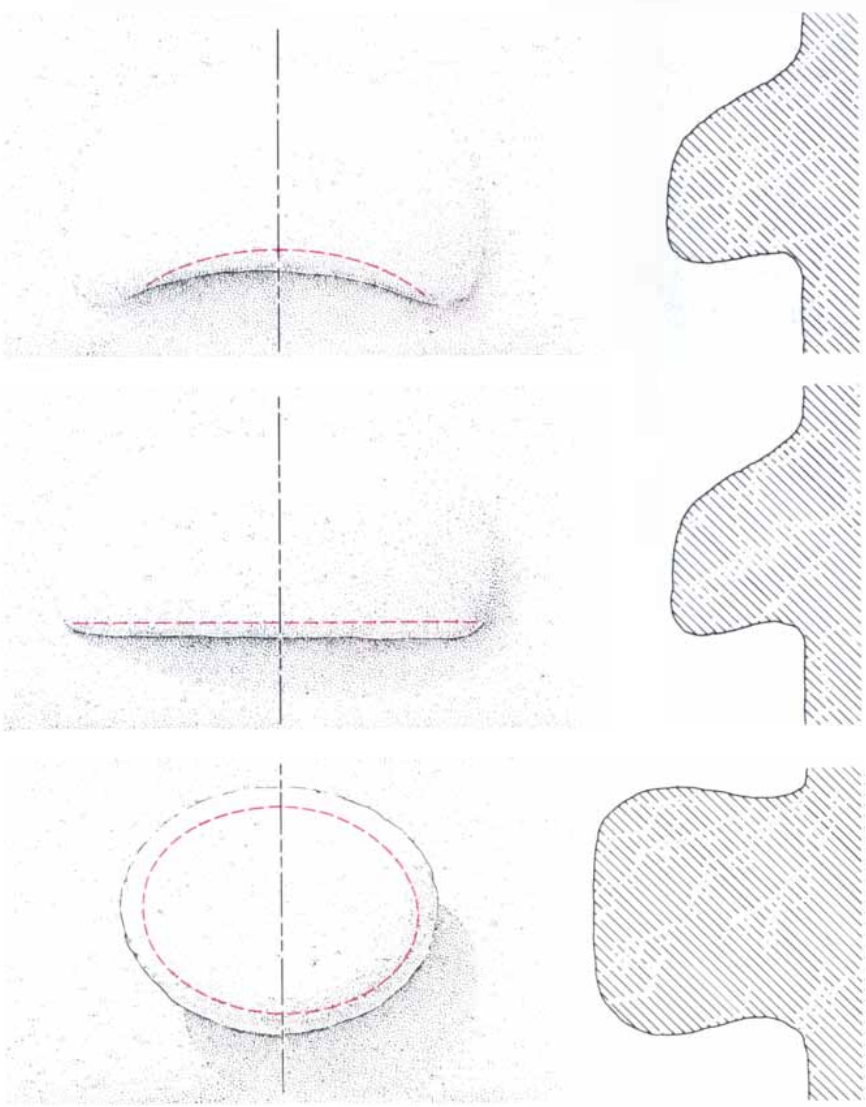
Once the edges have been drafted the block can be turned over. The small hammer is put aside, and the mason takes the heavier hammer again to dress the new face. On my experimental block I dressed two faces after the first one while trying out several more hammers that had a weight of between 3.5 and four kilograms. When I finished, I had a block that was mostly dressed. The entire process, from squaring the block to drafting five edges and finishing three sides, took no more than 90 minutes. My experiment shows that stones can be mined, cut and dressed using simple tools in a way that takes little time or effort. The next question is whether these are the methods the Incas actually employed.

The physical evidence that the Incas used techniques similar to mine is abundant. On the stones of all Inca walls, regardless of the type of rock, one finds scars resembling the scars left by my pounding on the experimental block. If the block is of limestone, there is a whitish discoloration in or around the scar. The white spots undoubtedly indicate a partial metamorphosis of the limestone resulting from the heat generated by the impact of the hammerstone. On every stone I examined the pit scars are smaller toward the edge of the stone than in the center, which suggests that the hammers used to work the edge were smaller than those used on the center of the face. Additional evidence comes from the contemporary commentator Garcilaso de la Vega, known as "the Inca." De la Vega, the son of a conquistador and an Inca princess, wrote in 1609 that the Incas "had no other tools to work the stones than some black stones... with which they dress the stone by pounding rather than cutting."

Perhaps the most intriguing questions of all concern not quarrying or dressing but the way the great stones were fitted to each other so precisely. Masonry joints are of two main types: bedding joints and lateral joints. The bedding joints are the seams through which most of the weight of a block is transmitted to the course, or row of stones, below. The lateral joints are seams between stones in the same course; little or no weight is transmitted through them. Here I shall be concerned mostly with the bedding joints.

After examining many Inca walls I concluded that when the walls were built, the bedding joints of a new course were cut into the top of the course already laid below. The stones generally had faces that were slightly convex, and the depressions that were cut to accommodate the upper stones are therefore concave. Wherever a wall has been dismantled one can clearly see the concave depressions in the remaining courses, making it appear that the removed stones have left precise impressions of their bottom surfaces [*see top illustration on opposite page*]. These concave depressions refute a hypothesis often advanced in relation to Inca masonry: that neighboring stones were ground against each other to achieve the perfect fit. It is clear that grinding two surfaces against each other cannot yield perfectly matching concave-convex joints such as the ones I observed. How, then, was the wonderful fit achieved?

As in dressing the stone, I tried the fitting myself to learn how it was



PROTUBERANCES on Inca stone blocks take various forms that probably had specific functions. One type (*top*) is well suited for the application of levers. Another type (*bottom*) may have been used for tying ropes. A third type (*middle*) could have served both functions.

The fruits of Hitachi's ongoing research (from left to right): computed tomography, lasers, robots, a computer-scanned human eye, and optical memory disks.

RESEARCH

For seven and a half decades, Hitachi has conducted research with single-minded devotion. As one of the world's foremost operators of independent research facilities, Hitachi's original technologies have made the company an innovator of electrical and electronics products. Most recently Hitachi opened its 21st R&D center, the Advanced Research Laboratory. Its purpose: To develop long-range projects for applications in the 21st century and to carry on the Hitachi tradition of pioneering research in biotechnology, material science, software science, and other technologies upon which progress depends.

The results are in use all around you

Hitachi laser research has allowed high-density recording of business data in optical disk filing systems. Complex manufacturing operations have been automated thanks to advances in robotics, creating intelligent machines with sophisticated sensors to help them see and handle work pieces. New audio/video systems make use of Hitachi's digital technology for ultrahigh-fidelity sound and high-definition television pictures.

Each year, Hitachi devotes some $1 billion to research. Our efforts to further improve advanced technology involve the talents of 16,000 R&D personnel. And our desire to share the benefits of new knowledge with others has made available over 35,000 patents for licensing worldwide.

In fact, we are constantly coming up with innovations and new applications. One case in point: A device that focuses an ion beam just 0.5 microns in diameter on semiconductor surfaces, for microscanning and etching of circuits at temperatures in excess of 3,000°C.

These are just a few examples of Hitachi research at the forefront of high technology. In 1985 alone, six Hitachi products were cited by IR-100, one of the world's top research forums, as among the previous year's 100 best technical innovations—a rare honor for a single company to achieve.

The best of worlds is yet to come

Our vision of the future includes satellite networks employing laser telecommunications and remote sensing systems to link every corner of the globe and eventually the planets. We see tomorrow's society served by energy-generating fusion reactors, 300-mph trains floating above the ground on superconductive magnetic fields, 3-D color scanners for medical diagnosis and simulation of operations—and much, much more.

We'd like you to share in the benefits of our scientific research, covering the next generation of microchips, new ceramics, information processing systems, and other high-tech innovations. For improved business efficiency. For a higher quality of life. Two goals we've pursued for 75 years as part of our commitment to a better world through electronics.

WE BELIEVE RESEARCH IS THE LEADING EDGE OF PROGRESS

done. The experiment entailed the block of andesite from the dressing experiment and a larger block into which the bedding joint was to be cut. I started by putting the smaller block on the larger one and tracing its outline. I removed the smaller block and, using the outline as a guide, pounded out a depression that matched the overall shape of the bottom of the smaller stone. The pounding produced much dust, which had to be whisked away. The dust is annoying because it dampens the hammer blows, but it is also quite useful. When the upper block is put in place again, it leaves an impression of its lower surface in the dust. Where the fit is tight the dust is compressed and where the fit is loose it is not. After the stone is removed again one pounds away at the places where the fit is tight, which are indicated by the compressed areas. By repeating the process one can achieve as close a fit as one wants.

The same technique can be applied to form the lateral joints. The new block to be added to the course is laid against the blocks already in place, and concave depressions are cut out of the blocks in place. The lateral joints differ from the bedding joints in that the close fit observed from the front of the wall is sometimes only a few centimeters deep and the interior of the joint is filled with rubble. In many instances, however, the lateral joints are fitted with the same care as the bedding joints over the entire plane of joining.

It appears that the Inca technique of fitting the blocks together was based largely on trial and error. It is a laborious method, particularly if one considers the size of some of the huge stones at Saqsaywaman or Ollantaytambo. What should be kept in mind, however, is that time and labor power were probably of little concern to the Incas, who did not have a European notion of time and had plenty of tribute labor from conquered peoples at their disposal. Furthermore, my experiments show that with a little practice one develops a keen eye for matching surfaces, so that the time needed for constructing a joint is greatly reduced. In favor of my method it should be emphasized that it works and that it does not postulate any tools other than those for which there is evidence. Moreover, it has the support of at least two 16th-century writers. One of them, Jose de Acosta, a Jesuit priest who traveled with the Spanish conquerors and is considered a highly reliable observer, wrote in 1589: "All this was done with much manpower and much suffering in the work, for to fit one stone to the other, until they were adjusted, it was necessary to try the fit many times."

I think my experiments provide a reasonable account of how the Inca masons quarried their stones, shaped them and fitted them together. How the stones were transported to the building site and how they were handled at the site, however, are questions that have not yet yielded completely to investigation.

In the handling of the stones a variety of protuberances carved on the face of the block undoubtedly had a significant role. The protuberances come in several sizes and shapes [*see illustration on page 101*]. Generally they are found on the lower part of a block that has been set in place. The projections may have served as points to which ropes could be attached or to which the force of a lever could be applied. The projections were apparently cut only at the building site and served specifically for the purpose of handling stones there. Since none of the blocks abandoned along the transport routes have protuberances, it would seem that the projections did not have a role in bringing the blocks to the building site.

How were the blocks transported? Some preliminary evidence comes from blocks strewn about at Ollantaytambo. On these blocks one can ob-

serve a peculiar polish marked by more or less parallel longitudinal striations. Both the polish and the striations seem to be the result of dragging the blocks from the quarries to the construction site. The direction in which the block was dragged can readily be determined from the marks. If the surface is inspected closely, one finds irregularly shaped areas that have not been polished because they are slightly recessed. These regions generally have a sharp boundary on one side and a fuzzy, gradual boundary on the other. When the stone was being dragged, the sharp edge was at the front and the diffuse edge was at the rear. Gravel from the roadbed would have accumulated at the back of the depression and been ground between the block and the road, yielding the smoothed area at the trailing edge.

Other evidence from the blocks helps to fill out our picture of the transport process a bit. The polish is found only on the broadest of the block's faces, suggesting that the stones were dragged in their stablest position. The blocks in the quarry show no polish and the extent of the polished surface increases with distance from the quarry. The presence of the polish tends to refute the suggestion that the Incas moved the larger stones on rollers or skids. The presence of drag marks does not exclude the possibility that rollers or skids were employed on the uphill parts of the ramps, but no material evidence of such implements has been found.

If the blocks were dragged along the ramps, the Incas must have devoted considerable labor power to the task, particularly for the largest stones. The force required to drag any block depends on the coefficient of friction between the stone and the material of the ramp, the slope of the ramp and the weight of the block. I determined the coefficient of friction experimentally and measured the slope of the ramp at Ollantaytambo as being about 10 degrees. The largest block at Ollantaytambo weighs about 140,000 kilograms. I have calculated that it would take a force of some 120,400 kilograms to pull such a block up the ramp. If a man can pull consistently with a force of 50 kilograms (which may be an overestimate), it would have taken some 2,400 men to get the block to the top of the ramp. That figure agrees in order of magnitude with the account of the 16th-century writer Cieza de Leon, who observed that of the 20,000 men assigned to the construction of Saqsaywaman, 6,000 were delegated to the transport detail.

The foregoing account seems reasonable, yet it raises significant questions that I have not been able to answer so far. The Inca ramps were only from six to eight meters wide, and I have not been able to propose plausible solutions for two problems posed by this narrowness. One is how 2,000 men or more could have been harnessed to the block so that each was contributing to the pull. The other is how the crowd of workers was arranged on the cramped road. These are only two of the unsolved problems concerning the transport of the blocks. Among the others are the techniques for tying the ropes to the blocks and the methods for maneuvering the huge stones.

Moreover, the stones from Rumiqolqa were probably not dragged at all. Unlike the blocks from Kachiqhata, those from Rumiqolqa were finely dressed before they left the quarry. No drag marks are found on them, and it seems unreasonable to think that a finely dressed face would be dragged on a stone ramp. How then were the dressed blocks transported? This question and many others remain to be answered before the final account of Inca stonemasonry can be written. Yet by experiment and observation some of the fundamental questions about the quarrying, dressing and fitting of the stones have now received answers.

William Herschel and the Making of Modern Astronomy

He discovered thousands of stars and nebulas through telescopes that he himself built. His observations and theories expanded the bounds of astronomy to include the study of objects beyond the solar system

by Michael Hoskin

A typical astronomy textbook of the mid-18th century has chapters on such topics as time and celestial coordinates, and pages devoted to descriptions of the sun, moon and planets and their orbits. There is virtually nothing, however, on stellar clusters, nebulas or the large-scale structure of the universe. In contrast, textbooks published a century later deal with such subjects as a recognized part of astronomy.

The expansion in astronomical knowledge was largely brought about by the achievements of one man: William Herschel. His skill as a craftsman advanced the art of telescope construction; his dedication as an observer yielded unprecedentedly comprehensive catalogues of nebulas and stars, and his boldness as a theorist provoked the scientific study of galactic evolution. The accomplishments of this one man compelled astronomers of the 19th century to widen the compass of their field to include active study of celestial bodies outside the solar system and indeed outside our galaxy.

Yet astronomy was not Herschel's primary career. He was trained as a musician, becoming sufficiently accomplished on the oboe to join a regimental band in his native city, Hanover. In 1757, the year Herschel turned 19, the French occupied Hanover and he fled to England. He maintained himself in his new homeland first by copying music and then as a performer, conductor and composer. (He also anglicized his given name, Friedrich Wilhelm, to William.) In 1766 he was appointed organist at the fashionable Octagon Chapel in the city of Bath. It was a secure position, and although Herschel had a variety of musical duties, he was at last able to indulge his awakening intellectual interests.

Having enjoyed a monograph on the mathematical theory of harmony by Robert Smith, a professor of astronomy at the University of Cambridge, Herschel turned to Smith's previous work, a popular textbook on practical optics published in 1738. The book, *A Compleat System of Opticks,* exposed Herschel to the art of telescope making; it also contained descriptions of what could be seen in the heavens with the aid of such instruments. The practical aspect of astronomy fascinated Herschel, and he decided to try his own hand at building telescopes.

Herschel began by constructing refracting telescopes from lenses and tubes of assorted lengths, but the difficulty of handling long refractors led him to turn to reflecting telescopes, which had the added advantage that their apertures could be made wider than those of refractors. Because aperture size determines how much light a telescope can collect, faint (and distant) objects can be better examined with reflectors.

Herschel had a neighbor who made a hobby of grinding and polishing mirrors for telescopes, and he arranged to buy the neighbor's stock of tools, equipment and unfinished mirrors. Guided by Smith's book on optics, Herschel taught himself the art of finishing mirrors of speculum metal (an alloy of copper and tin) by trial and error. By the fall of 1773 he had begun to mount his own telescopic mirrors. He soon became proficient, and by January of the following year he had made a reflector with a 5½-foot focal length, a very respectable size by contemporary standards.

As the scale of Herschel's instruments increased, his ambitions grew, but the large disks he wanted to grind into concave mirrors for the telescopes he envisioned could not be cast by local foundries. Not one to give up easily, Herschel converted the basement of his house into a foundry. While he was perfecting the technique of manufacturing progressively larger mirrors, he was also making progressively better eyepieces, some of them capable of hundredfold and even thousandfold magnification. (Indeed, many of Herschel's contemporaries refused to believe his telescopes were capable of such magnifications until one of his instruments was put through a side-by-side comparison with the best instrument the Royal Greenwich Observatory had at its disposal.)

In 1776 he completed his first telescope of 20-foot focal length, which had a primary mirror 12 inches in diameter. The telescope tube was crudely slung from a pole and had to be manhandled to face in roughly the right direction. The observer peered through the eyepiece at the top of the tube while perching precariously on a ladder. Such an awkward arrangement prompted Herschel to design new telescope mountings. Two years later he had perfected a stand for his smaller telescopes that enabled him to control the motions of the telescope mechanically, by means of a system of pulleys, hinges, grooves and gears, without looking away from the eyepiece.

Herschel later incorporated many of the same mounting features into a large stand for his second 20-foot reflector, which had a mirror 18 inches in diameter. The observer stood safely on a platform with the fine-focusing and positioning controls at hand. When necessary, the entire structure could be turned around by a single workman.

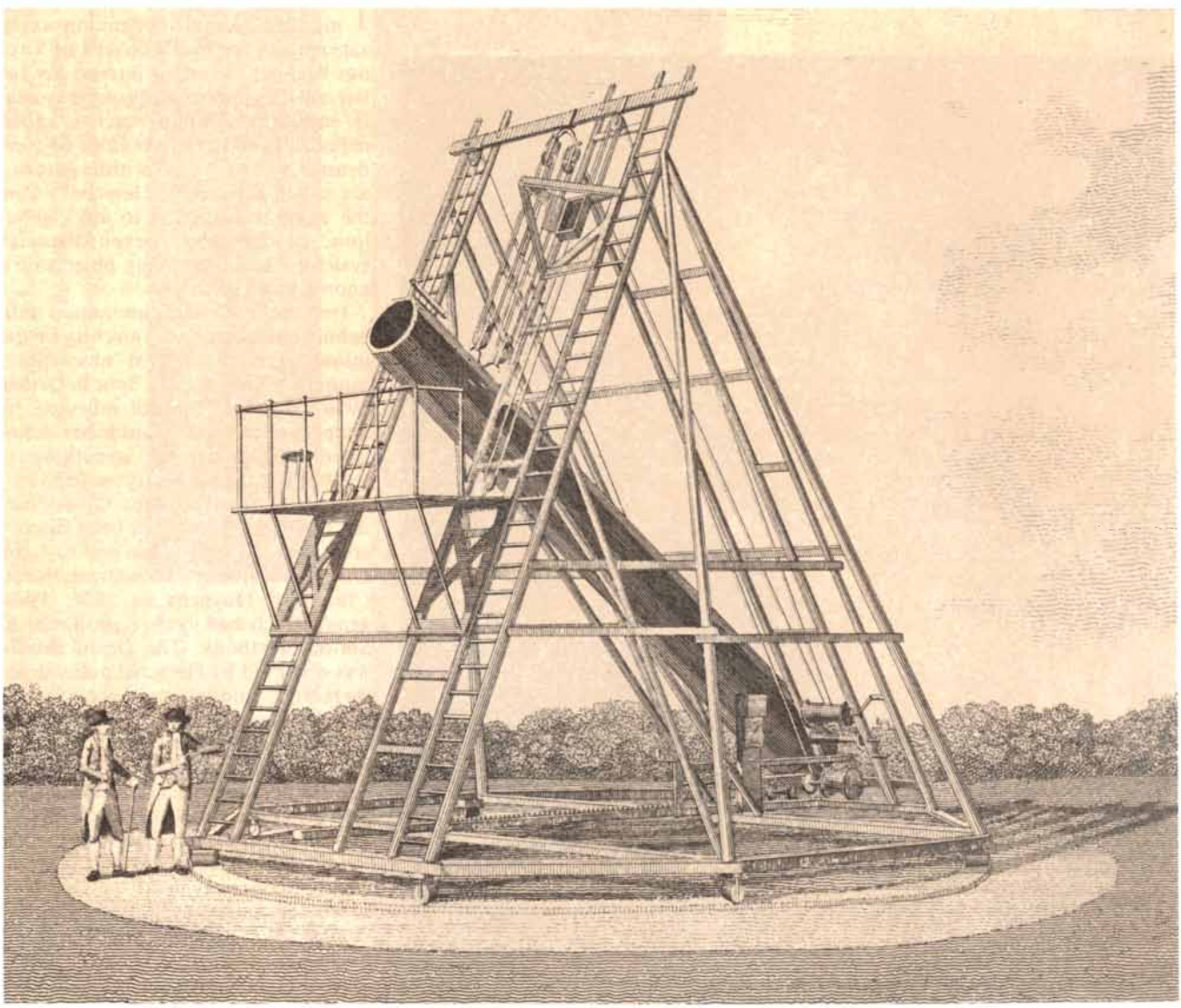
This instrument, which he called the "large" 20-foot reflector, was the mod-

el for Herschel's greatest achievement as telescope maker: a 40-foot telescope of four-foot aperture. Although it was one of the technical wonders of the 18th century, it was too cumbersome to use for frequent observation and proved to be an unsuccessful scientific instrument.

Although Herschel did sell some of his telescopes to supplement his income, the primary motivation for constructing them was a desire to personally observe objects beyond the reach of conventional astronomical instruments. In spite of his demanding job as organist and his time-consuming avocation as telescope builder, Herschel devoted countless hours to familiarizing himself with all the celestial objects his unmatched telescopes brought into view.

He assumed the role of a natural historian, examining and noting the position of every specimen he came across that was brighter than a given magnitude. Like a naturalist categorizing thousands of animal or plant species, Herschel soon had to confront the problem of cataloguing the objects he had identified. Never before in the history of astronomy had anyone observed as many celestial objects as Herschel did through his powerful telescopes, let alone attempted to sort and classify them.

Herschel tackled the task with the same thoroughness and determination he had shown in building telescopes. For example, he discovered and recorded some 848 double stars (pairs of stars whose angular separation is small), with which he hoped to gauge stellar distances by means of parallax measurements. (When he reexamined some of them later, he found instances where the stars had moved around each other. This was the first direct evidence that attractive forces operate outside the solar system, as Newton had assumed but not proved.) Herschel also compiled catalogues of "the comparative brightness of stars," listing the stars in the order of decreasing apparent brightness so precisely that in future years even a slight variation in the luminosity of any star would



"LARGE" 20-FOOT REFLECTOR, completed by Herschel in 1783, had a primary mirror 18 inches in diameter at its base. The most significant advance represented by the telescope was its unique mounting. The observer could stand on the platform at the mouth of the telescope, regardless of the telescope's declination, and have the fine-adjustment controls within arm's reach. The entire structure could be turned around by a single workman. Although the telescope was originally designed as a Newtonian reflector, with a small, flat mirror reflecting the image sideways into an eyepiece, Herschel could not accept the consequent loss of light transmission. He eliminated the small mirror and instead peered directly down the tube through an eyepiece fixed to the inside rim.

manifest itself by throwing the sequence out of order.

Herschel made a number of discoveries within the solar system as well. One of them did not require a telescope at all: he effectively discovered infrared rays while recording the temperature indicated by a thermometer exposed to each color of the sun's spectrum. He noted that he in fact got the highest temperature reading just beyond the red.

Another serendipitous observation won Herschel world fame and led to his liberation from musical chores. On the evening of March 13, 1781, he was engaged in his latest and most thorough survey of the entire visible sky when he encountered an object he instantly recognized as being no ordinary star: it was not a point of light but a shining disk, whose apparent size increased in proportion to the telescopic power he applied. When he examined it again a few days later, he found it had moved; it was a member of the solar system, and presumably a comet in spite of the absence of a cometary tail. Because Herschel lacked the expertise to define the object's position accurately when he announced his discovery, professional astronomers (who were less skilled as observers and whose instruments were inferior) were exasperated by the difficulty of finding the supposed comet. When it eventually was located and its orbit was determined, it proved to be a new planet, the one we know as Uranus and the first to be discovered since antiquity.

ENGRAVED PORTRAIT OF HERSCHEL depicts him holding a sketch of the planet Uranus and two of its satellites, all of which he discovered. Herschel originally named the planet Georgium Sidus (Georgian Star) in honor of the ruling British monarch, George III. Although the discovery of Uranus won Herschel worldwide fame and secured him a royal stipend, his main astronomical interest lay in star clusters and nebulas and their relation.

Herschel became an international celebrity overnight and was thereupon elected a fellow of the Royal Society of London. After judicious lobbying in court circles he was awarded a royal pension by King George III, his only duties being to live near Windsor Castle and to be on call as the royal family's resident astronomer. In appreciation Herschel dubbed the planet Georgium Sidus (Georgian Star) and referred to it thereafter by that name. (In France many astronomers continued to call the planet Herschel in honor of the discoverer until the middle of the 19th century, when the name Uranus, suggested by the contemporaneous German astronomer Johann Elert Bode, finally prevailed.)

It is somewhat ironic that Herschel attained formal recognition as an astronomer for his discovery of Uranus because his major interest lay far beyond the confines of the solar system in mysterious milky patches called nebulas. (Today only regions of low-density gas and dust within galaxies are called nebulas. In Herschel's time the word was applied to any "nebulous," or fuzzy, object beyond the solar system; it included many objects now known to be galaxies.)

Herschel's lifelong fascination with nebulas is attested to by an entry on the initial page of his first observation journal: "Saw the lucid Spot in Orions Sword, thro' a $5\frac{1}{2}$ foot reflector; its Shape was not as D^r Smith has delineated in his Optics; tho' something resembling it; being nearly as follows." The "lucid Spot" was the Great Nebula in Orion, which had been discovered and roughly sketched by the Dutch astronomer and mathematician Christiaan Huygens in 1656; Huygens' sketch had been reproduced in Smith's textbook. The Orion nebula was observed by Herschel many times thereafter, and on one occasion he noted: "There is a visible alteration in the figure of the lucid part."

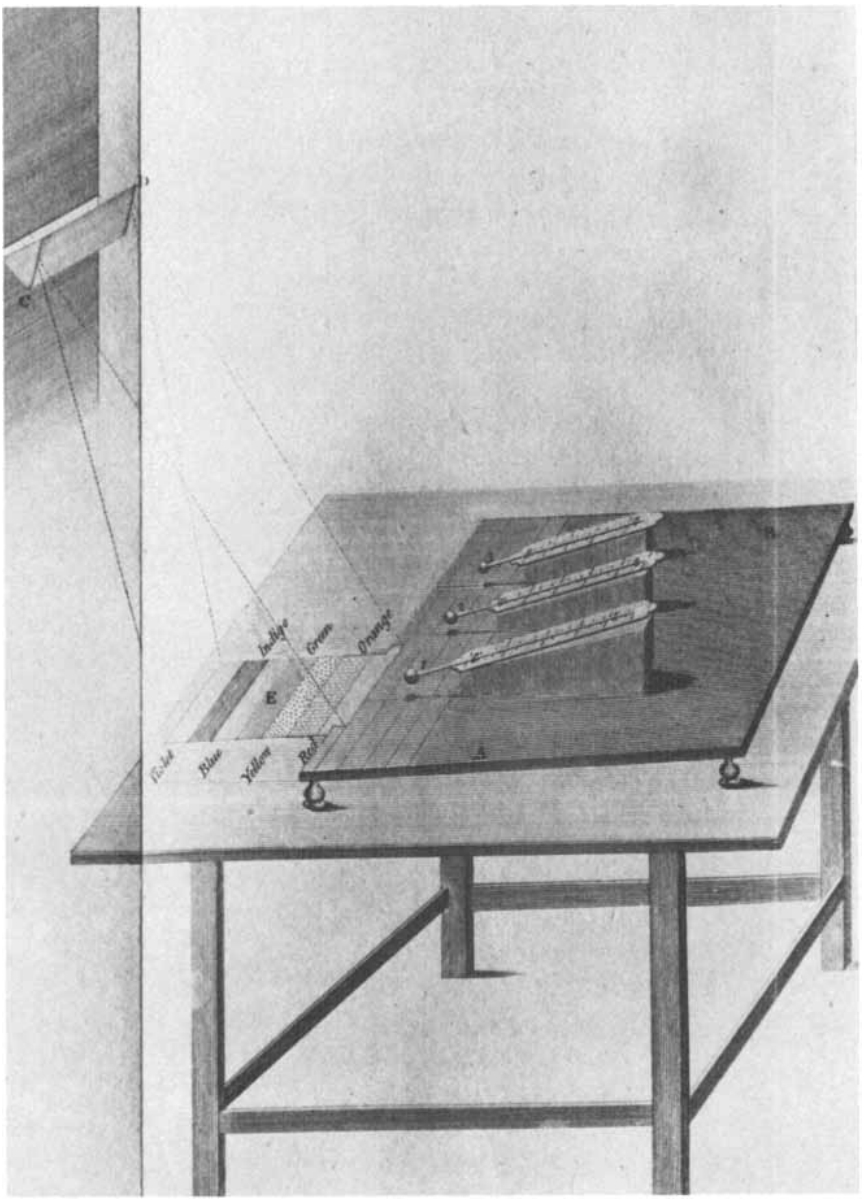
Stargazers of the early 18th century had come across a number of these mysterious objects. Preeminent among the astronomers was Charles Messier, the French comet hunter, who regarded the permanent milky patches as a source of confusion in his sweeps for comets; in 1780 he compiled a list of 68 such objects so that other astronomers would not mistake them for comets. Herschel, who found nebulas to be worthy of study in their own right, acquired Messier's list from a friend and proceeded to investigate the objects enumerated. His telescopes revealed that many of them were simply clusters of stars, but some seemed

to be truly nebulous and of a different physical nature.

It was while he was searching for other such objects that Herschel made his first major discovery in stellar astronomy: "A curious Nebula, or what else to call it I do not know. It is of a shape somewhat oval, nearly circular, and with this power [a magnification of 460 times] appears to be about 10 or 15 [seconds of arc] in diameter.... The brightness in all the powers does not differ so much as if it were of a planetary nature, but seems to be of the starry kind." Over the course of the years Herschel found several more objects of that type. He called them planetary nebulas, a term astronomers still use today. Those mysterious glowing disks were to puzzle Herschel, and astronomers from abroad who made the pilgrimage to his home were often shown a planetary nebula and asked to give their opinion of its nature.

Herschel decided that to advance the study of nebulas further he would have to examine considerably more specimens; his large 20-foot telescope would be ideal for the purpose. He set himself to search the entire sky as visible from England, recording the position and description of as many nebulas as the reflector's 18-inch mirror could reach. For the next 20 years he would spend nights in the cold and damp by the Thames River, sweeping the sky strip by strip and shouting out positions and descriptions of nebulas to his sister Caroline, his devoted assistant throughout his career in astronomy. It was one of the most heroic campaigns in the history of observational astronomy, and it resulted in two catalogues of 1,000 nebulas each and a third one listing 500.

After Herschel's death his son extended the effort into the southern skies. The combined Herschel catalogues formed the "Catalogue of Nebulae and Clusters of Stars," ultimately published in 1864; they were later enlarged by J. L. E. Dreyer into the New General Catalogue (NGC), which is still commonly referred to by astronomers today.

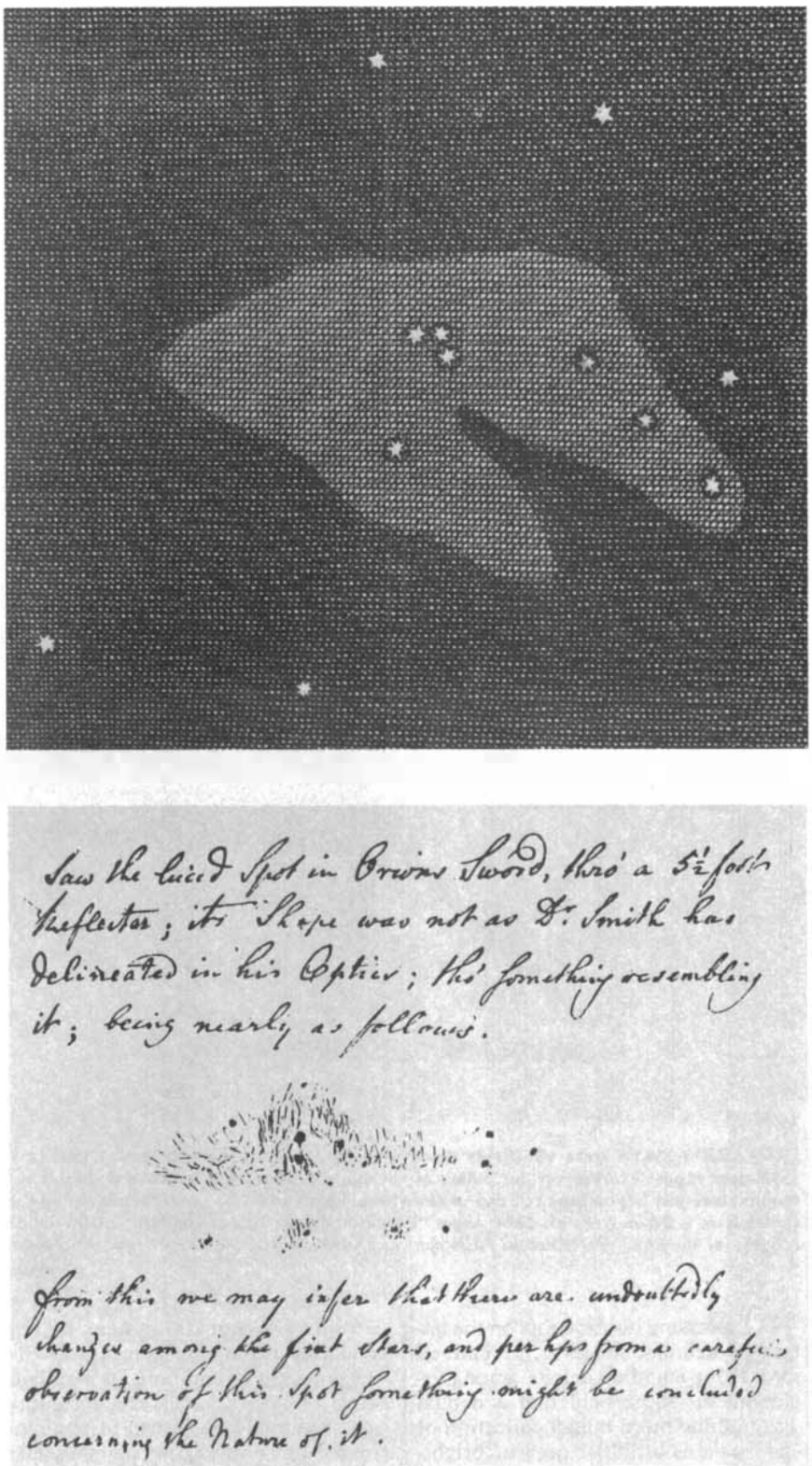


INFRARED RAYS were effectively discovered by Herschel when he noted that thermometers exposed to the various colors of the sun's light actually registered the highest temperature just beyond the red end of the visible spectrum. The experimental apparatus shown here is taken from his 1800 paper "Experiments on the refrangibility of the invisible rays of the Sun," published in *Philosophical Transactions of the Royal Society of London*.

Herschel's work in nebular astronomy extended beyond observation and cataloguing: he theorized freely on the nature of the puzzling objects and their significance in the scheme of the universe. Indeed, the yearning to explain what he saw through his telescopes was what had driven him to observe and catalogue nebulas in the first place. As Herschel himself put it, "a knowledge of the construction of the heavens has always been the ultimate object of my observations."

In expressing his thoughts on the nature of nebulas, however, he entered into a long-standing debate among astronomers. Some held that a nebula was nothing more than a collection of stars seen as a diffuse area of brightness because of its distance from the earth, much like the Milky Way. Others believed nebulas consisted of what was referred to as nebulosity, a luminous fluid that differed from the substance of stars. Because Herschel thought he had perceived changes in the shape of the Great Nebula in Orion, he maintained that it could not be composed of stars. If a nebula were an extremely distant star system, he reasoned, even the smallest angular displacement of its components would involve movement over vast distances, and stars could not travel fast enough to account for the Orion nebula's variation in shape. On the other hand, Herschel's telescopes had shown that many of Messier's "nebulas" were star clusters. How then could one distinguish between the two appearances of nebulosity, the real and the illusory?

Herschel thought he could tell if a nebula was composed of nebulosity or of stars according to its aspect. Some nebulas had a smooth, "milky"



ORION NEBULA (M42) was depicted in Robert Smith's *A Compleat System of Opticks* (*top*) and was drawn by Herschel (*bottom*) on March 4, 1774, in his observing journal. Although Smith's book was published in 1738, the picture of the nebula was actually reproduced from a sketch done by the Dutch astronomer and mathematician Christiaan Huygens in 1656. Herschel was observant enough to note right away that "its Shape was not as D. Smith has delineated in his Optics." In the course of further examination over several years Herschel thought he detected changes in the nebula's shape. He reasoned that the Orion nebula, unlike others he observed through his instruments, could not be an extremely distant cluster of stars but must instead be composed of "nebulosity," or luminous fluid.

appearance; he believed they were probably true nebulas. Others had a mottled appearance; he believed they were star clusters that could be resolved into their constituent stars by a sufficiently powerful telescope. These insights were the basis of his first major theoretical paper to the Royal Society, "Account of some observations tending to investigate the construction of the heavens," which he read to the society on June 17, 1784.

Just five days later Herschel trained a telescope on M17 (object number 17 in Messier's list), also known as the Omega Nebula. To his consternation he found that it seemed to contain both kinds of nebulosity, "milky" and "resolvable." "It is not of equal brightness throughout, and has one or more places, where the milky nebulosity seems to degenerate into the resolvable kind.... Should this be confirmed on a very fine night, it would bring on the step between these two nebulosities which is at present wanting, and would lead us to surmize that this nebula is a stupendous Stratum of immensely distant fixed stars some of whose branches are near enough to us to be visible as resolvable nebulosity, while the rest runs on to so great a distance as only to appear under the milky form."

Herschel now began to suspect that the difference between milky and resolvable nebulosity was not physical but rather was due to distance. He considered his suspicions confirmed by the configuration of M27, the Dumbbell Nebula. Stars could be made out along its axis and at the center of its two outer globular regions, but the entire object was enveloped by a diffuse aura. Herschel explained the curious structure as arising purely from distance effects: the nebula was a huge comet-shaped conglomeration of stars, its nucleus pointing toward the earth and its tail flaring behind it into space. The stars that were distinguishable were those in the nucleus; the others constituted the unresolvable aura [*see illustration on opposite page*].

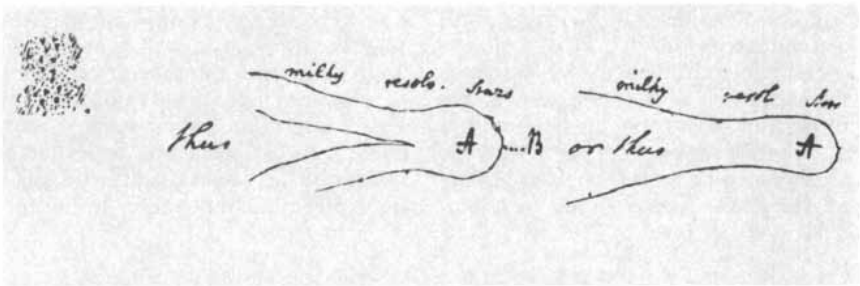
In spite of the evidence to the contrary that he himself had noted in the Orion nebula, Herschel's confidence in the existence of a luminous fluid visible as milky nebulosity quickly began to wane. His new opinion was bolstered by his observation that the area of sky surrounding a nebula was generally devoid of stars. "It appeared to me remarkable that in and about the place where the many Nebulae began there was an uncommon scarcity of stars so that many fields were totally without a single star," he wrote in his observing book. "If these Nebulae

should be clusters of stars it should seem as if they were collected together from the neighbouring spaces [presumably as a result of gravity's attractive force]."

Herschel theorized that all nebulas (except perhaps the planetary nebulas, which appeared to be exceptionally uniform in brightness) were composed of stars, gathered together over long periods of time by the force of gravity. To present this theory convincingly in a second paper to the Royal Society he simply suppressed all mention of the changes he thought he had detected in the Orion nebula, for these were incompatible with his new cosmogony. Instead the Orion nebula was now presented as a star system, but one so distant that it could not be resolved into stars even by Herschel's telescopes. Since it was so distant, and yet was spread across such a wide region of sky, it must be a star system of enormous extent, one that might well "outvie our Milky Way in grandeur."

This proposition left many questions unanswered. What were the planetary nebulas, and how did they fit into the cosmogony? What was the ultimate fate of star systems condensing under gravity? Perhaps, Herschel suggested, the planetary nebulas are star systems in the final stages of gravitational collapse, and so "the stars forming these extraordinary nebulae, by some decay or waste of nature, being no longer fit for their former purposes... may rush at last together, and, either in succession, or by one general tremendous shock, unite into a new body." The final implosion, he maintained, could explain the new "star" recorded in 1572 by the Danish astronomer Tycho Brahe. Herschel's suggestion underscores his awareness that the force of gravity could do more than explain the stable orbits of planets and their satellites. He realized that the evolution of nebulas also had to be considered in the context of an all-pervading attractive force.

Herschel's picture of the dynamic processes that shape star clusters was incomplete, however. How were stars formed in the first place, and what happened to the matter involved in the gravitational collapse? We can be sure that Herschel was not entirely satisfied with a cosmogony in which all nebulas were simply star clusters. A solution to his difficulty manifested itself on November 13, 1790. That evening, in the course of his routine sweeps for nebulas, Herschel came on "a most singular phenomenon! A star of about the 8th magnitude, with a faint luminous atmosphere." The object was the planetary nebula NGC 1514, which has an unusually prominent central star. (Because it was not a faint disk of uniform light, Herschel in fact classified it not as a planetary nebula but as a "nebulous star.") How was he to explain it? The answer was suddenly clear to him: it was a star condensing out of a cloud of luminous fluid under the action of gravity. The revelation meant he would have to accept the existence of nebulosity and retract his theory that all nebulas are star clusters disguised by distance, but at least he could incorporate planetary nebulas into the revised cosmogony.



DUMBBELL NEBULA (M27) (*left*) and two alternative explanations of its appearance (*right*) were drawn by Herschel in his journal in 1784. He hypothesized that a huge, distant, comet-shaped star cluster (*A*) would look much like the Dumbbell Nebula to an observer (*B*) in front of it: the nearest stars would be seen individually, those farther away would be seen as "resolvable" nebulosity and those farthest away would be seen as "milky" nebulosity. Herschel had previously believed that the milky nebulosity was a luminous fluid quite unlike stars; M27 convinced him that this kind of nebulosity was merely the result of distance.

Herschel now believed the stellar evolutionary cycle began with a cloud of thinly scattered nebulosity that gradually breaks up into a number of smaller and denser clumps under the action of gravity. In the process of condensing, these clumps become first amorphous nebulas and then planetary nebulas. Most of the luminous substance constituting the nebulas eventually coalesces into stars, although some of it is dissipated. Such newly formed stars then congregate as a result of mutual gravitational attraction, evolving from a loose cluster of stars into a tightly packed globular cluster. The last stage of the cycle is reached when the cluster collapses cataclysmically. The luminous material that is dispersed throughout the universe from such collapses, together with the luminosity that is constantly given off by stars and nebulas, will here and there collect into clouds of nebulosity, and so the cycle can begin anew.

As the pieces that made up the jigsaw puzzle of the universe slowly began to fall into place, Herschel even included the creation of planets in his cosmogony. He theorized that small clouds of nebulosity would sometimes be attracted toward a star and assume the form of a comet. Every time the comet passed the star some of its material would fall into the star and replenish it, while the heat of the star in turn would help to congeal the material in the comet. After many such passes the comet would be transformed into a planet. In this way Herschel enhanced the unity of his universe by having stars and planets arise jointly out of the same primordial matter.

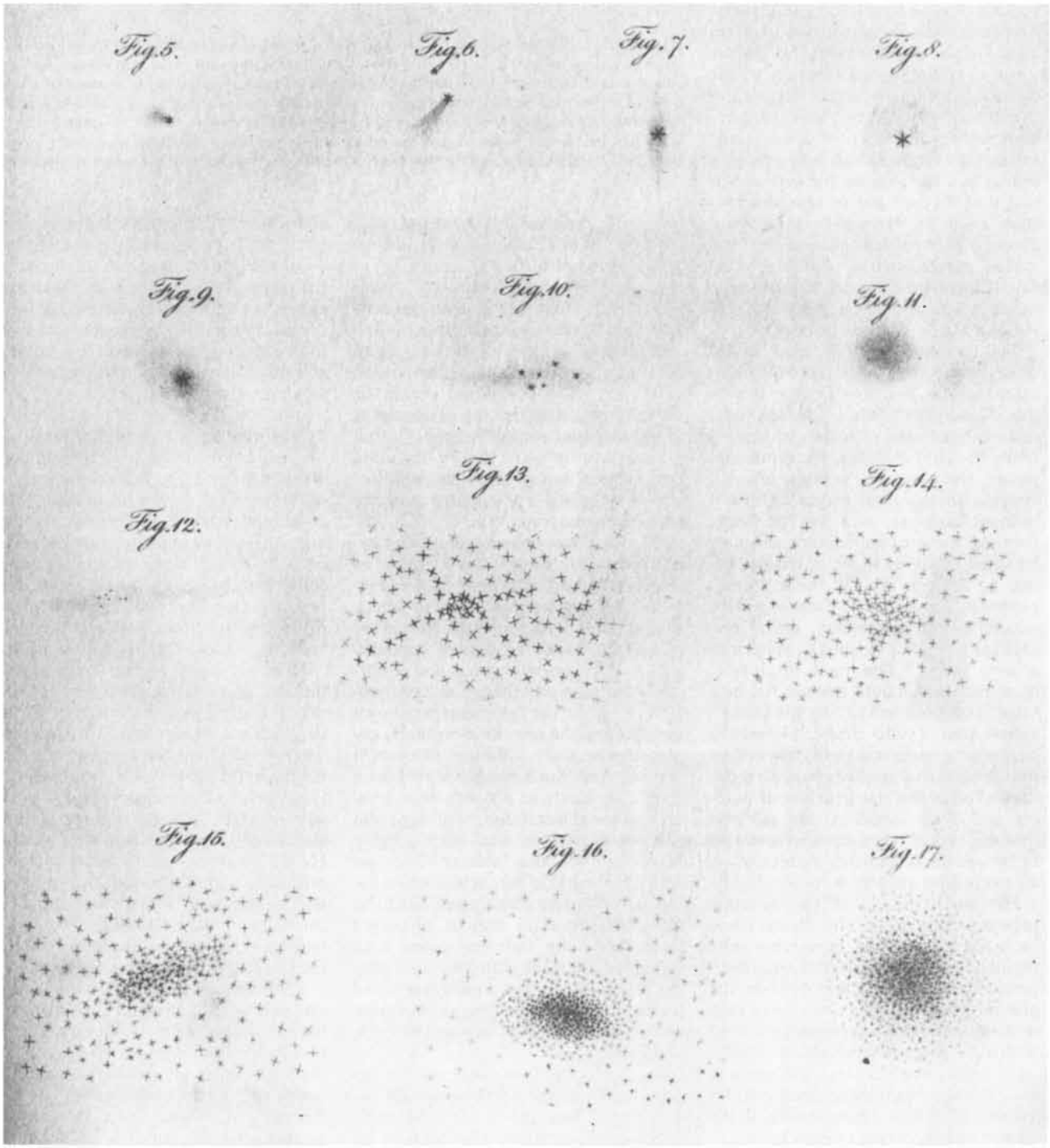
The breadth and novelty of his theories, coupled with his naive style of writing, kept Herschel at the center of controversy during his lifetime. His contemporaries in the Royal Society argued over whether he was a genius or a charlatan. Some of the society's fellows were openly hostile, perhaps because Herschel, unlike most great observers, speculated on what he saw. Indeed, he took it as his duty to do so.

Herschel's methods and many of his theoretical investigations were ingenious. For example, he showed how statistics could be applied in astronomy when he counted the number of stars visible to the observer in various directions and, by assuming a constant density of stars, plotted a three-dimensional outline of our local star system. He was also the first to find a systematic pattern in the "proper," or individual, motions of stars; he explained this as being the result of the solar system's motion through space in the direction of the constellation Hercules.

Although many of Herschel's conclusions were legitimately criticized by his contemporaries, no one could deny his success in building huge telescopes, the importance of his discoveries both inside and outside the solar system and the value of his immense catalogues of double stars and nebulas. By virtue of these nontheoretical achievements he came to enjoy almost automatic right of publication in *Philosophical Transactions of the Royal Society of London.* Through this medium his theories of

"the construction of the heavens" were disseminated.

His papers broached such topics as the evolution of stars and planets, the formation of galaxies and the nature of nebulas, which were all to become an accepted part of astronomy. Many of Herschel's speculations foreshadowed modern cosmological theories: they certainly displayed a perceptive appreciation of the critical role gravity plays in the workings of the universe. His most lasting legacy, however, is to be found not in particular results of his galactic and extragalactic explorations but rather in the fact that he dared to explore these regions of the cosmos in the first place. Herschel thereby ensured that the intellectual horizons of every subsequent generation of astronomers would extend far beyond the solar system that had monopolized the attention of astronomers until his time.



CERTAIN STARS, NEBULAS AND CLUSTERS were selected by Herschel from among the thousands he had catalogued to illustrate different stages in his theory of stellar evolution. The drawings shown here accompanied a paper by Herschel read to the Royal Society of London in 1814. The paper presents his theory that individual stars arise out of congealing accumulations of nebulosity and "grow" by absorbing any nebulosity that is drawn to them. The stars thus formed are then gathered into clusters by the mutually attractive force of gravity. The final stage is represented by a dense globular cluster (*bottom right*), which collapses into itself.

THE AMATEUR SCIENTIST

A homemade device for testing particle scattering; experiments in zero gravity

by Jearl Walker

In 1909 Ernest Rutherford suggested an experiment that led to the first modern model of the atom. Hans Wilhelm Geiger, a German postdoctoral researcher working in Rutherford's laboratory, had been studying the passage of alpha particles through a metal foil. Most of the positively charged particles passed through the foil directly, striking a detector screen on the other side. A few particles were deflected slightly by electrical forces as they went through atoms in the foil. The deflections never amounted to more than a few degrees because the particles carried appreciable energy and momentum.

Nevertheless, Rutherford suggested to Geiger that Ernest Marsden, then a student working with Geiger, look hard for alpha particles deflecting at greater angles. No one, including Rutherford, expected that any would be found. Within days, however, Geiger returned to Rutherford with the startling news that some of the particles were scattered at large angles, even directly back toward the source. Rutherford later commented: "It was quite the most incredible event that has ever happened to me. It was almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you."

The deflection of alpha particles in that experiment is now called Rutherford scattering. It revealed that the contemporary model of the atom was wrong. This plum-pudding model, as it was often called, envisioned a roughly uniform distribution of positively and negatively charged particles. If an alpha particle penetrated such an atom, it would be attracted by the negatively charged particles and repelled by the positively charged ones. The net deflection, if any, would be small.

When Geiger announced that some alpha particles were deflected considerably, Rutherford realized they must be experiencing a strong electric repulsion when they entered an atom. That meant the charged components of an atom are not uniformly distributed. Instead the positively charged ones are collected in a highly compact core called the nucleus and negatively charged ones orbit the nucleus at distances that are large compared with the diameter of the nucleus.

Imagine following an alpha particle as it penetrates an atom made according to Rutherford's model. As it approaches the atom it experiences essentially no electric force because the atom as a whole is neutral. Once it passes through the orbits of the electrons, it begins to be repulsed by the nucleus because both the particle and the nucleus carry a positive charge. Suppose it travels directly toward the nucleus. The continuously increasing repulsion slows the particle to a momentary stop and then propels it back along the path of entry. The angle of deflection is 180 degrees.

An alpha particle entering an atom along a path offset from the nucleus is deflected less. The larger the offset, the smaller the angle of deflection. Since the nucleus is tiny, a small deflection is likelier than a large one. That is why Geiger and Marsden originally observed small deflections.

The distribution of the particles striking the detection screen is described in terms of the density of strikes per unit of screen area. The density depends on the scattering angle. The major strength of Rutherford's model for the atom was that it correctly predicted how the strikes would vary in density. Although several earlier workers had proposed similar solar-system models of the atom, Rutherford is credited with the first modern model because of his success in explaining the scattering of the alpha particles.

In last year's finals of the International Science and Engineering Fair, Rudy Timmerman of Wickes, Ark., won first place in the physics division with his experiments on the scattering of alpha particles from a thin gold foil. He set out to repeat the basic features of the experiment by Geiger and Marsden. In particular he wanted to verify Rutherford's mathematical prediction concerning the angular dependence of the scattering.

Timmerman relied on a detection scheme developed by Charles W. Leming of Henderson State University in Arkansas. Leming's apparatus, which is now manufactured commercially by the Daedalon Corporation (35 Congress Street, Salem, Mass. 01970), incorporates a special type of film that is sensitive to the impact of alpha particles. When it is developed, it has a small hole at every point where a particle struck it.

Timmerman constructed his own version of the apparatus. It consisted of three aluminum plates held by four long bolts. In each plate he drilled a small central hole. Above the hole in the top plate he positioned a source of alpha particles. Below the hole in the central plate he mounted a metal foil. On the upper surface of the bottom plate he placed the detection film.

He expected that the alpha particles would pass through the holes in the top two plates, forming a beam. When the particles passed through the foil, most of them would continue undeflected, but some would scatter into a small region surrounding the center of the detection film. Timmerman planned to measure the distribution of holes around the center in order to determine how the density of particles depended on the scattering angle. Although his plan was simple, he had to overcome several problems before he finally obtained the data he submitted at the fair.

Leming supplied Timmerman with the detection film. Leming also worked out a procedure for developing the film. After exposure the film is kept for 24 hours in a 2.5-molar solution of sodium hydroxide that is maintained at a temperature of 40 degrees Celsius (104 degrees Fahrenheit). Timmerman built an apparatus for developing the film. He puts the film and the sodium hydroxide solution in a canister that is lowered into a water bath inside a bucket. An aquarium heater warms the water and hence the contents of the canister. A cover and some insulation around the bucket slow the loss of heat to the room. A thermometer mounted on the inside of the bucket monitors the water temperature.

Before beginning his experiments on scattering Timmerman tested the strength of his source of alpha particles by putting a small piece of film

near the source for a few minutes. He developed the film and put it in a microfilm machine to magnify it. The film had hundreds of holes, indicating that the source was strong enough for his plans.

He then positioned the source and a new sheet of film in his apparatus but still did not insert the metal foil. The apparatus was placed in a bell jar, which was then evacuated with a pump so that the alpha particles would not be scattered by air molecules. After five minutes the film was removed, developed and examined. Again Timmerman found numerous holes spread over the center of the film. The apparatus seemed to be suitable for the scattering experiments.

Timmerman's next move was to put his thin gold foil in the apparatus. (He had acquired the foil from a painter of window signs.) The apparatus was evacuated and the film was exposed for two hours. When he looked at the film later, he discovered that the vacuum had been lost during the test. After he developed the film he found it was covered with holes, indicating that the alpha particles had scattered from air molecules and heavily showered the film.

Timmerman concluded that the bell jar surrounding the scattering apparatus was leaking air and therefore would have to be evacuated several times during a test. He tried to rig an automatic pump control by attaching an electronic switch to an input port of a home computer, which he programmed to turn on the pump every 30 minutes. The arrangement resulted in a burned-out switch. Timmerman rigged a switch to control a relay on the pump, but the relay began to stick. Eventually he discovered that the leak was not in the bell jar but in the vacuum pump. From then on he closed off the hose from the pump as soon as the air had been removed from the bell jar.

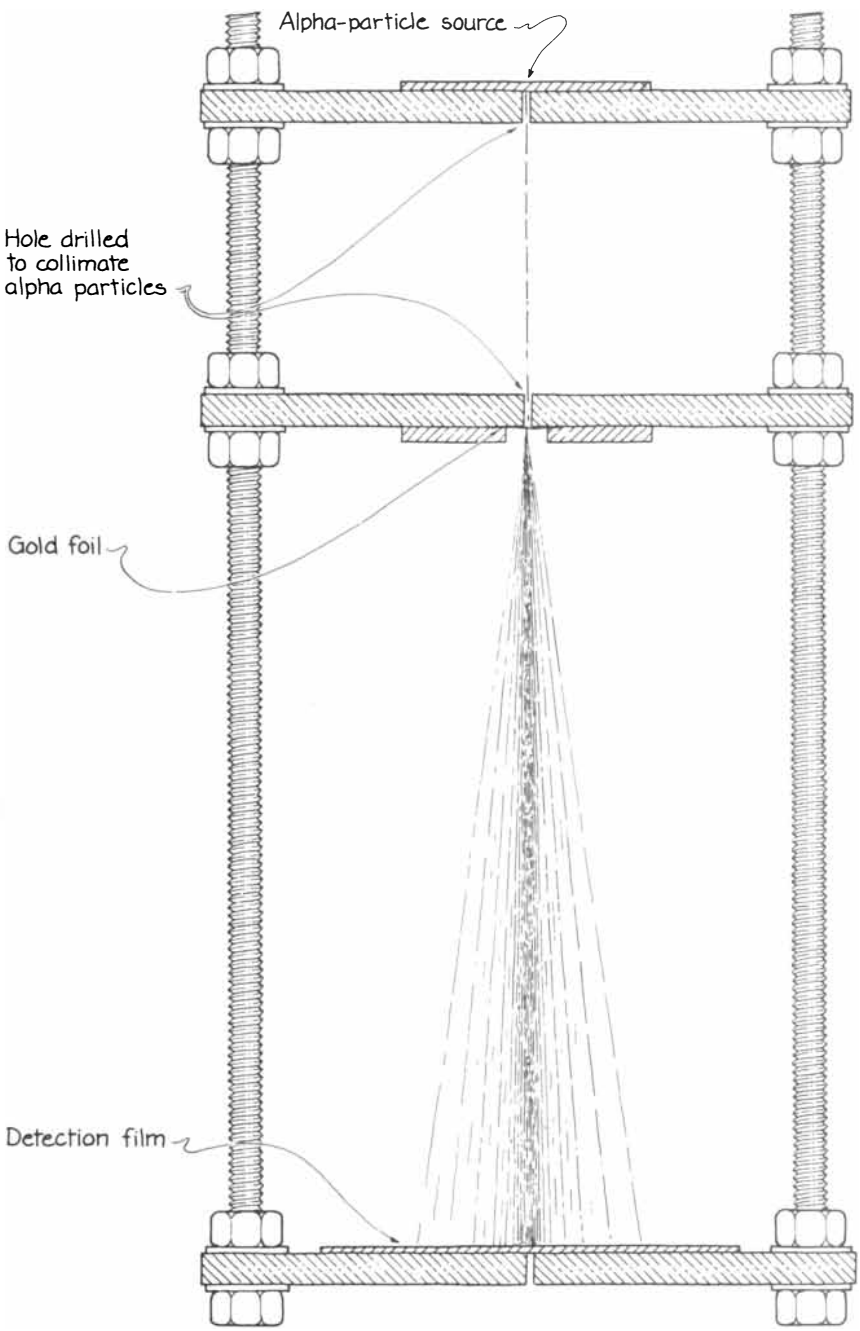
The next problem was a malfunction of the aquarium heater that resulted in the ruin of several sheets of film. Unable to find a replacement for the heater, Timmerman wired it to an interface with the computer. The computer was to monitor a temperature probe in the water bath and turn on the heater as needed. This procedure failed because the temperature probe did not function properly. Timmerman then decided to buy a new aquarium heater.

With a fresh supply of film he set out to determine the proper exposure time for his experiments. Seven hours seemed to work well. When a sheet of exposed film was placed in the microfilm machine, he could easily see the holes created by the alpha particles. Measuring the distribution of the holes with the microfilm machine proved to be too difficult, and so he decided to try a more automatic procedure incorporating his computer.

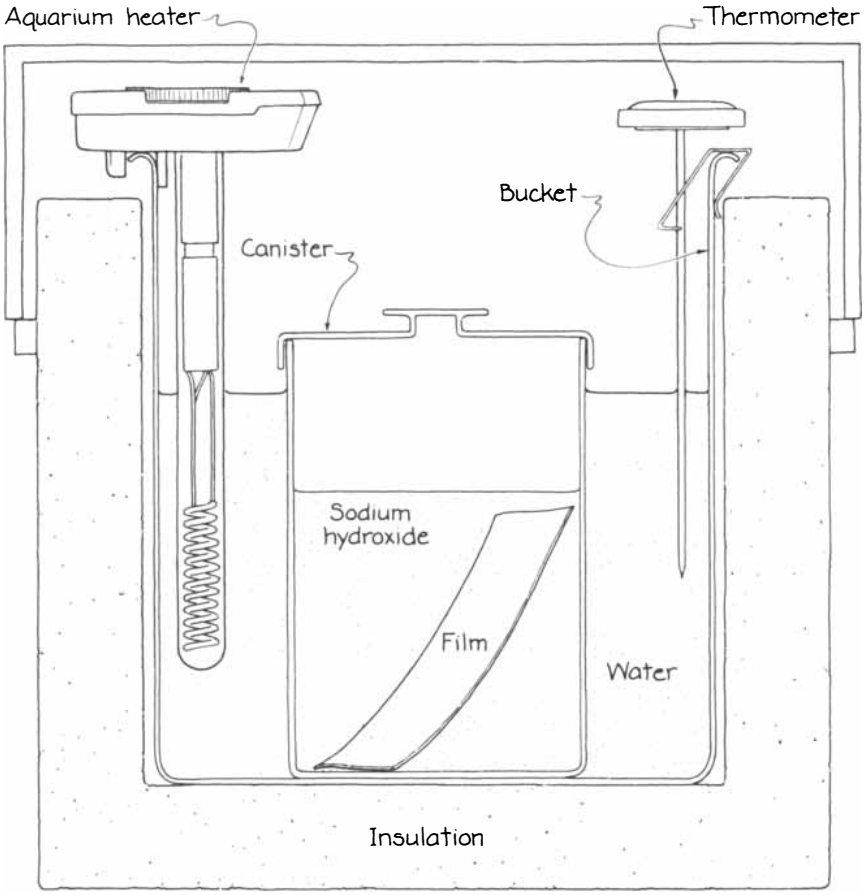
He enlarged his view of the holes in the film by projecting the film onto a wood screen he built. The holes appeared as bright spots. To chart the location of the spots Timmerman pivoted a wood arm around one of the lower corners of the screen. The arm bore a slide with a small aperture. He rotated the arm and moved the slide until a bright spot fell on the aperture.

Wires in the slide and the arm constituted a voltage divider. When the slide was near the pivot of the arm, the voltage across the wires was low. It increased as Timmerman moved the slide to the far end of the arm. Hence the voltage across the wires was proportional to the distance between the aperture in the slide and the pivot point of the arm.

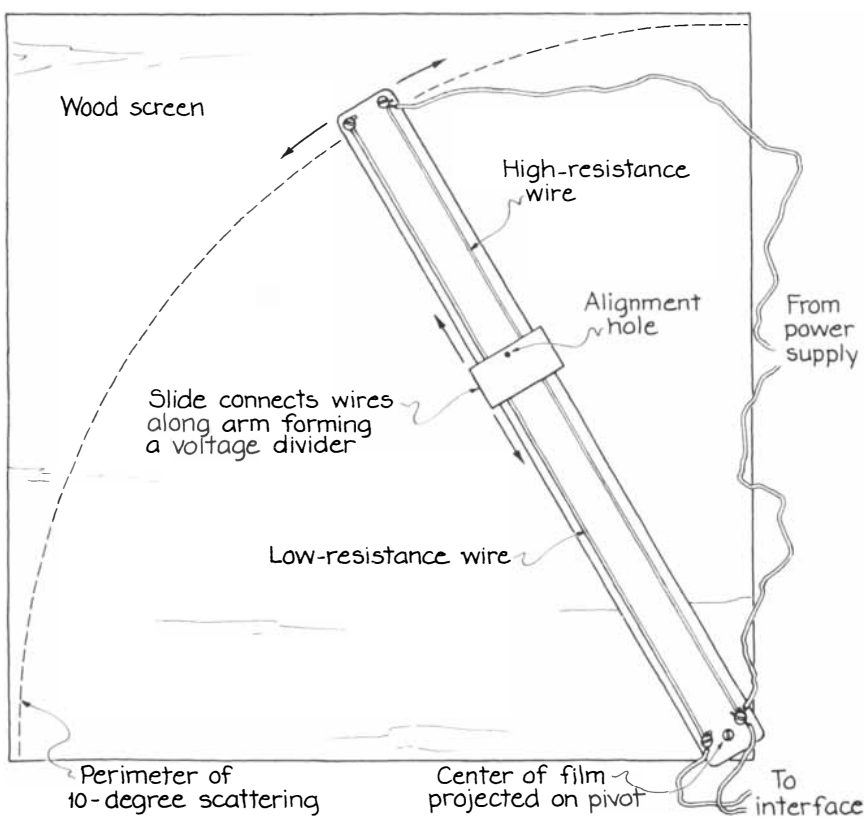
Timmerman arranged for the center of the film, the point reached by undeflected alpha particles, to project directly onto the pivot. He next moved the arm and the slide until the aperture in the slide was aligned with a bright spot. The voltage across the wires was



Rudy Timmerman's particle-scattering apparatus



How the special film is developed



Timmerman's device for recording the locations of holes in the film

then proportional to the distance between the center of the film and the point where the particle responsible for the bright spot struck the film.

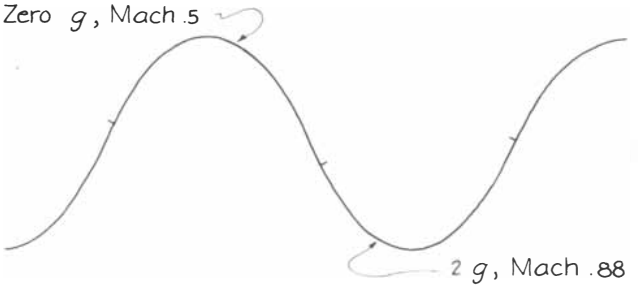
Timmerman was interested in the distribution of alpha particles within a scattering angle of 10 degrees. Therefore he set up his apparatus so that a bright spot from a particle deflected at 10 degrees fell at the outermost position of the slide. With the help of James A. Wisman of the University of Arkansas at Fayetteville, Timmerman constructed an interface between the voltage divider and his computer. The interface sectioned the maximum voltage from the divider into 256 steps. Thus each step corresponded to 10/256 degree of scattering.

To measure the scattering distribution on a film Timmerman moved the arm and slide to each bright spot in the projection. When the aperture in the slide was aligned with a spot, he triggered a switch so that a signal was sent to the computer from the interface. The signal indicated which voltage step (and so which angle of scattering) should be assigned to the spot. The computer recorded the information.

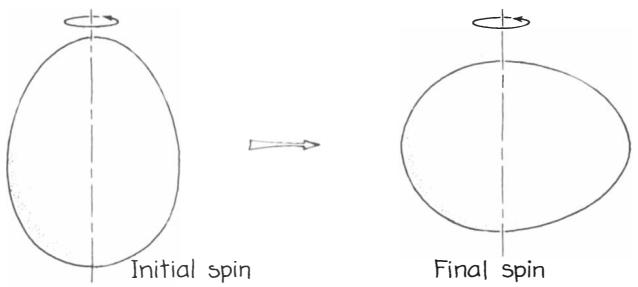
Timmerman then wrote programs that calculated the density of holes in the film and printed a histogram of the density versus the angle of scattering. Not satisfied with the results, he repeated the experiments with longer exposure times (10 hours and 18 hours). These tests generated such an abundance of holes in the film that hours were needed to measure their locations. Timmerman also corrected for a background of holes in the film, which were presumably due to alpha particles unrelated to his source. He tried various "best fit" functions to the data to find the angular dependence of the scattering.

Timmerman concluded that the angular dependence seemed to follow Rutherford's prediction. Nevertheless, he still had reservations about the collimation of the particle beam before it reached the foil, and so he tested his conclusion with two procedures. He weighed the gold foil in order to determine its thickness. Then with Rutherford's formula for scattering he calculated the thickness from his measurements of the density of holes in the film. The two findings agreed to better than an order of magnitude.

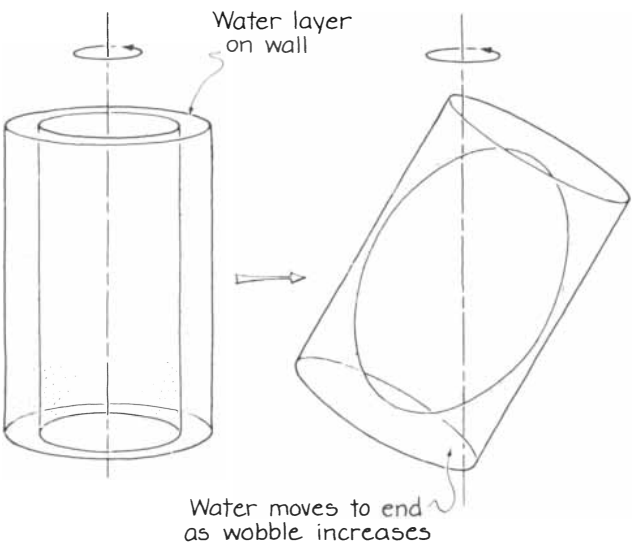
Gravity often modifies the behavior of solids and liquids in subtle ways. One technique for eliminating its influence is to study materials in free fall. Donald R. Pettit of the Los Alamos National Laboratory and astronaut Joseph P. Allen of the Johnson Space Center in Houston have recently



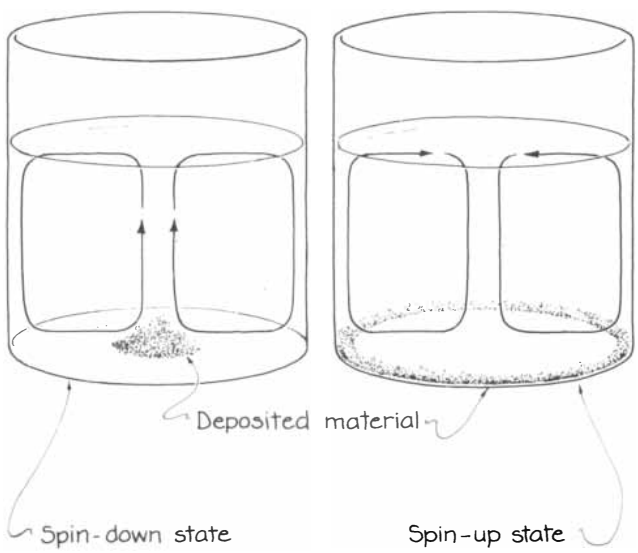
The flight path of the airplane for zero-gravity experiments



The egg experiment



A container of water rotating in zero gravity



Secondary flow in normal gravity

done experiments in free fall in the cargo space of an airplane belonging to the National Aeronautics and Space Administration. They were assisted by Robert K. Williams of the Johnson Space Center.

The airplane flew a series of between 40 and 60 vertical parabolic loops. As the craft neared the top of a loop with a speed of about Mach .5 (half the speed of sound in the surrounding air) the occupants began free fall. The floor of the airplane actually fell away from them, but the sensation was that gravity suddenly vanished. This state is referred to as zero *g*, where *g* symbolizes the normal strength of gravity.

About 20 seconds later the free fall ended near the bottom of the loop. For the next 50 seconds the airplane pushed upward on the occupants, creating the sensation that gravity was twice its normal strength, a state referred to as 2 *g*. The speed of the airplane at the bottom of the dive was about Mach .88. The brief time of free fall and the subsequent need to protect oneself and the equipment from the 2-*g* phase limited the studies to transient phenomena.

In addition to their serious research Pettit and Allen had time for a few recreational experiments. One experiment involved the stability of an egg spinning about its long axis. Try standing an egg on a table and spinning it on either end. A hard-boiled egg spins stably for some tens of seconds, but a fresh egg quickly becomes unstable, falls over and spins for a while about its short axis until friction from the table drains all its energy.

Pettit investigated the spin of eggs at zero *g*. On each egg he marked a line from end to end to enhance the visibility. Then he carefully spun each egg (in the air) about its long axis with as little initial wobble as possible. The hard-boiled egg continued to spin stably throughout the zero-*g* phase of the loop. The fresh egg completed about two revolutions and then abruptly began to spin about its short axis. Apparently the fluid in the egg was set in motion by the initial rotation, even at zero *g*. The fluid motion increased the wobble, making the egg spin about its short axis.

Pettit then did a similar experiment with a closed, transparent container partially filled with water. He released the container at zero *g* while giving it a spin about its long axis. The container soon began to wobble appreciably, but it never stabilized into rotation about its short axis before the end of zero *g*.

Pettit and Allen also studied fluid flow in rotating systems. Consider a cylindrical container of water that is placed at the center of a turntable. When the turntable begins to rotate, the wall of the container drags the water in a circle. Eventually the water circulates around the center of the container at the speed of the turntable. During the transition the water is said to be in a spin-up state. Suppose the turntable abruptly stops. The circulation slows and eventually stops. In this phase the water is said to be in a spin-down state.

During spin-up and spin-down an additional flow arises in the water. This secondary flow results from unequal pressures created in the water by the primary flow around the center. In spin-up the secondary flow is downward along the center line, outward along the bottom, upward along the wall and then inward along the top surface. In spin-down the secondary flow is reversed. Evidence for secondary flow is seen in the motion of tea leaves when the tea is in spin-down. The leaves, which initially are strewn over the bottom of the cup, are forced to

the center and then abandoned in a pile by the upward flow of water.

What about secondary flow in zero *g*? Pettit partially filled the transparent, closed container with water, adding half a teaspoon each of waterlogged sawdust and aluminum glitter. (The glitter is available in hobby and art-supply shops.) The sawdust and glitter served as tracers for the secondary flow.

In normal conditions of gravity the secondary flow of spin-down caused the glitter to collect like tea leaves in a small pile at the center of the container. The sawdust circulated in a ring just above the bottom center until near the end of spin-down; then it collapsed onto the pile of glitter. In spin-up the glitter moved to the wall first, followed by the lighter sawdust. At zero *g* Pettit released the container while spinning it. The water was in spin-up. The sawdust and glitter moved to the wall as before, but this time they did not collect along the bottom edge. The glitter was pressed against the wall and the sawdust moved up along the wall.

To generate spin-down Pettit made the container gyrate and then held it stationary. The sawdust and glitter moved along with the expected secondary flow, but they failed to pile up on the bottom.

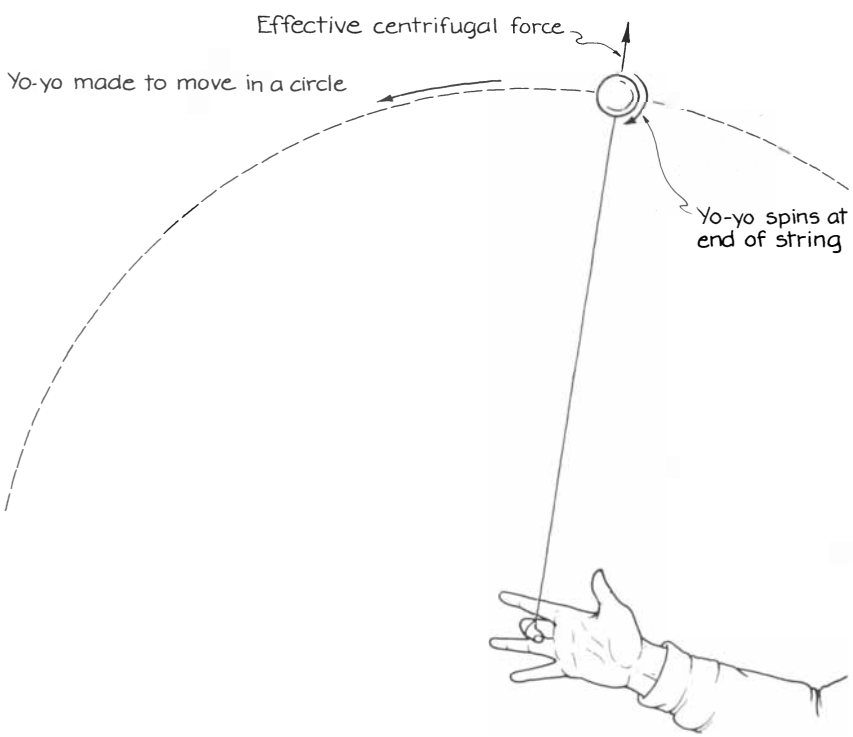
Apparently the secondary flow in spin-up and spin-down is the same at zero *g* as it is at normal gravity. In the effective absence of gravity, however, the sawdust and glitter are no longer confined to the bottom of the container. The secondary flow can carry them upward at the center of the container in spin-down and at the wall in spin-up.

Pettit and Allen did another experiment with the container. When water circulates about a container's long axis at normal gravity, the top surface is concave. Pettit and Allen wondered how the shape would change as the effective gravity varied. They found that at 2 *g* the concave surface was shallower. At zero *g* the concavity deepened enough to force all the water into a layer along the wall.

In a final experiment Pettit and Allen tested a yo-yo at zero *g* and at 2 *g*. They wondered if it could be made to spin at the end of its string, a trick called sleeping. At normal gravity you must let the yo-yo fall gently to the end of its string to minimize the usual bounce. Gravity holds the yo-yo there while it spins loosely in the loop around the spindle.

At 2 *g* Pettit easily made the yo-yo sleep. At zero *g* it refused to sleep even with a gentle toss. It always bounced. The only way Pettit could get it to sleep was to throw it outward and then pull on the string. That made the yo-yo circle around his hand. The resulting effective centrifugal force kept the yo-yo at the end of the string.

Pettit is interested in more experiments that might be done at zero *g*. If you have any ideas, write to him at the Los Alamos National Laboratory, MS P952, Los Alamos, N.M. 87545.



A BUILDING SITE LIKE NO PLACE ON EARTH.

Some structures needed in space are just too big to launch in one piece. Too large and too fragile even to stand alone on Earth, intricate sections can be brought up on successive shuttle flights, plucked from the orbiter by station robot arms, assembled into massive structures by the space station crew, and released to their own orbits. The miracle of microgravity will make light work of such space construction.

One example is the 20-meter diameter Large Deployable Reflector infrared telescope NASA is planning for the mid-1990s. Consisting of some 100 pieces–large mirrors and supporting structures–when assembled and deployed, the device will permit a variety of deep space investigations.

Without the manned space station, structures of this size would not be possible.

We think that's a very good reason to build a building site like no place on Earth.

With Eastman Kodak and NASA, McDonnell Douglas is planning the precision techniques needed for construction in space. We are teamed with Honeywell, IBM, Lockheed and RCA for major definition and design work on the station.

MCDONNELL DOUGLAS

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

COMPUTER RECREATIONS

CLASSIC CHESS PROBLEMS. Kenneth S. Howard. Dover Publications, Inc., 1970.

COMPUTER GAMESMANSHIP. David Levy. Simon and Schuster, 1984.

DIOXIN

THE CHEMICAL SCYTHE—LESSONS OF 2,4,5-T, AND DIOXIN. Alistair Hay. Plenum Press, 1982.

ACCIDENTAL EXPOSURE TO DIOXINS. Edited by F. Coulston and Francesco Pocchiari. Academic Press, 1983.

DIOXINS IN THE ENVIRONMENT. Edited by Michael A. Kamrin and Paul W. Rogers. Hemisphere Publishing Corp., 1985.

THE *T* CELL AND ITS RECEPTOR

H-2 COMPATIBILITY REQUIREMENT FOR T-CELL-MEDIATED LYSIS OF TARGET CELLS INFECTED WITH LYMPHOCYTIC CHORIOMENINGITIS VIRUS: DIFFERENT CYTOTOXIC T-CELL SPECIFICITIES ARE ASSOCIATED WITH STRUCTURES CODED FROM *H-2K* OR *H-2D*. Rolf M. Zinkernagel and Peter C. Doherty in *The Journal of Experimental Medicine,* Vol. 141, No. 6, pages 1427–1436; June 1, 1975.

ANTIGEN-INDUCIBLE, *H-2*-RESTRICTED, INTERLEUKIN-2-PRODUCING T CELL HYBRIDOMAS. LACK OF INDEPENDENT ANTIGEN AND *H-2* RECOGNITION. John W. Kappler, Barry Skidmore, Janice White and Philippa Marrack in *The Journal of Experimental Medicine,* Vol. 153, No. 5, pages 1198–1214; May 1, 1981.

BINDING OF IMMUNOGENIC PEPTIDES TO Ia HISTOCOMPATIBILITY MOLECULES. Bruce P. Babbitt, Paul M. Allen, Gary Matsueda, Edgar Haber and Emil R. Unanue in *Nature,* Vol. 317, No. 6035, pages 359–361; September 26, 1985.

QUANTUM CHEMICAL REACTIONS IN THE DEEP COLD

CHEMICAL REACTIONS AT VERY LOW TEMPERATURES. V. I. Goldanskii in *Annual Review of Physical Chemistry,* Vol. 27, pages 85–126; 1976.

QUANTUM MOLECULAR TUNNELING IN CHEMICAL REACTIONS. Vitalii I. Goldanskii in *Chemica Scripta,* Vol. 13, No. 1, pages 1–9; 1979.

FACTS AND HYPOTHESES OF MOLECULAR CHEMICAL TUNNELLING. Vitalii I. Goldanskii in *Nature,* Vol. 279, No. 5709, pages 109–115; May 10, 1979.

THE TUNNEL EFFECT IN CHEMISTRY. R. P. Bell. Chapman and Hall, Ltd., 1980.

SEISMIC IMAGES OF PLATE BOUNDARIES

SEISMIC MULTICHANNEL TOWED ARRAYS IN THE EXPLORATION OF THE OCEANIC CRUST. Manik Talwani, Paul L. Stoffa, Peter Buhl, Charles Windisch and John B. Diebold in *Tectonophysics,* Vol. 81, No. 3–4, pages 273–300; January 20, 1982.

ORIGIN OF SEAWARD-DIPPING REFLECTORS IN OCEANIC CRUST OFF THE NORWEGIAN MARGIN BY "SUBAERIAL SEA-FLOOR SPREADING." John C. Mutter, Manik Talwani and Paul L. Stoffa in *Geology,* Vol. 10, No. 7, pages 353–357; July, 1982.

MULTICHANNEL SEISMIC IMAGES OF THE OCEANIC CRUST'S INTERNAL STRUCTURE: EVIDENCE FOR A MAGMA CHAMBER BENEATH THE MESOZOIC MID-ATLANTIC RIDGE. John C. Mutter and North Atlantic Transect (NAT) Study Group in *Geology,* Vol. 13, No. 9, pages 629–632; September, 1985.

THE HEART AS AN ENDOCRINE GLAND

LE DEGRÉ DE GRANULATION DES CARDIOCYTES AURICULAIRES. ÉTUDE PLANIMÉTRIQUE AU COURS DE DIFFÉRENTS APPORTS D'EAU ET DE SODIUM CHEZ LE RAT. J. P. Marie, H. Guillemot and P. Y. Hatt in *Pathologie Biologie,* Vol. 24, No. 8, pages 549–554; October, 1976.

THE NATRIURETIC HORMONE AND ITS POSSIBLE RELATIONSHIP TO HYPERTENSION. H. E. de Wardener and G. A. MacGregor in *Hypertension: Physiopathology and Treatment,* edited by J. Genest, O. Kuchel, P. Hamet and M. Cantin. McGraw-Hill Book Company, 1983.

EFFECTS OF SYNTHETIC ATRIAL NATRIURETIC FACTOR ON RENAL FUNCTION AND RENIN RELEASE. John C. Burnett, Jr., Joey P. Granger and Terry J. Opgenorth in *American Journal of Physiology,* Vol. 247: *Renal, Fluid and Electrolyte Physiology,* Vol. 16, No. 5, pages F863–F866; November, 1984.

THE HEART AND THE ATRIAL NATRIURETIC FACTOR. M. Cantin and J. Genest in *Endocrine Reviews,* Vol. 6, No. 2, pages 107–127; Spring, 1985.

RECONSTRUCTING BIRD PHYLOGENY BY COMPARING DNA'S

THE PHYLOGENY AND CLASSIFICATION OF BIRDS BASED ON THE DATA OF DNA-DNA HYBRIDIZATION. Charles G. Sibley and Jon E. Ahlquist in *Current Ornithology,* Vol. 1, pages 245–292; 1983.

THE PHYLOGENY AND CLASSIFICATION OF THE NEW WORLD SUBOSCINE PASSERINE BIRDS (PASSERIFORMES: OLIGOMYODI: TYRANNI). Charles G. Sibley and Jon E. Ahlquist in *Neotropical Ornithology,* edited by P. A. Buckley, Mercedes S. Foster, Eugene S. Morton, Robert S. Ridgely and Francine G. Buckley. American Ornithologists' Union Monograph No. 36, 1985.

INCA STONEMASONRY

INCA CULTURE AT THE TIME OF THE SPANISH CONQUEST. John Howland Rowe in *Handbook of South American Indians,* edited by Julian H. Steward. Bureau of American Ethnology Bulletin 143, Vol. 2, Smithsonian Institution, 1946.

INCA ARCHITECTURE. Graziano Gasparini and Luise Margolies, translated by Patricia J. Lyon. Indiana University Press, 1980.

INCA QUARRYING AND STONECUTTING. Jean-Pierre Protzen in *Journal of the Society of Architectural Historians,* Vol. 44, No. 2, pages 161–182; May, 1985.

WILLIAM HERSCHEL AND THE MAKING OF MODERN ASTRONOMY

ON THE POWER OF PENETRATING INTO SPACE: THE TELESCOPES OF WILLIAM HERSCHEL. J. A. Bennett in *Journal for the History of Astronomy,* Vol. 7, Part 2, No. 19, pages 75–108; June, 1976.

HERSCHEL'S SCIENTIFIC APPRENTICESHIP AND THE DISCOVERY OF URANUS. J. A. Bennett in *Uranus and the Outer Planets,* edited by Garry E. Hunt. Cambridge University Press, 1982.

WILLIAM HERSCHEL, BATH, AND THE PHILOSOPHICAL SOCIETY. Roy Porter in *Uranus and the Outer Planets,* edited by Garry E. Hunt. Cambridge University Press, 1982.

THE AMATEUR SCIENTIST

THE DISCOVERY OF SUBATOMIC PARTICLES. Steven Weinberg. W. H. Freeman and Company, 1983.

də·'rek·shən

Direction

- *The line or course along which something is moving; e.g., a company surpassing $11 billion in sales, a leader in growth markets, zeroed in on the future.*
- *A governing or motivating process; aware management guiding highly resourceful and diverse areas of commercial and government business.*
- *A channel or course of action, like record performance ten straight years.*

Rockwell International is moving fast and smart into the decade ahead. We're leading the way in the development of products for high-technology growth markets.

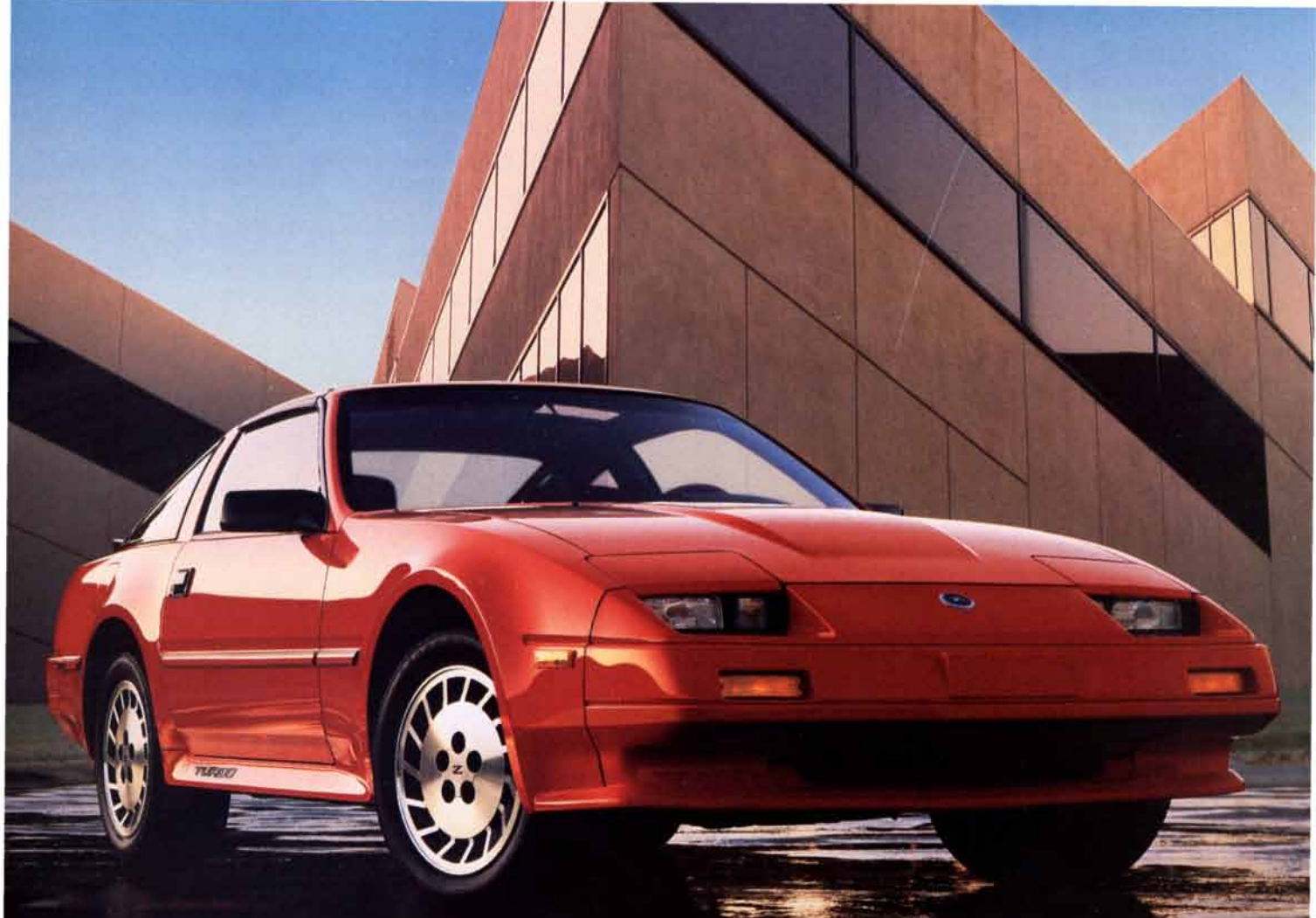
One example of that leadership is the work of our Allen-Bradley subsidiary, which is providing the automation and the electronic systems and controls for the revitalization of American industry.

And so it is with Rockwell's commitment to each of its five core business areas. Over the last five years, we have allocated more than $2.5 billion for capital expenditures. And we're continuing the pace for '86. To learn more about us, write: Rockwell International, Department 815S-7, 600 Grant Street, Pittsburgh, PA 15219.

Rockwell International

...where science gets down to business

Aerospace / Electronics / Automotive
General Industries / A-B Industrial Automation



THE NEW NISSAN 300 ZX PACE CAR FOR THE PERFORMANCE GENERATION.

Very few automobiles in the world have generated the excitement and emotional involvement associated with the Z-car. Keeping this in mind, it's not surprising to read that Motor Trend Magazine called the 300 ZX, "the best all-around Z-car ever built."

For 1986, Nissan has taken one more step in the thoughtful evolution of a classic.

At the heart of this Z is a 3-liter V-6 that is actually eight inches shorter and 15 percent lighter than the 280 ZX. Yet the turbo model puts out 11% more power; a rousing 200 horsepower. That power gets to the road by way of an electronic control system that gives you the most efficient transmission of power at any speed. Combined with shocks you adjust electronically from the cockpit, the result is startling.

In addition, an electronic monitoring system keeps track of spark plugs that fire 42 times a second at 5000 RPM and a micro computer controls the fuel injection system making the Z a marvel of functional electronic wizardry.

Outside, fender flares, housing wider tires, were integrated into the body. The air dam was extended and rocker panel extensions were added to reduce air turbulence under the car. All this, plus a wider track results in better handling than ever.

Inside, a choice of electronic or analog instrumentation is offered, along with every conceivable luxury, including a resounding 80-watt, 6-speaker stereo system.

The 300 ZX, turbo or fuel injected. Once you get inside a Z, a Z will get inside of you.

NISSAN

BELT YOURSELF

THE NAME IS NISSAN